Techies and Firm Level Productivity *

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

September 20, 2024

Abstract

We study the effect of techies—engineers and other technically skilled workers—on firm productivity. We use French administrative data and surveys to show that techies have STEM skills, and are associated with innovation, technology adoption and diffusion within firms. Using structural econometric methods, we estimate a positive and economically significant causal effect of techies on firm-level Hicks-neutral productivity in both manufacturing and non-manufacturing industries. This effect extends beyond R&D to ICT and Other techies. The effect of R&D techies is large in manufacturing firms, while productivity gains in non-manufacturing are primarily attributed to non-R&D techies.

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. An online appendix is available at https://sites.google.com/view/jamesharrigan/research/techies-and-firm-level-productivity.

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

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

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

1 Introduction

Engineers and other technically trained workers (techies) have long been recognized as fundamental in driving productivity. For example, engineers are at the heart of modern endogenous growth theory, as highlighted by Romer (1990). Indeed, economies that allocate more of their talent to engineering (versus lawyers) tend to grow faster, as shown by Murphy et al. (1991). The importance of techies in enhancing productivity has also been emphasized in the economic history literature.[1] In this paper, we study the role of techies in enhancing firm-level productivity. We show that techies raise firm-level productivity and that this effect extends beyond techies who do research and development (R&D). Techies that work with information and communication technology (ICT) and other technical tasks equally affect firm-level productivity. We also show that their effect is important not only in manufacturing but also in the non-manufacturing sector.

We start by providing a comprehensive description of techies based on precise administrative and survey datasets from the French National Institute of Statistics and Economic Studies, INSEE. We identify techie workers by using the French occupational classification (INSEE, 2003). Techie jobs are distinguished from other occupations by INSEE because they are related to the installation, management, maintenance, and support of technologies within firms: ICT, product and process design, long-term R&D activities, and other technology-related tasks. We show that techies are also distinguished from other workers by their STEM (science, technology, engineering, and math, including computer science) diplo-

[1] Kelly et al. (2014) and Ben Zeev et al. (2017) highlight the importance of the British apprentice system during the British Industrial Revolution in supplying the basic skills needed for technology adoption. Similarly, Kelly et al. (2023) show that the British Industrial Revolution started in areas where technically trained mechanics were abundant, and Hanlon (2022) shows how the emergence of "professional" engineers underpinned the Industrial Revolution. Maloney and Valencia Caicedo (2017) construct a dataset of engineer intensity for the Americas and U.S. counties around 1880 and show that this intensity helps predict income today. An early study of demand and supply of techies is Blank and Stigler (1957).

mas, skills, and experience. We also show that they adopt, manage, and diffuse technology within firms.

Techies are not homogeneous, and we classify them based on their specializations: R&D, ICT or Other technology-related occupations. This allows us to distinguish the impact of these three different types of techies on firm-level productivity. Importantly, R&D techies are much more common in manufacturing than they are in non-manufacturing, while the reverse is true for ICT techies. Therefore, limiting the focus to R&D techies alone does not provide an accurate picture of the overall influence of techies across industries.

A large literature has studied the role of R&D expenditure in shaping firm, industry and national outcomes. Our firm-level analysis uses the wage bill of R&D workers instead of total R&D expenditures, which is not a limitation for two reasons. First, we show that most of R&D expenditure in France is on wages, and by a large margin, compared to other R&D-related expenditures. Consistent with this, R&D wages are highly correlated with total R&D expenditures at the firm level. Second, non-wage R&D expenditures are included in our measure of the firms' purchased inputs and capital.

The right way to measure firm-level productivity differences is contentious, but there is broad consensus that these differences are very large. There is much less consensus about, to echo the title of the influential survey by Syverson (2011), what *determines* productivity differences. As noted by De Loecker and Syverson (2021), only a few papers have tried to answer this question in a structural way, which requires a methodology that permits both consistent estimation of firm-level productivity and its causal determinants. Our paper adds to this literature in two dimensions: we are the first to jointly study the impact of R&D, ICT, and other techies on firm-level productivity, and also the first to study firms in non-manufacturing in addition to manufacturing. This broadened focus allows us to paint a more complete picture of the overall influence of techies on firm-level productivity.

Our analysis of the survey data is complementary to the structural econometric analysis. The three surveys that we analyze (one at the individual level and two at the firm-level) allow us to study the qualifications and tasks of techies, and how techies are correlated with firm-level innovation effort and outcomes. This lends credence to the structural analysis, which is based on administrative data. We use the administrative data to construct a firm-level unbalanced panel of manufacturing and non-manufacturing firms from 2011 to 2019. The panel includes data on firms' inputs (capital, labor by detailed occupation, and expenditure on materials) and revenue, as well as an indicator for exporting. We use the panel to estimate structural models of firm-level Hicks-neutral total factor productivity (TFP) and the causal effect of techies and exporting on productivity. We use two recent structural production function estimators, due to Grieco et al. (2016) (hereafter, GLZ) and to Gandhi et al. (2020) (hereafter, GNR), which have different advantages and disadvantages for our application that we discuss below.

Our econometric strategy is based on two assumptions. First, techies affect Hicks-neutral TFP with a lag. Second, techies affect output only through their impact on future productivity, and not through any contemporaneous contribution to factor services that affect current output. This is analogous to the way economists usually think about current investment spending, which doesn't increase current output but increases future output through its impact on future capital stock. This is also how economists usually think about R&D expenditure, affecting only future outcomes. We use a flexible specification of the firm's productivity process, which permits us to make causal statements about the effects of firms' employment of R&D, ICT and other techies, as well as export status. We control for exporting in all of our specifications, which has been argued to have a separate effect on productivity De Loecker, 2013. In Section 6.3 we show that our results are not sensitive to the inclusion of managers in the equation explaining productivity.

We find that firms that employ techies have substantially higher future productivity than those who do not. The presence of techies leads to 4 or 5 percent higher productivity a year later, with a long run effect of over 45 percent in both manufacturing and non-manufacturing firms. Our analysis confirms the importance of R&D techies for productivity in manufacturing firms, as in Doraszelski and Jaumandreu (2013). In addition, we find that the positive impact of techies on productivity is not limited to R&D. ICT and other techie workers also positively impact productivity in manufacturing and non-manufacturing industries. Interestingly, R&D techies do not significantly contribute to the productivity of non-manufacturing firms.

TFP is defined as real output per unit of real inputs. However, our data reports revenue rather than real output, and expenditures on materials rather than quantities—a typical feature of firm-level datasets and of productivity studies. We address these challenges by applying the estimator of Grieco et al. (2016), which was developed for such datasets.

The GLZ estimator rests on three main assumptions. First, it assumes that all firms in an industry have the same constant elasticity of substitution (CES) production function. Second, it restricts returns to scale to be constant. Third, it assumes that both materials and labor inputs fully and flexibly adjust in response to current productivity shocks. We examine the sensitivity of our results by extending the methodology of Gandhi et al. (2020) to our setting, where real output is not observed.

Unlike GLZ, GNR imposes no functional form restrictions on the production function and does not require constant returns to scale. Furthermore, GNR's flexibility in accommodating labor as a "dynamic" (predetermined in period t) input is particularly attractive given the labor market institutions in France. We employ two versions of GNR: one in which both labor and materials are "static" inputs, as in GLZ, and another, in which labor is dynamic and does not respond to current productivity shocks. However, our application of GNR comes with two drawbacks: (1) it assumes that real materials input quantities are known—while they are not, and (2) we can only identify the impact of techies on productivity up to an unknown constant parameter. Despite the differences between the estimators, our estimates of the impact of techies on productivity using the GNR methodology are qualitatively similar to those we obtain using GLZ.

Our assumption that techies don't affect current output but do affect future productivity is key to our research design. We examine the validity of this assumption by considering the simple null hypothesis that techies are no different than other workers and reject this null in favor of the alternative that our baseline assumption is a better fit to the data. We also show that our inferences about the effect of techies are robust to a nonlinear adjustment process and to a re-classification of Other (not R&D nor ICT) techies as regular labor.

Related research. A small literature examines the impact of techies' impact on output, employment structure, and productivity at the firm level. The motivation for this literature is stated succinctly by Tambe and Hitt (2014): "the technical know-how required to implement new IT innovations is primarily embodied within the IT workforce". Similarly, Deming and Noray (2018) show that, in their words, "STEM jobs are the leading edge of technology diffusion in the labor market". While the literature on firm-level impacts of investment in IT and in R&D is vast, it rarely studies the importance of those key workers who install, manage and diffuse IT and other technologies within the firm.

A lack of firm-occupation-level data in most administrative and survey datasets has hampered firm-level research on this proposition. An exception is Harrigan et al. (2021), which uses detailed occupational data (including data on techies) for the entire French private sector from 1994 to 2007. Harrigan et al. (2021) show that employment growth is higher in French firms with more techies and also that more techies lead to within-firm skill upgrading. Lichtenberg (1995) and Brynjolfsson and Hitt (1996) find that IT labor has a positive output elasticity, a result confirmed on later data by Tambe and Hitt (2012). Using a remarkable dataset that tracks the movement of IT workers across firms, Tambe and Hitt (2014) find what they interpret as evidence for knowledge spillovers across firms through the channel of techie mobility. Recently, Brynjolfsson et al. (2024) use U.S. online job postings to create measures of the firm's IT usage. None of these papers structurally estimate the impact of techies on productivity, nor do they study the different tasks that techies perform (e.g., IT versus R&D). Hall et al. (2013) consider the separate roles of R&D and ICT investments in a sample if Italian firms. Differently from us, they study the impacts on sales per worker, not productivity *per se* and, moreover, they do not take a structural approach to estimating productivity and the impact of firm decisions on productivity. Similar to our work, they too study the correlations of R&D and ICT investment on innovation. However, they do not study the role of other productivity-enhancing efforts and non-IC Technologies, which we capture with our Other Techies category.

We rely on recent advances in the methodology of estimating firm-level productivity and its determinants. This literature was initiated by Olley and Pakes (1996) (OP) by estimating production functions and associated firm-specific, time-varying Hicks-neutral total factor productivity differences. Other key methodological papers in this literature include Levinsohn and Petrin (2003) (LP) and Ackerberg et al. (2015) (ACF), a set of techniques which we will refer to as OP/LP/ACF. The common thread that runs through these papers is that they apply the "control function" approach for identifying the production function. Countless papers have applied the OP/LP/ACF methodology to estimate TFP, but the study of the determinants of firm-level TFP is remarkably sparse.

Two pioneering papers that study the determinants of firm-level TFP are De Loecker (2013) (exporting) and Doraszelski and Jaumandreu (2013) (expenditure on R&D). We discuss these papers below, as our methodology relies on their insights. The methodology of Doraszelski and Jaumandreu (2013) requires observing real inputs and outputs, a specific functional form for the production function, and assumptions on labor flexibility. As discussed above, our applications of GLZ and GNR address these limitations in our setting, in different ways.

Two serious concerns have recently been raised for the control function approach. First, Gandhi et al. (2020) identify a weak instruments problem. Second, Ackerberg et al. (2021) show that the control function approach suffers from a "weak moments" problem, where the GMM objective function admits multiple solutions with equal value of the problem. These problems are not present in the GLZ and GNR estimators, which further motivates us to apply them, rather than the OP/LP/ACF approach.

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, comparing the relative advantages of the GLZ and GNR estimators. There we also 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.[2] We complement this information with survey data to characterize techies and their roles in firms. We highlight key features of the data here and relegate other details to the online appendix O.1.

2.1 The composition of labor within firms

Our employment data comes from the administrative source DADS.[3] Firms must report wages, hours paid, occupations, and sector of activity. Our labor input measure is hours paid. We checked the sensitivity of our results to adjusting labor input for quality by multiplying hours of lower-paid, less-qualified workers by their wage ratio to higher-paid, qualified workers, as in Gandhi et al. (2020). Fox and Smeets (2011) show that this lowers

[2] 2011 is the first year for which our data are available and 2019 is the last year pre-COVID19 pandemic.
[3] *Déclaration Annuelle de Données Sociales.*

productivity dispersion, but it has no qualitative effect on our results. The sample includes firms in 17 industries across manufacturing and non-manufacturing.[4]

The DADS provides nearly 500 detailed 4-digit occupational codes, classified using the detailed French PCS system (INSEE, 2003). From these, we identify 56 occupations as techies. As shown in Section 3, these workers differ from others in their education, training and tasks. Techies are distinguished from other occupations by INSEE because they are related to the installation, management, maintenance, and support of technologies within firms: ICT, product and process design, long-term R&D activities, and other technology-related tasks. In summary, techie employment reflects firms' technology investment.

For analytical purposes, we group the techie occupations according to three categories defined by us: *R&D techies*, *ICT techies* and *Other techies* (see Table O1). These categories are based on the definitions and descriptions of the 4-digit categories in INSEE (2003).

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 are made between technicians and other occupations.[5] Beyond what is suggested by their occupational titles, 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.

Our classification of techies into R&D and ICT techies is unambiguous and is based on a reading of the occupational definitions. 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"

[4]We exclude sectors with minimal hours worked or where estimation failed to converge.

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

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, despite being notably different from production and fabrication activities, they *optimize* the productivity of workers in those fields. In our baseline results below, we include Other techies along with R&D and ICT techies as drivers of productivity, but we also report results that treat Other techies as ordinary workers who contribute to current output rather than improve productivity with a lag; our results are not sensitive to this.

Hours worked in non-techie occupations are assumed to contribute directly to current output, as is standard in the structural productivity estimation literature.

2.2 Balance sheets and exporting

Firm balance sheet data is sourced from the FARE dataset (2011–2019).[6] This data is based on firms' tax declarations. We use total revenues, material expenditures, and the series needed to construct firm capital stocks. Appendix O.1 describes the source data and explains how we construct firm-level capital stocks.

French Customs data is used to create an export status indicator for each firm-year.

2.3 Survey data

We motivate and complement the structural estimation of techies' impact on productivity by collecting information from three survey data sources, which offer insights into techies' roles in firms. First, we provide information on education in STEM fields (Science, Technology,

[6] *Fichier Approché des Résultats Ésane*

Engineering, and Math, including Computer Science) and STEM training of techie workers from the Training and Professional Qualification (TPQ) survey. The survey collects data in 2015 on the specialization of the highest degree obtained by the individual, including degree specializations and post-degree training.

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 ICT survey, which informs on the relationship between ICT training and technology diffusion within firms.

Appendix O.1 provides detailed datasets descriptions. Both ICT and R&D surveys can be linked to the administrative datasets since they use the same SIREN firm identifier.

3 Facts about techies

This section uses both DADS and survey data to present a comprehensive overview of techie workers, focusing on their education, training, and role in the adoption, mediation, and diffusion of technology within firms. We verify that the surveys are representative of the private sector data. We present key descriptive results here, and relegate additional findings and analysis for each of the following five facts to the Appendix.

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

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

	Overall	Manufacturing	Non-Manufacturing	% techie wage bill in manufacturing
	(1)	(2)	(3)	(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

Source: DADS. Columns (1), (2) and (3) report the wage bill share of Techies or their sub-categories in the private sector overall, within manufacturing and within non-manufacturing industries, respectively. Column (4) reports the share of the Techie wage bill (or sub-categories thereof) that is in manufacturing.

In column (4) we see that most (62.6%) expenditures on techies are in manufacturing, while around 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. Again, this motivates studying the non-manufacturing sector when studying the impact of techies on firm-level productivity.

Fact 2. Techies have more STEM education and training than other occupations. We use the 2015 Training and Professional Qualification (TPQ) survey to examine whether techies have more STEM education and training compared to other workers. The TPQ survey provides detailed data on individuals' highest degree specialization and any subsequent training. We classify these degrees and training and build an indicator for STEM (see Appendix). The TPQ survey has 26,861 individuals with valid observations.

As we report in Table A1, 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 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. These numbers are very similar for the important skilled occupation of administrative and commercial managers (PCS 37). These findings support the idea that "*the technical know-how required to implement new IT innovations is primarily embodied within the IT workforce*" in a firm (Tambe and Hitt, 2014), and that their role is distinguished 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 excludes much of the ongoing expenditure and managerial attention that firms devote to technology adoption and ICT use. 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 related reasons. First, total materials are included as an input to production, and it is not possible to extract R&D materials costs from total materials. Second, R&D capital expenditure is part of the firm's total investment, which is used to construct firm capital stocks. In addition, capital investment often occurs in "spikes", potentially overestimating productivity-enhancing efforts during investment years and underestimating them in other periods.

As shown above, firms employ many scientists and engineers in non-R&D occupations.[7] Moreover, performing R&D requires hiring techies. Thus, the techie wage bill provides a broader and potentially more precise measure of firm-level efforts toward technology adoption and diffusion than R&D expenditures.

The R&D survey shows that labor costs account for most of R&D spending (67%),

[7] Barth et al. (2017) find that 80 percent of U.S. private sector scientists and engineers worked outside R&D occupations in 2013.

especially "in-house" R&D (75%), while the median firm spends nothing on external R&D services (Table A2). 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.[8] Similarly, Schweitzer (2019) finds that in 2014, labor costs accounted for 60 percent of aggregate R&D spending in France.[9]

One potential threat to our approach that treats firm-level techies as an indicator of firm-level technological sophistication is that firms can purchase ICT, R&D, and other technology-related consulting services. Table A2 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 three percent of total 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.

Fact 4. Techies are positively associated with the diffusion of ICT skills within firms. We match the ICT survey to the administrative DADS data and find that only 18% of firms offer training to improve ICT skills, including for ICT workers (Table A7). We then analyze the relationship between techies and ICT training using a linear probability model. Our regressions control for firm size, and include sector and year-fixed effects.[10]

We interpret ICT training as an investment in worker skills. We find a strong association between the likelihood of offering ICT training and the employment of techies, even after controlling for firm size. This relationship is particularly strong for ICT techies, while R&D and other techies are less strongly associated with ICT training.

[8] 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)"

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

[10] The results are reported in Table A8. 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.

Fact 5. Techies are positively associated with patenting and innovation. The R&D survey provides data on firms' patent filings and product and process innovation. We examine the relationships between patenting, innovation, R&D spending, and techies, by matching the survey outcomes with the information on techies from the DADS data.

As expected, we find that 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. By 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 A9).

We estimate regressions with either patenting or innovation as dependent variables, and control for industry-by-year fixed effects and firm size. We find that patenting correlates positively with all types of R&D expenditures in the R&D survey: internal or external, wages or other expenses (Table A10). The strongest correlate for innovation and patenting is R&D labor costs in internal R&D. When we match the R&D survey to the DADS we find a positive correlation between the techie wage bill and firms' patenting (Table A11). This correlation is strongest for R&D techies, with a smaller correlation for ICT techies.

Using the matched dataset we also find that techies are positively related to both product and process innovation (Table A12). Interestingly, the R&D and ICT techie wage bills are similarly correlated with product innovation (although in non-manufacturing the relationship for ICT techies is not statistically significant). In contrast, Other techies are uncorrelated with product innovation. The R&D and Other techie wage bills are positively related to process innovation (although in non-manufacturing the relationship for R&D techies is not statistically significant). In contrast, ICT techies are not associated with process innovation.

The analysis here, with further details in the online appendix, shows a clear relationship between techies and innovation. Different techie roles emerge: R&D techies are linked to both product and process innovation, ICT techies to product innovation, and Other techies to process innovation.

Our findings align with Hall et al. (2010), who link R&D to both product and process innovation. Like other studies on R&D's impact on productivity, we do not distinguish the specific innovation channel. Arora et al. (2017) highlight how U.S. corporate research drives innovation and patenting, with productivity gains tied to researcher quality, often reflected in wages. Techies likely mediate many process innovations and represent the "organizational capital" essential to ICT adoption (Brynjolfsson and Hitt, 2003). ICT investments foster organizational changes within firms such as business processes and work practices (Bresnahan et al. (2002)) and span of control (Bloom et al. (2014)), both of which may enhance productivity (Brynjolfsson and Hitt (2000)).

4 Why don't all firms employ techies?

The previous section shows that techies are essential to adopt, mediate, and diffuse technology at the firm level, potentially enhancing productivity. Nonetheless, we also show that relatively few firms employ techies, which raises an obvious question: why don't more, if not all, firms employ techies? This mirrors a well-known observation in trade economics—exporting can boost productivity, yet few firms export. Following Melitz (2003), the consensus explanation for this phenomenon is fixed costs: firms choose to export only when the extra revenue from exporting exceeds the fixed costs of exporting. Alternatively, the variable costs of exporting may make it unprofitable for high-cost firms, as shown by Melitz and Ottaviano (2008). Here we sketch a simple theoretical framework that makes a similar point about techies and gives a rationale for a constant elasticity relationship between techies and productivity. We do not estimate this model; rather we use it here to make a few simple theoretical points.

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

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

Fixed costs of employing positive techies are κ_f and their wage 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 intuitive conclusions in the online appendix O.2. 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 meaningful increase in productivity, compared to their cost. Finally, when the optimal amount of techies is positive, it is increasing in initial productivity.

These predictions are in line with Brynjolfsson et al. (2023), who find that larger firms, benefiting more from IT investments, have a higher incidence of such investments. An additional implication is that exporting firms that face higher demand, are more likely to employ techies. This motivates us to control for exporting in our structural analysis.

5 Methodology

We now describe our methodology for estimating the impact of Techies on firms' productivity. Firm-level TFP is defined as real output per unit of real inputs (Caves et al., 1982). Most of the production function estimation literature assumes that all necessary real quantities are available. However, our data reports revenue rather than real output and expenditures on materials rather than quantities, which is the case in the large majority of productivity studies.[11] We build on two methodologies, proposed by Grieco et al. (2016) (hereafter, GLZ) and an extension of Gandhi et al. (2020) (hereafter, GNR), that address these data issues in different ways. Both methodologies have drawbacks and advantages, which we discuss

[11] For a discussion of the challenges that such a data environment poses for estimation, see De Loecker and Goldberg (2014).

below.

Our research strategy rests on the assumption that that techies affect output only through their impact on *future* Hicks-neutral productivity, not through any contemporaneous contribution to factor services that affect current output. This assumption is analogous to the common assumption in the literature that investment in R&D has no contemporaneous effect on output, but raises future output through its contribution to capital (Doraszelski and Jaumandreu, 2013). Similarly, Beaudry et al. (2016) use a model with cognitive labor affecting future output only through its effect on organizational capital.

Both methodologies lead to isomorphic estimation equations (3) and (4). Here we follow the steps according to GLZ and refer the reader to the online appendix O.3 for the derivation of GNR. We start with specifying the firm's production function $Q_{ft} = \Omega_{ft} F(\mathbf{x})$, for some input vector $\mathbf{x}$ and Hicks-Neutral productivity Ω. Taking logs this becomes

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

where all lower case letters denote logs of upper case variables and functions. Firms face an industry-level demand function, $Q_{ft} = B_t P_{ft}^{-\eta}$, 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 $\rho \in (0,1)$.[12] Taking logs this becomes

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

In our data we do not observe output in quantities, but only firm revenues (sales). *Anticipated* revenue when making decisions based on time t information is given in logs by $E_t(r_{ft}) = q_{ft} + p_{ft}$. To this we add u_{ft}, which is an *unanticipated, ex post* i.i.d. revenue shock that materializes after production decisions were made, such that observed revenue in logs is

[12]It is well know that this demand curve can be derived from an utility function that exhibits constant elasticity of substitution across horizontally differentiated varieties within a product category.

$r_{ft} = q_{ft} + p_{ft} + u_{ft}$. Thus, we allow for transitory firm-specific demand shocks, which give rise to *ex post* variable markups across firms.

By manipulating (1) and (2) and adding an equation for the evolution of log productivity we arrive at the two equations that frame our research strategy:

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

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

Equation (3) is a firm-level "revenue production function" that is common to all firms in the industry. While both ω and u affect revenue and are unobserved by the econometrician, the important distinction between the two is that the firm observes ω before making variable input choices—while u is not, i.e., it is unexpected. Equation (4) describes the evolution of log productivity, assuming a controlled Markov process. $\mathbf{z}_{ft-1}$ is a vector that includes techies, and ξ_{ft} is a shock to productivity that is realized after $\mathbf{z}_{ft-1}$ is chosen. Techies appear with a lag, and only in equation (4), not as an input to current production in (3). While our assumption that techies affect output only through their effect on future productivity is well-grounded, it is important to consider how our productivity measurement could go awry if techies directly increase current output. We return to this issue below.

5.1 Controlled Markov productivity.

Equation (4) is a generalization of the Markov productivity assumption made by the pioneering OP/LP/ACF methodologies. Before discussing the estimation of equation (4), it is important to clarify what is meant by a "controlled Markov process". In particular, how should any estimated effects of the elements of $\mathbf{z}_{ft-1}$ that are chosen by the firm, such as the employment of techies, be interpreted? The key is that the Markov assumption breaks realized productivity into expected and unexpected components, with the function g mapping ω_{ft-1} and other firm-level decision variables $\mathbf{z}_{ft-1}$ into expected future productivity,

$E_{t-1}\omega_{ft} = g(\omega_{ft-1}, \mathbf{z}_{ft-1})$. Thus, orthogonality of $\mathbf{z}_{ft-1}$ and ξ_{ft} in (4) is assured, which allows us to interpret the estimated effects of techies as causal.

De Loecker (2013) discusses how to interpret the learning-by-exporting effect in the context of a controlled Markov process. He emphasizes two things. First, it is *lagged* exporting that enters the Markov process, which is to say that productivity (more precisely, the shock to productivity ξ_{ft}) is realized after the exporting decision is made. Second, persistence in the exporting decision is controlled for by having lagged realized productivity in the controlled Markov equation. This means that the coefficient on exporting in $t-1$ captures the *incremental* impact on productivity in t. These arguments extend directly to our setting, where lagged employment of techies takes the place of lagged exporting.

In section 4 above, we presented a simple model of optimal techie choice. However, equation (4) can consistently estimate techies' effect on productivity even without a structural model of techie choice. Similar to Doraszelski and Jaumandreu (2013), our estimated effects of techies on productivity are conditional on the decision of firms to employ techies. That is, they capture the causal effect of the choice to employ techies for those firms that decided to do so.

As in both De Loecker (2013) and Doraszelski and Jaumandreu (2013), identification of the effects of firm choices on productivity is based on cross-sectional differences in productivity between firms that make a given choice and firms that do not. For example, consider two firms with the same lagged productivity and all other explanatory variables except that one firm chooses to employ techies and the other does not. If the firm with techies has higher productivity in the next period, the estimator attributes that to the firm's techies.

5.2 Estimation using the GLZ and GNR estimators

Our objective is to consistently estimate (4). As mentioned above, the data report revenue rather than real output and expenditures on materials rather than quantities. GLZ and our extension of GNR's estimator address these data issues in different ways.

Before describing the pros and cons of the GLZ and GNR methodologies, we discuss briefly why we do not apply the OP/LP/ACF methods, widely known as the "control function approach". Under this approach, the identification of (3) relies on the timing assumptions in equation (4). In this context, when $\mathbf{z}_{ft-1}$ as well as ω_{ft-1} affect expected productivity, De Loecker (2013) and Doraszelski and Jaumandreu (2013) make the important point that equations (3) and (4) must be estimated jointly.[13] GNR demonstrate that without sufficient input price variation the control function approach suffers from a weak instruments problem and is inconsistent. Ackerberg et al. (2021) demonstrate a more severe identification problem due to multiple global minima for the GMM optimization problem for the control function approach, which makes the estimates sensitive to the initial values given to the GMM search.

GNR and GLZ do not suffer from these problems. GNR circumvents these problems by identifying the output elasticities of variable inputs in a way that does not rely on the timing assumptions of OP/LP/ACF. After doing this, GNR jointly estimates (4) with the remainder of the production function. GLZ identifies (3) independently from (4) by making structural assumptions that we discuss below. When we apply the GLZ estimator, we use the productivity estimates that we obtain from (3) in a second step to identify (4).

The GLZ methodology. GLZ develop an estimator that overcomes the problem of missing material input quantities and instead uses expenditures on materials in a theory-consistent way. The GLZ estimator does not rely on (4) for identifying the production function. However, this comes with some additional structural assumptions. First, the GLZ estimator assumes that all firms in an industry have the same constant elasticity of substitution (CES) production function. Second, it restricts returns to scale to be constant. In addition, GLZ assume that a constant elasticity demand curve gives firm-level demand, a strategy first used by Klette and Griliches (1996).

Rather than using proxy methods as in the OP/LP/ACF methodology, GLZ use first-

[13] One reason is that decisions on $\mathbf{z}$ in $t-1$ may be correlated with investment decisions in $t-1$ that affect capital in t. Failing to control for $\mathbf{z}_{ft-1}$ when estimating the production function will lead to an inconsistent estimator.

order profit-maximizing conditions to eliminate productivity and unobserved materials input quantities from the estimating equation. The assumption about demand permits estimation of the demand elasticity, which is used to implicitly calculate firm-level prices and thus convert revenue into real output. Remarkably, the GLZ estimator can be computed by nonlinear least squares (NLLS), without recourse to instruments or assumptions about productivity's stochastic process. After estimating the industry production function and ρ, one can recover log productivity for each firm f and year t in a given industry, $\{\widehat{\omega}_{ft}^{GLZ}\}$.

Key to the derivation of the GLZ estimator is the assumption that at least two inputs, materials plus at least one type of labor, are chosen optimally after ω_{ft} is observed. In the literature, such inputs are referred to as "static", in contrast to "dynamic" inputs, such as capital, that are either predetermined or adjust only partially to realized ω_{ft}. Exploiting the CES functional form, GLZ manipulate the firm's optimality conditions and use expenditures on materials, expenditures on labor and quantities of labor input to derive materials quantities. The final timing assumption of GLZ is that an i.i.d. demand shock is realized after production occurs.[14]

Given $\{\widehat{\omega}_{ft}^{GLZ}\}$ for each industry, we estimate versions of (4) by OLS pooling across industries (separately for manufacturing and non-manufacturing). The simplest version specifies (4) as an AR(1) with industry i by year t fixed effects,

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

To account for the fact that $\{\widehat{\omega}_{ft}^{GLZ}\}$ is estimated in the first stage GLZ procedure, we bootstrap the entire two-stage procedure to compute the standard errors of the estimated parameters of (5). The bootstrap procedure samples firms f rather than individual f-t observations, so can be thought of as a way of clustering errors by firm.

The effect of techies on future productivity is identified from cross-section variation. In

[14]The analogous assumption in OP/LP/ACF is that realized productivity has two components, one observed only after all input decisions have been made.

order to compute the permanent level effect we divide β by $1-\lambda$.

The GNR methodology. GNR make no functional form restrictions, nor do they impose constant returns to scale. However, they do assume that all quantities, including materials inputs, are observed by the econometrician. GNR assume that ω_{ft} is a Markov process, which may be a controlled Markov, and that one or more of the inputs to production is static (that is, fully flexibly adjusted after productivity is realized). We explain in the online appendix O.3 how to include several flexible inputs.

Their estimator has two steps. First, by manipulating the first-order profit-maximizing conditions for the static input, they derive a relationship between the cost share of the flexible input in revenue and the output elasticity of that same flexible input. This relationship is estimated by non-linear least squares (NLLS). This is used to build the contribution of the flexible input to output and identify an error term that captures unexpected productivity shocks that materialize after decisions on demand for the static input occur. Both elements are used in the second step.

After subtracting the contribution of the flexible input from output, the second step uses GMM to identify the rest of the production function that does not rely on the flexible input jointly with the Markov process for productivity. The Markov assumption (4) is used to specify the necessary moment conditions and is also identified by the GMM estimator.

The baseline GNR estimator assumes that the data include physical output quantities. We use GNR's extension to the case in which only revenues are available to the researcher. Like GLZ, GNR's extension builds on Klette and Griliches (1996). However, since our sample spans only nine years (2011–2019), we are not able to precisely identify the demand curvature parameter ρ using the GNR methodology.[15] Therefore, we identify productivity scaled by the unknown demand curvature parameter $\rho \in (0,1)$: $\widehat{\rho\omega}_{ft}^{GNR}$. Once this is done, we pool across industries (separately for manufacturing and non-manufacturing) and estimate the

[15] In the GNR extension ρ is identified only from time variation; while the point estimates are in the $(0,1)$ range, with only nine years the standard errors are huge. GLZ identifies ρ using different moment conditions, which yield precise inference.

following AR(1) controlled Markov equation

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

where $\overline{\rho}$ is an average ρ_i across industries j. The key difference between (5) and (6) is that the parameter of interest β cannot be separately identified from $\overline{\rho}$. Consistent with $\rho \in (0, 1)$, in our estimates reported below we indeed find smaller coefficients in (6) compared to (5). As we do for the GLZ procedure, standard errors of the estimated parameters of (6) are computed using a bootstrap that samples firms.

Comparing GLZ to GNR. Our application of GNR is more general than GLZ because it does not make any functional form assumptions apart from constant elasticity demand, which is common to both estimators. In particular, GNR do not impose a CES production function nor constant returns to scale, while GLZ do. But this generality comes with two significant drawbacks given our data. First, the parameter vector of interest β in (6) is identified only up to a constant. Nevertheless, the signs and the relative magnitudes of the elements of $\beta\overline{\rho}$ are informative. Second, GNR assume that the researcher observes input quantities, but in our data they are not. To implement GNR we deflate expenditures on materials using an industry-specific materials input price series, as does most of the literature. GLZ show that this can bias estimates of productivity dispersion.

GNR has another important virtue compared to GLZ: it does not require labor to be a static input, which is appealing, given the labor market structure in France. France's labor market features both temporary and permanent employment contracts, and firing costs are high for both. In addition, we use a version of GNR to entertain the case in which both labor and materials are static inputs, as in GLZ. Thus, for each specification of (6), we estimate two versions of GNR: (1) labor is predetermined in period t and thus does not adjust to period t productivity shocks (like capital), and (2) labor is a static input and fully adjusts to period t productivity shocks (like materials inputs). The second version is closer to GLZ, while the first represents a distinct assumption about labor adjustment in the model.

Before moving to the results, we highlight that while the presentation of the controlled Markov in (5) and (6) is linear—other, richer specifications do not change our inferences in a qualitative way. For example, we report specifications in which we include a third order polynomial in lagged ω in Table O4 in the Appendix. We also experimented with non-linear specification of Techies, which lead to qualitatively similar inference. Results for interacting Techies with lagged productivity are presented in Table 6.

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). 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 three categories R&D, ICT, or Other techies.[16]

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) and that their prevalence and intensities vary by type. While Other techies have the highest wage bill shares on average in both manufacturing and non-manufacturing firms, R&D techies have

[16] In Harrigan et al. (2023), we also classify the techie occupations by their 2-digit categories, technical managers and engineers (PCS38) and technicians (PCS 47) as engineers may be more knowledgeable and skilled, and thus should matter more in the technology-enhancing and diffusion process. We show in the online appendix that both engineers and technicians positively affect productivity, although the engineers exhibit a greater effect than the technicians, both at the extensive and intensive margins.

Table 2: Descriptive statistics for estimation sample

	Percent with positive values	Mean conditional on positive values	Percentiles of techie wage bill shares on positive support, percent 10	25	50	75	90	IQR
Manufacturing								
Techies	71.8	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
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

higher wage bill shares in manufacturing and ICT techies have a higher average wage bill share in non-manufacturing firms.

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.[17]

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

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

labor input in quantities) is not defined for the knife-edge case of $\sigma = 1$ (i.e., a Cobb-Douglas production function). 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.[18]

We relegate the estimates of the "revenue production function" using the GNR methodology to online appendix O.4. 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.

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.[19] Formally, we estimate specifications of this general form:

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

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

[19]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.

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%.

where we also estimate specifications where we allow for multiple arguments for techies by their category. 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 O4, we report estimates of the controlled Markov process where we add $\omega^2_{f,t-1}$ and $\omega^3_{f,t-1}$. 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 4) and non-manufacturing industries (columns 5 to 8). Our analysis of non-manufacturing firms contrasts with most of the literature, which focuses on manufacturing.

Columns (1) and (5) 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 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 (6) 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 2 and 4 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

Table 4: Impact of techies on productivity – GLZ estimates

	Manufacturing				Non-Manufacturing			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$I\left(T_{ft-1} > 0\right)$	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)		
$I\left(T^{RD}_{ft-1} > 0\right)$			0.017***	0.011***			0.010*	-0.002
			(0.002)	(0.003)			(0.006)	(0.007)
$I\left(T^{ICT}_{ft-1} > 0\right)$			0.021***	0.014***			0.025***	0.015***
			(0.002)	(0.003)			(0.004)	(0.004)
$I\left(T^{OTH}_{ft-1} > 0\right)$			0.029***	0.011***			0.053***	0.018***
			(0.002)	(0.003)			(0.003)	(0.003)
T^{RD}_{ft-1}				0.069***				0.160*
				(0.023)				(0.088)
T^{ICT}_{ft-1}				0.101***				0.117***
				(0.036)				(0.021)
T^{OTH}_{ft-1}				0.113***				0.243***
				(0.010)				(0.015)
$I\left(x_{ft-1} > 0\right)$	0.009***	0.007***	0.002	0.003	0.008***	0.006**	0.006**	0.005**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
$\hat{\omega}_{ft-1}$	0.911***	0.913***	0.908***	0.911***	0.874***	0.875***	0.874***	0.876***
	(0.003)	(0.003)	(0.003)	(0.003)	(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

firms and 8 times larger in non-manufacturing.[20] These can be seen in Table 5, where we see that firms with the median intensity of techies are estimated to have 57.45% greater productivity in manufacturing, compared to 48.29% 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.

Columns (3), (4), (7) and (8) in Table 4 report estimates when techie workers are broken down by their detailed job descriptions. Both the presence and the intensity of R&D techies have a large impact on productivity in manufacturing. These findings corroborate the results

[20]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.

of Doraszelski and Jaumandreu (2013) for Spanish manufacturing firms.

However, techies' positive impact on productivity is not limited to R&D techie workers. In columns (3) and (7), 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.

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. 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)

	A. Impact Effects				B. Steady State Effects			
	Manufacturing		Non-Manufacturing		Manufacturing		Non-Manufacturing	
	$0-p50$	IQR	$0-p50$	IQR	$0-p50$	IQR	$0-p50$	IQR
Techies	4.03	2.38	5.05	3.77	57.45	31.00	48.29	34.50
R&D techies	1.46	0.49	0.20	0.88	17.72	5.66	1.63	7.35
ICT techies	1.60	0.26	1.99	1.09	19.59	2.99	17.20	9.17
Other techies	2.92	1.70	4.65	4.02	38.12	20.83	44.28	37.36

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), and (8) 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

Table 6: Impact of techies on productivity – General specification

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

Notes. Units are percent points.

the techie wage bill and lagged productivity, but the effects are not substantially different from the baseline results reported in Table 5. Table O5 in the online 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 and consider the role of managers. We then report results using the GNR estimator. In online appendix O.7 we report 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. Our methodology relies on the assumption 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 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 online appendix in Table O9. We see that the null hypothesis that the effect of techies on future productivity is zero is easily rejected. We 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, we 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), (7) and (8) 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.

We conclude 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)
$I\left(T_{ft-1}>0\right)$	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)		
$I\left(T_{ft-1}^{RD}>0\right)$			0.027***	0.012***			0.034***	0.023***
			(0.002)	(0.003)			(0.006)	(0.007)
$I\left(T_{ft-1}^{ICT}>0\right)$			0.024***	0.018***			0.055***	0.038***
			(0.002)	(0.003)			(0.003)	(0.004)
T_{ft-1}^{RD}				0.274***				0.296***
				(0.032)				(0.110)
T_{ft-1}^{ICT}				0.219***				0.188***
				(0.051)				(0.022)
$I\left(x_{ft-1}>0\right)$	0.000	0.002	-0.001	0.000	0.021***	0.022***	0.020***	0.022***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
$\hat{\omega}_{ft-1}$	0.915***	0.915***	0.914***	0.915***	0.878***	0.878***	0.878***	0.878***
	(0.002)	(0.002)	(0.002)	(0.002)	(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 $\{RD, ICT\}$ denote R&D, ICT, other techies, 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 their associations with innovative and productivity-enhancing activities (Section 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 (Table 4) for convenience, while columns (2) and (4) add lagged managerial labor to the controlled Markov equation. The results in Table 9 indicate that including lagged managers does not materially affect the estimated effects of lagged techies.

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

	Manufacturing		Non-Manufacturing	
	Baseline (1)	Managers (2)	Baseline (3)	Managers (4)
$I\left(T_{ft-1}>0\right)$	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)
$I\left(M_{ft-1}>0\right)$		0.002		0.027***
		(0.003)		(0.002)
M_{ft-1}		-0.063***		-0.021***
		(0.010)		(0.006)
$I\left(x_{ft-1}>0\right)$	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

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 when using the GNR estimator. One reason for doing this is a general robustness check. A second reason is that the GNR estimator allows us to relax the assumption that labor is a static input, which is an important consideration given that there are large firing costs in the French labor market. Table 10 reports the results when labor is assumed to be "static" (like materials, as assumed when implementing the GLZ estimator), and Table 11 reports the results for when labor is assumed to be "dynamic" (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).

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, 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. 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 techies is greater than that of R&D.

Some differences with Table 4 are apparent, however. For example, in Table 10 we do not

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)
$I(T_{ft-1} > 0)$	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)		
$I\left(T^{RD}_{ft-1} > 0\right)$			0.014***	0.012***			0.008***	0.007***
			(0.001)	(0.001)			(0.002)	(0.002)
$I\left(T^{ICT}_{ft-1} > 0\right)$			0.014***	0.012***			0.010***	0.006***
			(0.001)	(0.002)			(0.001)	(0.001)
$I\left(T^{OTH}_{ft-1} > 0\right)$			0.031***	0.026***			0.023***	0.014***
			(0.002)	(0.002)			(0.001)	(0.001)
T^{RD}_{ft-1}				0.019*				-0.016
				(0.011)				(0.025)
T^{ICT}_{ft-1}				0.017				0.036***
				(0.016)				(0.008)
T^{OTH}_{ft-1}				0.029***				0.057***
				(0.005)				(0.004)
$I(x_{ft-1} > 0)$	0.015***	0.014***	0.013***	0.013***	0.008***	0.007***	0.007***	0.006***
	(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***	0.932***	0.933***	0.933***	0.933***
	(0.005)	(0.005)	(0.005)	(0.005)	(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 $\hat{\omega}_{ft}$, log estimated productivity. $I(.)$ is the indicator function. T is the techie wage bill share, superscripts $\{RD, ICT, OTH\}$ denote R&D, ICT, other techies, 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

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. 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.

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)
$I\left(T_{ft-1}>0\right)$	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)
$I\left(x_{ft-1}>0\right)$	0.028***	0.027***	0.026***	0.027***	0.009***	0.009***	0.006***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
$\hat{\omega}_{ft-1}$	0.689***	0.687***	0.690***	0.687***	0.820***	0.820***	0.846***	0.845***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.007)	(0.007)	(0.008)	(0.008)
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 $\{RD, ICT, OTH\}$ denote R&D, ICT, other techies, 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

7 Conclusion and implications

Our paper has shown the key role of techies in raising firm-level productivity in both manufacturing and non-manufacturing firms in France from 2011 to 2019. A key 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 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 employment of techies as analogous to investment: 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 achieving this goal. Our findings about the key role of ICT and other techies suggest that educational, training and other policies that enhance the supply of techies will also have positive effects on growth. This suggests a trade-off between investing in R&D that may expand technological possibilities versus investing in workers' skills that may help firms adopt these new possibilities. We leave resolving this trade-off for future research.

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

Ackerberg, Daniel, Garth Frazer, Yao Luo, Yingjun Su et al., "Under-identification of structural models based on timing and information set assumptions," 2021. Working paper, https://ssrn.com/abstract=3717757 or http://dx.doi.org/10.2139/ssrn.3717757.

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

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, Luis Garicano, Raffaella Sadun, and John Van Reenen, "The distinct effects of information technology and communication technology on firm organization," *Management Science*, 2014, *60* (12), 2859–2885.

_ , Raffaella Sadun, and John Van Reenen, "Management as a Technology?," 2017.

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

Bresnahan, Timothy F, Erik Brynjolfsson, and Lorin M Hitt, "Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence," *The Quarterly Journal of Economics*, 2002, *117* (1), 339–376.

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

_ and Lorin M Hitt, "Beyond computation: Information technology, organizational transformation and business performance," *Journal of Economic perspectives*, 2000, *14* (4), 23–48.

_ and _ , "Computing productivity: Firm-level evidence," *Review of economics and statistics*, 2003, *85* (4), 793–808.

_ , 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.

Caves, Douglas W, Laurits R Christensen, and W Erwin Diewert, "The economic theory of index numbers and the measurement of input, output, and productivity," *Econometrica*, 1982, pp. 1393–1414.

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.

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.

_ , _ , and _ , "Techies and firm level productivity," Working Paper 31341, National Bureau of Economic Research June 2023.

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.

Appendix

Fact 1. Techies across industries. In addition to the information in the main text, we report some additional statistics here. In the DADS data 47% of manufacturing firms that employ R&D techies also have ICT techies; this proportion is similar in non-manufacturing firms, 44%. 96% of manufacturing firms with R&D techies also employ other techies; this proportion is slightly lower in non-manufacturing firms, 84%.

Fact 2. Techies have more STEM education and training than other occupations. We use data from the Training and Professional Qualification survey (TPQ survey: *Enquête formation et qualification professionnelle*) to assess whether techies have more STEM education and training than other occupations. Among the 26,861 individuals with valid data, 5.4% are Engineers (PCS 38) and 5.1% are Technicians (PCS 47), closely matching the shares in the DADS data.

Table A1 reports the results. Around 60% of techies have a STEM degree and/or additional training, with about 20% holding both. STEM degrees are more common among engineers (55%) than technicians (41%). In contrast, STEM education is rare in other occupations, with only 11% holding a STEM degree and fewer than 20% having any degree or training. These results confirm that techies have higher levels of STEM education and training than other occupations.

Table A1 also provides details on STEM degrees and training among large non-techie occupations. Fewer than 20% of upper managers have any STEM education, with even lower shares among middle managers and clerical workers. In contrast, over one-third of skilled industrial workers have some STEM education. However, these are mostly high school-level qualifications rather than university degrees. More than two-thirds of skilled industrial workers hold a professional baccalaureate (14%), a vocational school certificate (CAP, 29%), or a certificate of vocational proficiency (BEP, 15%).

Table A1: 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 .

Fact 3. Most R&D spending is on wages. The Annual Survey on the Means dedicated to Research and Development (R&D survey: *Enquête R&D Entreprises*) 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 firm. Conditional on reporting positive internal R&D, most R&D expenditures originate within the control of the firm. We show in Table A2 that expenditure on R&D is overwhelmingly spent within the firm, and the median firm spending nothing on external R&D.

Table A2: 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.

We use the R&D survey to analyze how much of the firm's R&D budget is spent on in-house, R&D labor costs. This is important because we cannot assume that all R&D costs in the firm are spent on labor. We show in Table A2 that R&D spending is mainly spending on wages, especially when R&D is done within the firm. In Table A3 we show that the external share of R&D spending is weakly correlated with overall R&D spending and strongly negatively correlated with the wage bill share of total R&D. We conclude that firms indirectly hire some R&D workers through external R&D spending, but not many: most R&D workers are employed by the firm paying for the R&D, and their wages make up the bulk of firm R&D spending.

Table A3: Correlations

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

Source: R&D survey.

Our main data analysis uses information on various types of techies from the DADS data to explain firm-level productivity. In Table A4, 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 A4: 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.

Fact 4. Techies are positively associated with the diffusion of ICT within firms. We use The Information and Communication Technology survey (ICT survey: *Enquête sur les technologies de l'information et de la communication et le commerce électronique – TIC entreprises*) to examine the link between techies and technology diffusion within firms, focusing on three key questions:

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?

Panel A of Table A5 shows that only 20% of firms provide ICT training. Firms with ICT workers are six times more likely (0.66/0.11) to offer such training. Panel B provides more detail on the exposure of different worker types to ICT training. Interestingly, 11% of firms without ICT workers still provide ICT training, suggesting external training.

Table A5: ICT Workers and ICT Training

	A. Offer ICT training?		B. Which workers get ICT training?			
	No	Yes	None	Only ICT	Only non-ICT	ICT & non-ICT
Employ ICT workers?						
No	0.89	0.11	0.89	0.00	0.11	0.00
Yes	0.34	0.66	0.34	0.18	0.12	0.35
Mean	0.80	0.20	0.80	0.03	0.11	0.06

Source: ICT survey.

We match the ICT survey data with the DADS sample and observe minimal discrepancies between the two datasets. Specifically, 10% of firms report having ICT workers in the DADS data, while 12% of firms report ICT workers in the ICT survey—a small difference. We further examine the relationship between having ICT workers in the survey and ICT tech workers (both and others) in the DADS data. As shown in Panels A and B of Table A6, the two measures are highly correlated Panel A 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). Panel B of Table A6 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 A6: ICT workers in the ICT survey and DADS dataset

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

Source: ICT survey.

We next ask if ICT techies are associated with training of workers in ICT. To answer this question, Table A7 repeats the analysis of Table A5 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 3.5 (= 0.49/0.14) times more likely to offer ICT training.

Table A7: ICT workers and ICT training

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

Source: Matched dataset.

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

Table A8 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 A7). Columns (1)-(3) show that firms with techies are significantly more likely to offer training. Firms with any techies are 6 percent more likely to offer training (Column 1). This effect is mainly driven by ICT techies, with a coefficient of 0.20, while R&D (0.06) and other techies (0.04) show smaller positive effects (Columns 2-3). In Columns (4)-(6), which include only firms with techies, the results show that a 10% higher techie expenditure increases the likelihood of offering ICT training by 5 percentage points, primarily driven by ICT techies.

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

Fact 5. Techies are positively associated with patenting and innovation. We describe the relationship between R&D spending, techies and patents, and innovation outcomes. The R&D survey reports whether firms introduced new or improved products, services, or processes and the number of patents filed during the year. As expected, 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 top four firm-year observations are around 2,000 patents. Responses to questions related to innovations are much less skewed, as seen in Table A9: only a quarter of firms say that they had no process or product innovations in the past year, while half had both.

Next, we analyze the relationship between patenting, R&D spending, and techies in two steps. First, we analyze firms' patenting and innovation activities using the R&D variables from the 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. The analysis is limited to firm-year observations with positive R&D expenditures. We use a negative binomial model for the number of patents filed and a linear probability model for innovation activities. The estimates have the interpretation of elasticities as the right-hand side variables are taken in logs. In both regressions we controls for firm size with the the non-techie wage bill–which turns out not to affect the results. All models include industry and year-fixed effects.

Table A8: 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.

Table A9: Innovation activity, share of firms

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

Source: R&D survey.

We report the results of the analysis of the R&D survey in Table A10. The results in columns (1) and (2) show a positive relationship between R&D spending and patents, with an elasticity of about 0.60, which remains stable when using the R&D wage bill. In column (3), breaking down R&D into wage and non-wage components confirms the positive link, highlighting the role of labor in R&D. In columns (4) to (12), R&D spending strongly correlates with innovation in both products and processes. The elasticity of the R&D techie wage bill to innovation is nearly five times that of non-wage R&D spending, emphasizing the key role of R&D workers in product innovation.

We now study the results in the matched sample in Tables A11 and A12. We include the firm's non-techie wage bill as a control for size, which turns out to be unimportant.

In Table A11, we report the results from the matched R&D and DADS datasets on the impact of techies on the number of patents. In column (1), we estimate the impact of techies and observe a striking similarity to the effect of total Research and Development (R&D) spending presented in Table A10. 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.

Our last statistical exercise in this section reports linear probability models for the three

Table A10: 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.

Table A11: Number of patents (results using the matched dataset)

			Manufacturing	Non-Manufacturing
Wage bill (log):	(1)	(2)	(3)	(4)
– 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.

innovation outcome indicator variables. The parameter estimates reported in Table A12 have the interpretation of semi-elasticities. Overall, Techies have a statistically significant positive relationship with the likelihood of innovation. This suggests that techies can lead to increased innovation in product development or process improvement.

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

Table A12: Innovation (results using the matched dataset)

Wage bill (log):	Innovation (1)	(2)	Manuf. (3)	Non-Manuf. (4)	Product Innovation (5)	(6)	Manuf. (7)	Non-Manuf. (8)	Process Innovation (9)	(10)	Manuf. (11)	Non-Manuf. (12)
– 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.

innovation, in both manufacturing and non-manufacturing industries.

Online Appendix

Techies and Firm Level Productivity
Harrigan, Reshef and Toubal

O.1 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 O1 reports our classification.

Table O1: 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)

[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.

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
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 O2 provides information on the list of diplomas and training that we group to identify individuals with education and training in science, technology, engineering, and math (STEM). .

Table O2: Mapping diplomas' specialties into STEM skills

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

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

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

O.2 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.

O.3 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 $\eta = 1/(1-\rho) > 1$, $\rho \in (0,1)$, *á la* Klette and Griliches (1996), as in GNR's Appendix O6-4 "Revenue Production Functions".

A firm that sets price P_{ft} sells quantity

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

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

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

Therefore, revenue is

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

Given an aggregate price index Π_t we have deflated revenues

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

The theory-consistent measure of B_t is given by

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

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.

O.3.1 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

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

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

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

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

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

The second step equation is

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

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

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

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

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

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

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

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.

O.3.2 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

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

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 :

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

O.3.3 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.

O.4 Production functions estimates

Table O3 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.

O.5 More lags of $\hat{\omega}_{ft}$

Table O3: 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 O4: 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)				
$I\left(T^{RD}_{ft-1}>0\right)$			0.015***	0.009***					0.013**	0.000		
			(0.002)	(0.002)					(0.006)	(0.007)		
$I\left(T^{ICT}_{ft-1}>0\right)$			0.018***	0.011***					0.025***	0.015***		
			(0.002)	(0.002)					(0.003)	(0.004)		
$I\left(T^{OTH}_{ft-1}>0\right)$			0.028***	0.009***					0.048***	0.012***		
			(0.002)	(0.003)					(0.003)	(0.003)		
T^{RD}_{ft-1}				0.071***						0.151		
				(0.023)						(0.092)		
T^{ICT}_{ft-1}				0.111***						0.118***		
				(0.037)						(0.022)		
T^{OTH}_{ft-1}				0.114***						0.251***		
				(0.010)						(0.015)		
$I\left(T^{38}_{ft-1}>0\right)$					0.028***	0.010***					0.046***	0.009***
					(0.002)	(0.003)					(0.003)	(0.003)
$I\left(T^{47}_{ft-1}>0\right)$					0.015***	0.005*					0.030***	0.019***
					(0.002)	(0.002)					(0.002)	(0.003)
T^{38}_{ft-1}						0.141***						0.271***
						(0.013)						(0.018)
T^{47}_{ft-1}						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

O.6 General Specification

Table O5 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 O5: 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

O.7 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 (O6) while the results using the GNR estimator are presented in Tables (O7) and (O8).

Table O6: Impact of techies on productivity – GLZ estimates (Adjusting for labor input quality)

	Manufacturing				Non-Manufacturing			
	(1)	(2)	(3)	(4)	(7)	(8)	(9)	(10)
$I(T_{ft-1}>0)$	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)		
$I\left(T^{RD}_{ft-1}>0\right)$			0.018***	0.014***			0.008	-0.003
			(0.002)	(0.003)			(0.006)	(0.007)
$I\left(T^{ICT}_{ft-1}>0\right)$			0.019***	0.014***			0.026***	0.017***
			(0.002)	(0.003)			(0.004)	(0.004)
$I\left(T^{OTH}_{ft-1}>0\right)$			0.035***	0.018***			0.060***	0.024***
			(0.002)	(0.003)			(0.003)	(0.003)
T^{RD}_{ft-1}				0.040				0.126
				(0.025)				(0.098)
T^{ICT}_{ft-1}				0.081**				0.103***
				(0.038)				(0.022)
T^{OTH}_{ft-1}				0.102***				0.242***
				(0.010)				(0.015)
$I(x_{ft-1}>0)$	0.008***	0.006**	0.001	0.002	0.008***	0.006**	0.006**	0.006**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
$\hat{\omega}_{ft-1}$	0.913***	0.916***	0.911***	0.914***	0.876***	0.877***	0.876***	0.878***
	(0.003)	(0.003)	(0.003)	(0.003)	(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 $\{RD, ICT, OTH\}$ denote R&D, ICT, other techies, 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 O7: Impact of techies on productivity – GNR estimates assuming labor to be static (adjusting for labor input quality)

	Manufacturing				Non-Manufacturing			
	(1)	(2)	(3)	(4)	(7)	(8)	(9)	(10)
$I\left(T_{ft-1} > 0\right)$	0.037***	0.029***			0.026***	0.016***		
	(0.001)	(0.001)			(0.001)	(0.001)		
T_{ft-1}		0.037***				0.056***		
		(0.003)				(0.003)		
$I\left(T^{RD}_{ft-1} > 0\right)$			0.013***	0.012***			0.005***	0.004**
			(0.001)	(0.001)			(0.002)	(0.002)
$I\left(T^{ICT}_{ft-1} > 0\right)$			0.012***	0.011***			0.011***	0.007***
			(0.001)	(0.001)			(0.001)	(0.001)
$I\left(T^{OTH}_{ft-1} > 0\right)$			0.031***	0.026***			0.025***	0.014***
			(0.001)	(0.001)			(0.001)	(0.001)
T^{RD}_{ft-1}				0.014				-0.016
				(0.009)				(0.025)
T^{ICT}_{ft-1}				0.016				0.037***
				(0.016)				(0.008)
T^{OTH}_{ft-1}				0.027***				0.063***
				(0.004)				(0.004)
$I\left(x_{ft-1} > 0\right)$	0.014***	0.013***	0.012***	0.012***	0.005***	0.004***	0.004***	0.004***
	(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***	0.934***	0.935***	0.934***	0.935***
	(0.002)	(0.002)	(0.002)	(0.002)	(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 $\hat{\rho\omega}_{ft}$, log estimated productivity. $I\left(.\right)$ is the indicator function. T is the techie wage bill share, superscripts $\{RD, ICT, OTH\}$ denote R&D, ICT, other techies, engineers and technicians respectively, x is the value of firm exports. Industry-year fixed effects included in all columns. Bootstrap standard errors clustered by firm in parentheses. *** denotes p-value $\leq$ 0.01, ** p-value $\leq$ 0.05, * p-value $\leq$ 0.10

Table O8: Impact of techies on productivity – GNR estimates assuming labor to be predetermined (adjusting for labor input quality)

	Manufacturing				Non-Manufacturing			
	(1)	(2)	(3)	(4)	(7)	(8)	(9)	(10)
$I(T_{ft-1} > 0)$	0.027***	0.019***			0.012***	0.010***		
	(0.002)	(0.002)			(0.001)	(0.001)		
T_{ft-1}		0.041***				0.015***		
		(0.006)				(0.004)		
$I\left(T^{RD}_{ft-1} > 0\right)$			0.003**	-0.000			0.002**	0.004**
			(0.001)	(0.002)			(0.001)	(0.002)
$I\left(T^{ICT}_{ft-1} > 0\right)$			0.007***	0.003*			-0.000	-0.002*
			(0.002)	(0.002)			(0.001)	(0.001)
$I\left(T^{OTH}_{ft-1} > 0\right)$			0.024***	0.017***			0.010***	0.008***
			(0.002)	(0.002)			(0.001)	(0.001)
T^{RD}_{ft-1}				0.034***				-0.020
				(0.013)				(0.016)
T^{ICT}_{ft-1}				0.057***				0.017**
				(0.021)				(0.007)
T^{OTH}_{ft-1}				0.043***				0.011***
				(0.007)				(0.003)
$I(x_{ft-1} > 0)$	0.026***	0.026***	0.025***	0.026***	0.007***	0.007***	0.004***	0.005***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
$\hat{\omega}_{ft-1}$	0.680***	0.678***	0.681***	0.678***	0.839***	0.839***	0.843***	0.842***
	(0.021)	(0.021)	(0.020)	(0.021)	(0.009)	(0.009)	(0.008)	(0.008)
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 $\{RD, ICT, OTH\}$ denote R&D, ICT, other techies, engineers and technicians respectively, x is the value of firm exports. Industry-year fixed effects included in all columns. Bootstrap standard errors clustered by firm in parentheses. *** denotes p-value $\leq$ 0.01, ** p-value $\leq$ 0.05, * p-value $\leq$ 0.10

O.8 Sensitivity

Table O9: 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