The Minimum Wage in Firms' Organizations: Productivity Implications*

Nicholas Lawson† Claire Lelarge‡ Grigorios Spanos§

Abstract

Taking advantage of a unique empirical setting in which different minimum wages coexisted in France during 2003-06, we document that higher minimum wages are associated with smaller firms, flatter hierarchies, more training, and higher revenue productivity per worker. We construct and calibrate a general equilibrium model of optimal hierarchies that rationalizes these facts via firms' optimization of costly labor and knowledge. These two channels generate large endogenous labor productivity gains which strongly contribute to mitigating the aggregate output loss. In relative terms, this loss is amplified in economies with better communication technologies and attenuated with better information or problem-solving technologies.

Keywords: Minimum Wage, Firms' Organizations, New Technologies (ICT, AI), Productivity
JEL Classification: D21, D22, D24, J23, J30

*This work was partially funded by a subsidy allocated by France Strategie (5%), the TANED Idex chair (Université Paris-Saclay, 10%) and the TECHNOFIRMS ANR grant (85%). We would like to thank (without implicating) Luis Garicano, our discussant at the CEPR Paris 2023 Symposium, Timothy Van Zandt, our discussant at the BSE 2024 Summer Forum, Eliza Forsythe, our discussant at the 2024 NBER SI, as well as Ghazala Azmat, Emma Duchini, Miguel Espinosa, Emeric Henry, John Van Reenen, Maria Guadalupe, Florian Mayneris, Jean-Marc Robin, Alexandra Roulet, Raffaella Sadun, Chad Syverson, members of the French low-wage commission (chaired by Gilbert Cette), and participants in the June 2023 Centenary Symposium (Princeton Industrial Relations Section), the HEC 2023 "Firm organization and capabilities" workshop, CEPR Paris 2023 Symposium, BSE 2024 Summer Forum, ETEPP 2024, SED 2024, NBER SI 2024, and seminars at the Universities of Guelph, Dalhousie, Paris-Saclay and at the OECD(ECO).

†Dalhousie University. E-mail: nicholas.lawson@dal.ca. Most of the work on this paper was completed while Prof. Lawson was at UQAM.

‡Paris-Saclay University, RITM, CREST and CEPR. E-mail: claire.lelarge@universite-paris-saclay.fr

§GSEM. E-mail: grigorios.spanos@unige.ch

1 Introduction

For decades, labor economists have been studying the effects of the minimum wage on the economy. To date, however, little is known about how minimum wages impact the overall internal organization of labor within firms, even though they unambiguously alter the relative prices of factors (labor) and hence the relative cost of different work organization strategies. These in turn determine the roles assigned to workers with different skills, the structure of wages and, ultimately, the productivity of firms (Garicano, 2000; Caliendo and Rossi-Hansberg, 2012). In this paper, we propose a unified framework to examine these questions and investigate how the ongoing wave of communication and AI-based technologies might affect the relevant trade-offs.

Our objectives are threefold. First, we take advantage of a unique empirical setting in recent French economic history, where different minimum wage schedules coexisted between 2003 and 2006. Combined with extremely detailed matched worker-level and firm-level data, this setting allows us to investigate the empirical relationships between the tightness of the minimum wage constraint, firms' organizational strategies and productivity. Second, we propose a structural model of optimal hierarchies which delivers qualitative as well as quantitative predictions about how firms' internal organizations respond to increases in the minimum wage, allowing us to examine the implications of the model for workers' skills, wages and firms' productivity. We carefully calibrate the model to salient moments of the French data and verify that it delivers predictions that are compatible with our richer set of estimation results. Finally, we rely on this model to perform simulations and show that firms' endogenous organizational strategies and their implications for aggregate productivity are key to mitigating the impact of (moderate) minimum wage constraints. Importantly, we also consider different technological scenarios to determine whether large technological shifts that are likely to affect firms' organizational decisions, such as the current wave of new AI-based technologies, as well as the massive investment in communication technologies during the recent Covid-19 pandemic, are likely to confirm or invalidate these findings.

Our empirical analysis first focuses on a specific period of the French economy, which was characterized by the coexistence of several legal minimum wages, each applying to different sets of firms. This situation emerged as a consequence of the 1998 and 2000 "Aubry" laws (named after the Minister of Labor at the time), which mandated a reduction of the legal workweek from 39 to 35 hours from January 19$^{\text{th}}$, 2001 in firms with more than 20 employees, and on January 1$^{\text{st}}$, 2002 for all other firms. Different regimes of GMR ("Garantie Mensuelle de Rémunération", Aeberhardt et al., 2016) were created, tied to the timing of each firm's agreement to implement the workweek reduction, in order to preserve workers' purchasing power. Our period of interest is the subsequent period between 2003 and 2006, during which the government implemented a process of reconvergence of the various GMRs towards a single national minimum wage. Over this period, legal working hours had stabilized to 35 hours in a large sample of companies, which were however – and this is more interesting for us – differentially exposed to increases in the minimum wage (GMR) applicable to them. This setting opens the door to powerful double- and triple-difference strategies, whereby our mechanisms of interest are identified from firms' differential exposure to minimum wage evolutions, while addressing their endogenous sorting across GMRs. In order to address the potential endogeneity of some of the components of labor costs, we also propose an instrumental variable strategy, wherein their non-manipulable component serves as the instrumental variable.

The French data allow us to accurately capture the evolution of French production hierarchies (Caliendo et al., 2015), their worker-level training decisions as well as their performance in terms of sales and productivity. We obtain the following results. First, the increase in labor costs is transmitted to all but the highest level of the hierarchy of firms that are subject to increases in their legal minimum wage. Second, relative to other firms, they significantly decrease their size in terms of hours worked by flattening their overall hierarchical organization, and by contracting the lower levels of their internal hierarchical organizations. The firm-level estimated elasticity of total hours worked to the cost of labor at the minimum wage level is -0.15.[1] In addition despite the rise in the relative cost of labor compared to other factors of production, firms do not significantly increase their capital intensity, nor do they appear to be outsourcing more of their production. Conversely, their workers enroll more often in training programs, suggesting that the skill content of jobs increases. Lastly, these adjustments are associated with significant increases in "revenue" productivity (Foster et al., 2008; Hsieh and Klenow, 2009).

In a second step, we construct a quantitative model of firms' organization to assess whether the organizational responses observed in the data are likely to qualitatively and quantitatively rationalize the observed firm-level productivity increases, and to investigate further general equilibrium implications of minimum wages. Our framework builds on the knowledge-based management hierarchy framework initially introduced by Garicano (2000) and extended by Caliendo and Rossi-Hansberg (2012) to an economy with heterogeneous firms.[2] In this framework, production requires labor to be combined with knowledge. The optimal firm organization has a pyramidal structure composed of one layer of production workers, and one or many layers of managers with increasing levels of knowledge (Garicano, 2000). The benefit of this structure is that it prevents knowledgeable managers from dealing with the most frequent problems and allows them to devote their time to less common problems, which they have a comparative advantage in addressing (Garicano, 2000; Garicano and Zandt, 2013; Garicano and Hubbard, 2007). The optimal number of layers to use in production is endogenously determined by firms. Their decisions are based on the costs and benefits of introducing an additional layer of management to the firm (Caliendo and Rossi-Hansberg, 2012): from the perspective of the firm, adding a new layer of management is similar to occurring an additional fixed cost, because the firm has to hire an additional knowledgeable manager. However, provided that the firm is producing at a sufficiently large scale, an additional layer reduces the marginal costs of production, because workers become more efficient at solving problems with the help of the new manager. Hence, adding a layer of management is only profitable to firms producing at a sufficiently large scale, ie. to those benefiting from favorable demand conditions.

We introduce to this analytical framework the feature that the minimum wage constrains the optimal decisions of firms: an increase in the minimum wage increases the wage floor and the costs of hiring low-skilled production workers, especially when training costs are large relative to managerial costs. This in turn distorts the way firms organize their production process in terms of hierarchical layers and knowledge across layers, which further affects both firms' productivity and the structure of wages within firms. Our theoretical set-up thus generate a rich set of counterfactual predictions. In particular, one of the largest direct responses of firms to higher minimum wages is to raise the skill level of their production workers at the bottom of hierarchies, which means increasing their level of training. As a result, managers become less useful, so firms tend to select flatter organizations, saving on the "fixed costs" (associated with hierarchical levels) mentioned above and mitigating the cost associated with the minimum wage constraint. In addition, firm sizes decrease and the industrial fabric is more fragmented. Finally, theory-based indicators allow a detailed quantitative analysis of the implications of minimum wages on aggregate productivity (Caliendo et al., 2020).

[1] This result does not imply that the overall impact of the minimum wage on aggregate employment is negative, since our regressions do not allow us to accurately measure the impact on firms' entries and exits. Our theoretical model allows us to fill this gap.

[2] The origins of this literature go back decades, to an analysis of span of control in a multi-layer firm in Rosen (1982), and even further to papers on team production in Marschak and Radner (1972) and a basic model of span of control in Lydall (1959). See Garicano and Rossi-Hansberg (2015) for a comprehensive review.

We calibrate the resulting model to salient moments of our data and show that the model is fully able to rationalize, both qualitatively and quantitatively, our regression results. To that end, we solve for the equilibrium for different values of the minimum wage, in particular: (1) the baseline value (corresponding to the initial calibration to the French data); (2) a no minimum wage scenario, which appears to numerically coincide with the case where the minimum wage is decreased by 4% as compared with the baseline scenario; (3) a scenario where the minimum wage is 4% higher than the baseline, which corresponds to the real increase experienced by GMR2 between 2003 and 2006 (a "medium" case); (4) a scenario where the minimum wage is 8% higher than the baseline, which corresponds to the real increase experienced by GMR1 between 2003 and 2006; and lastly (5 and 6) two scenarios where the minimum wage is 16% and 24% higher than the baseline, which corresponds to two and three times the real increase experienced by GMR1 between 2003 and 2006, respectively. The latter case corresponds roughly to the shock induced in the US by the Fair Labor Standards Act in 1966 (Bailey et al., 2021). We obtain that output losses are strongly mitigated by (revenue) productivity increases caused by endogenous organizational responses of firms. Market share reallocations between firms are significant, but the pure selection channel (Melitz, 2003) only plays a minor role. Workers benefit from large skill increases and pocket the associated productivity gains in the form of higher wages across all hierarchical layers. Our framework therefore implies that minimum wages, when set at moderate levels, are likely to induce firms to invest more in workers' human capital and create higher-quality jobs, as in Acemoglu (2001) but via an entirely different channel. In addition, the model allows for an in-depth inspection of the detailed organizational channels through which the minimum wage affects the organization of production and overall, aggregate productivity. These adjustments are made through extensive organizational redesign: most firms get flatter, with negative average impact on their size. Our simulations show that these organizational margins of adjustment make it possible to divide the impact of minimum wage constraints by a factor of four relative to a counterfactual economy where only skills, but not firms' hierarchical organizations, could adjust.

Importantly, our set-up also contains a rich description of the technological environments of economies, by introducing as many as four different parameters capturing different dimensions of technologies that are directly relevant to firms' organizational choices: first, technologies augmenting homogeneously the productivity of all workers, second, communication technologies, third, information technologies, and fourth, technologies affecting the distribution of the difficulty level of production problems. These four dimensions allow us to document the likely impact of two on-going waves of technologies, namely, the implications of the huge spike of investment in communication technologies which occurred during the Covid-19 pandemic (Gibbs et al., 2023; Barrero et al., 2021; Bloom et al., 2022), and the implications of the AI-based wave of technologies, which are likely to resemble information or problem-solving technologies (Mullainathan and Obermeyer, 2021; Kleinberg et al., 2017; Noy and Zhang, 2023). We find that the relative cost of minimum wage constraints is amplified in economies benefiting from more efficient communication technologies. The intuition for this result is the following: more efficient communication technologies tend to increase the overall efficiency of verticalized organizations, as in Garicano and Rossi-Hansberg (2006). As organizational flattening is the most effective mitigation strategy in the face of minimum wages, we obtain unsurprisingly that the latter prevent economies from fully benefiting from the gains generated by efficient communication technologies, amplifying the cost of minimum wage constraints in terms of aggregate output. Importantly, our analysis thus contributes to clarifying why the empirical literature obtains mixed results about the productivity impact of remote work and the intense use of communication technologies which accompanies it: rather positive in Barrero et al. (2021), Bloom et al. (2022) or Angelici and Profeta (2024) and rather negative in Morikawa (2023) or Gibbs et al. (2023). Indeed, our framework shows that the impact might be heterogeneous across firms depending on their internal organization, as in Caliendo et al. (2015), and that the overall impact is likely to be modulated by labor market institutions, such as minimum wages.

On the other hand, our results are more reassuring regarding the likely impact of AI-based technologies. Assuming that the latter are likely to take the form of improvements in information or problem-solving technologies, we obtain that they are likely to, first, reduce the cost of strategies involving increases in workers' skills, and second, expand the set of problems that bottom, production workers are able to solve autonomously. These two mechanisms tend to mitigate the aggregate output cost associated with moderate minimum wage constraints, but they become ineffective when technological shocks or minimum wage constraints become large. Overall, our results imply that the current waves of technologies are likely to require a thorough re-examination of cost-benefit analyses of labor market institutions and regulations.

Our paper is related to several strands of the literature. As previously mentioned, it contributes directly to the research program documenting how corporate organizations affect economies (Garicano and Rossi-Hansberg, 2015). We do so in three important dimensions. First, we apply the toolbox to a new application, in labor economics, by investigating how firms' organizational strategies mediate the impact of a major labor market institution: the minimum wage. The latter directly enters firms' cost minimization programs which determine jointly firms' labor demand and optimal organizational choices. We show that firms' endogenous organizational adjustments are extremely powerful at dampening the impact of this constraint and that it directly translates into higher firm-level indices of productivity. Secondly, we take this mainly theoretical toolbox and turn it into an operational quantitative framework, which we also calibrate carefully to static moments of the French data. This is a critical step to provide a credible quantification of the organizational mechanisms at play as well as their firm-level and aggregate implications. We show that the endogenous response of firms' organizations to the minimum wage constraint is quantitatively capable of rationalizing why empirical estimates of the impact of the minimum wage on aggregate employment or output are often low.[3] Indeed, our regression analyses show that in the French data, this empirical result goes hand in hand with additional adjustments in terms of the number of hierarchical layers and in productivity when minimum wages increase. These regression results accurately match the model's predictions, both qualitatively and quantitatively, lending credibility to the economic relevance of the organizational channel. Finally, in a third contribution, we draw on our quantitative model to provide simulations and assess the potential impact of the current waves of technologies.

In addition, our paper relates to the vast literature documenting the different dimensions of the impact of the minimum wage. In addition to the papers mentioned above, those focusing on firm-level responses in terms of profit and productivity are particularly relevant to our research question. Draca et al. (2011) document that the introduction of the UK national minimum wage in 1999 led to wage increases as well as decreases in firms' profitability. Rizov et al. (2016) and Riley and Rosazza Bondibene (2017) show that it also led on average to increases in productivity, while both Hau et al. (2020) and Mayneris et al. (2018) document that managerial practices might be an important driver of these productivity gains. Our theoretical and quantitative framework is able to fully reconcile all these patterns, since it predicts opposite impacts on average profits and productivity, respectively, for the range of increases in the minimum wage observed in the

[3]See for example the classic paper by Card and Krueger, 1994 or more recently Bailey et al., 2021, Cengiz et al., 2019 or Dustmann et al., 2021), although methodologies and precise quantifications remain at the center of an intense debate (eg. Neumark et al., 2014 vs. Card et al., 1993). The most common explanations suggest that firms' monopsony power reduces wages below their competitive level, so that minimum wage increases only reduce their rents (Manning, 2011; Azar et al., 2023; Wiltshire et al., 2023), or that price pass-through offsets higher labor costs (Ashenfelter and Jurajda, 2022; Bello and Pesaresi, 2024; Sorkin, 2015; Harasztosi and Lindner, 2019; Renkin et al., 2022; Aaronson et al., 2008). However, these candidate mechanisms are unable to rationalize the observed patterns of flattening corporate organizations. We leave the joint analysis of the potential interaction between market power, on both the input and output sides, and optimal corporate organizational strategies for future research (Hau et al., 2020), since the lack of firm-level price data to be inserted into our regression and calibration exercises (for the period of interest) unfortunately complicates the feasibility of such extensions.

data. Furthermore, it shows that selection and composition effects only become significant when minimum wage constraints are extremely stringent. More generally, our paper relates to the extremely large literature studying the effects of minimum wages on employment and wage distributions.[4]

The remainder of the paper is structured as follows. Section 2 describes the French labor market institutions, with a focus on the minimum wage. Section 3 then presents the empirical setting, our main identification strategies and our regression results describing how French firms adjusted to the large increases in their minimum wage which occurred between 2003 and 2006. In Section 4, we present our theoretical framework and Section 5 details our calibration strategy as well as our simulation results.

2 Institutional Background

2.1 A Brief History of the French Minimum Wage

The minimum wage in France dates back to the end of World War II. It was first introduced into law in 1950 along with post-war reforms that restored workers' freedom to negotiate their wages through collective bargaining agreements (loi n° 50-205 of February 11, 1950).[5] Today, the minimum hourly wage in France is the Salaire Minimum Interprofessionnel de Croissance (SMIC). It applies uniformly across metropolitan France and covers workers in the French labor market, with the notable exception of civil servants.[6] The latter represent around 20% of the workforce,[7] but their career paths are strictly regulated and take place in administrative environments that are largely disconnected from the for-profit sector, such that we exclude them from our analysis. The minimum wage is updated annually according to the French Labor Code, which during our period of analysis states that: (i) changes to the minimum wage are to be indexed to changes in the national consumer price index, (ii) the annual increase in the purchasing power of the minimum wage cannot be less than half of the increase in the purchasing power of the average hourly wages of blue collar workers, and (iii) that the minimum wage can be raised to a higher level than that resulting from the application of conditions (i) and (ii). As a result, over the period 1998 to 2006, the SMIC was updated every July 1st using the following formula:

$$\frac{SMIC_{t+1}}{SMIC_t} = \frac{CPI_{t+1}}{CPI_t} + \frac{1}{2} \cdot \left(\frac{SHBO_{t+1}}{SHBO_t} - \frac{CPI_{t+1}}{CPI_t} \right) + \Delta_{t/t+1} GOV \tag{1}$$

where $SHBO_t$ denotes the average nominal hourly wage of blue collar workers (Salaire Horaire de Base Ouvrier) in period t, CPI_t denotes the national consumer price index in period t, and $\Delta_{t/t+1}GOV$ denotes discretionary increases to the minimum wage that the government can implement between periods t and $t+1$.

[4] Beyond those aforementioned, many extremely detailed studies could be cited, among which: Lee (1999); DiNardo et al. (1996), Derenoncourt and Montialoux (2020) and Forsythe (2023) for the case of the US or Dickens et al. (1999); Dickens and Manning (2004); Machin et al. (2003); Stewart (2004); Metcalf (2008); Stewart (2012) for the case of the UK.

[5] Concerned that wages would be too low in sectors with weak unions, the government introduced the Salaire Minimum Interprofessionnel Garanti (SMIG), which determined the minimum hourly compensation in metropolitan France. Initially, the SMIG varied across metropolitan France: it was based on the standard budget of a single person living in Paris, and adjusted for workers living in the different regions of France. Moreover, following France's high inflation in 1951 and 1952, the SMIG was indexed to the consumer price index, so as to maintain minimum wage workers' standard of living (loi n° 52-834 of July 18, 1952). It was further reformed over the years, yet it continued to increase at a slower pace than average hourly wages, increasing inequality between minimum wage and other workers. As a result, it was replaced in 1970 by the Salaire Minimum Interprofessionnel de Croissance (SMIC), which remains in place until today (loi n° 50-205 of January 2, 1970).

[6] There are a few groups of workers, such as workers below the age of 18 and interns, that are able to earn wages below the SMIC. These workers are outside of the scope of our analysis.

[7] https://www.insee.fr/fr/statistiques/5392034?sommaire=5392045

2.2 Workweek Reform: the Rise and Fall of Multiple Minimum Wages

Our analysis focuses on a specific period of the French economy, which was characterized by the coexistence of several legal minimum wages, each applying to different sets of firms. This situation emerged as a consequence of the 1998 and 2000 "Aubry" laws (loi n° 98-461 of June 13, 1998 and loi n° 2000-37 of January 19, 2000), which mandated a reduction of the legal workweek from 39 to 35 hours from January 19th, 2001 in firms with more than 20 employees, and on January 1st, 2002 for all other firms.[8] To ease our exposition, we separate this event into two sub-periods: the emergence of multiple legal minimum wages in the French economy and their convergence to a single minimum wage.

The first sub-period ranges between 1998 and 2002 and corresponds to the actual implementation of the workweek reduction within firms. Most interesting for our question of interest, while the Aubry laws mandated a reduction in the French legal workweek, they also stipulated that firms entering into such agreements had to preserve the monthly income of workers paid at the minimum wage.[9] This implied that a total of five distinct schedules of minimum wages were created, each associated with the actual date of the workweek reduction agreement: between 15/06/1998 and 30/06/1999 (GMR1), between 01/07/1999 and 30/06/2000 (GMR2), between 01/07/2000 and 30/06/2001 (GMR3), between 01/07/2001 and 30/06/2002 (GMR4), or later than 01/07/2002 (GMR5). The acronym GMR stands for "Garantie Mensuelle de Rémunération" and reflects the fact that each GMR was designed to maintain the income of workers who were paid at the minimum wage, despite the reduction in work time.[10] Furthermore, each GMR was based on the SMIC at the initial date of the workweek reduction agreement and once in place, each GMR was also updated on every July 1st, according to the following rule:

$$\frac{GMR^k_{t+1}}{GMR^k_t} = \frac{CPI_{t+1}}{CPI_t} + \frac{1}{2}\left(\frac{SMBO_{t+1}}{SMBO_t} - \frac{CPI_{t+1}}{CPI_t}\right) + \Delta_{t/t+1}GOV^k. \tag{2}$$

where GMR^k_t denotes the minimum wages of workers employed in firms in GMR k, $SMBO_t$ denotes the average nominal *monthly* wage of blue collar workers (Salaire Mensuel de Base Ouvrier) in period t, and $\Delta_{t/t+1}GOV^k$ denotes discretionary increases to the minimum wage in GMR^k_t that the government can implement between periods t and $t+1$. The essential difference with Equation 1 is that GMRs were indexed to blue-collar workers' *monthly* wages, not their *hourly* wages.

The rules governing how the different GMRs and the SMIC were to be updated imply that these different minimum wages had different trajectories over the period 1998 to 2002. This is illustrated in Panel (A) of Figure 1, which plots the different minimum wages across time. From this figure, it is evident that until 2002, each GMR increased at a much slower pace than the SMIC. This is a direct result of the decrease in the legal workweek, which implied that the aggregate *monthly* income of blue collar workers (SMBO) evolved at a much slower pace than their *hourly* income (SHBO). Panel (A) of Figure 1 also illustrates that since they were based on the same rule, and since discretionary increases were the same until the year 2002, the different GMRs were moving in parallel over this sub-period and thus did not converge.

The second sub-period ranges between 2003 and 2006 and corresponds to the convergence of the different GMRs and the SMIC. Indeed, the Aubry law of 2000 stipulated two things: (i) a policy to converge the different minimum wages had to be designed and approved before the end of the year 2002, and (ii) that all

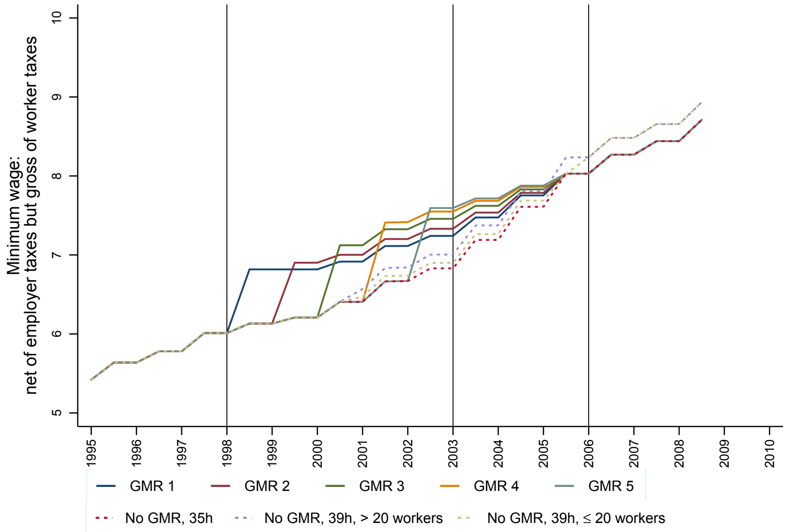
[8] The media named this law after the (left-wing) minister who led the reform. Refer for example to the following online report published by the French Senate for an overview of the Aubry and Fillon laws: https://www.senat.fr/rap/np05_35/np05_3511.html.

[9] Note that this implies that the hourly wage (net of employer payroll taxes) increased by 39/35 -1 ≈ 11.4% upon the adoption of the 35 hour workweek.

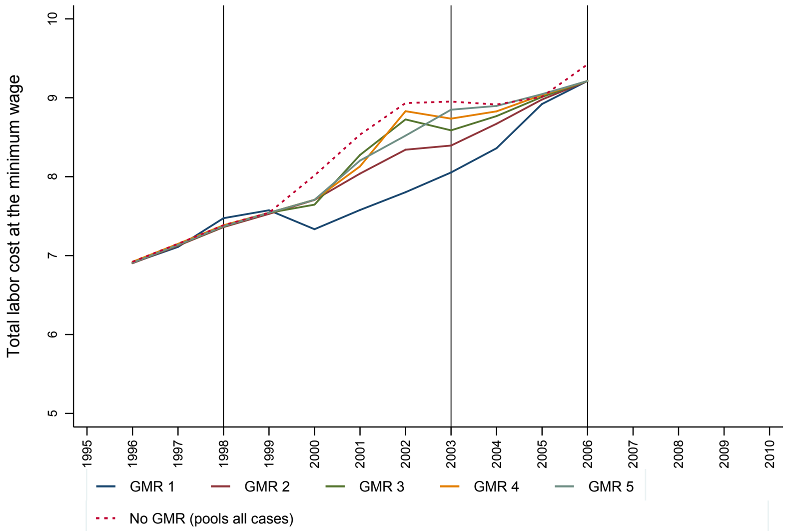
[10] GMRs are thus defined in "net" terms, i.e. net of social payroll taxes - whereas the relevant concept for firms is the total cost of labor (incorporating such taxes). We come back to this important aspect below.

Figure 1: Official Minimum Wage and GMRs

(A) Net of Employer Payroll Taxes



(B) Gross of Employer Payroll Taxes - Total Labor Costs



(C) Average Changes in (ln) Total Labor Costs at GMR, across GMR, 2003 to 2006

	GMR1	GMR2	GMR3	GMR4	GMR5
In nominal terms	0.135	0.093	0.070	0.053	0.040
In real terms	0.079	0.037	0.014	-0.003	-0.016
(Standard deviation)	(0.027)	(0.018)	(0.014)	(0.013)	(0.000)

Notes: This figure depicts the evolution of the official hourly minimum wage and the five temporary GMRs over our analysis period. The five GMRs were created between July 1998 and July 2002, and their re-convergence to the official minimum wage took place between July 2003 and July 2005. Panel (A) presents the evolution of all GMRs and the minimum wage (SMIC) in terms of nominal wages net of social charges at employer level, but gross of social charges at worker level: the latter correspond to the concept of directly regulated wages (see https://www.ipp.eu/baremes-ipp/marche-du-travail/salaire_minimum/gmr/). Panel (B) supplements this information by constructing the average *total* labor cost of a job paid at the corresponding GMR (or minimum wage). The necessary data (DADS) are only available on an annual basis, so temporal evolutions are slightly less precise than in panel (A). Figure (B) and the first row of Table (C) are expressed in nominal terms, while the last row adjusts for an economy-wide inflation rate of 5.6% from 2003 to 2006 (Insee online data, accessed April 2024).

GMR schedules and the baseline minimum wage should converge no later than July 1st, 2005. Yet, how this would happen was not detailed in the initial 1998 law and was left for the ruling government to decide. In fact, even the timing specified in the law was not completely respected: the policy to converge the different minimum wages was only approved on January 17th, 2003, with the passing of the "Fillon" law (loi n° 2003-47 of January 17, 2003), named after the new right-wing government's minister.[11] While there were many options available, the new government decided that the highest minimum wage at that time, GMR5, would only evolve according to the consumer price index, so as to preserve the purchasing power of the corresponding workers. All other GMRs and the SMIC followed the same path and additionally received annual discretionary boosts in order to progressively converge to GMR5 on July 2005. In sum, between July 2003 and July 2005, these boosts amounted to 40 euros per year for the SMIC, 18 euros per year for GMR1, 13.5 euros for GMR2, 7 euros for GMR3, and 2.25 euros for GMR4.

It is clear from the above developments that the 2003 to 2006 period is particularly interesting, as it allows us to compare the behavior of firms that were exposed to differential and arguably unanticipated changes in their minimum wage schedule, once we take into account their potential self-selection into different GMR classes over the 1998-2002 period (Aeberhardt et al., 2016). There is however one last remaining important difficulty to address, namely that the concept of labor cost that is relevant for our research question is not well captured by the net wages that are plotted in Panel (A) of Figure 1. This is particularly important in our setting, as firms received different types of job-level subsidies to reduce their legal workweek to 35 hours. To be precise, while the Aubry laws mandated a reduction in the French legal workweek, they also put into place financial incentives to encourage firms to sign collective bargaining agreements containing a reduction in working time (RWT) to 35 hours. These incentives were in the form of job-level lump-sum or payroll tax subsidies that were designed to soften the associated increase in labor costs, brought about by the decrease in the workweek.[12]

The total cost of employing a worker is unfortunately not directly available in the data. A key task of our project is precisely to simulate it at the worker level from the labor and social security laws. All details are given in Section 3.2. The most important point is that, from 2003 onwards, the scale determining the gap between total labor costs and net wages was, on the one hand, largely unanticipated and, on the other, stable over time, despite the fact that the Aubry subsidies initiated during the 1998-2002 period were, themselves, heterogeneous between firms (notably according to size, even within GMR groups). In Panel (B) of Figure 1, we plot our measure of the average total hourly cost of employing a minimum wage worker, across firms operating under the different GMRs. Panel (B) shows that the trajectories of the labor costs associated with minimum-wage workers were significantly different from the net wage trajectories of Panel (A), as subsidies tended to partially offset the cost to companies of reducing the working week to 35 hours. However, despite

[11] The right-wing government in power was the second Raffarin government, named after its Prime Minister Jean-Pierre Raffarin. It entered into office on June 17th, 2002, and replaced the first Raffarin government, elected on May 6th, 2002. Moreover, while the policy was approved on January 17th, 2003, it was announced earlier (see, for example, the following statements made by François Fillon, at the Meeting of the National Commission for Collective Bargaining in Paris on September 6th, 2002, available at: https://www.vie-publique.fr/discours/132763-declaration-de-m-francois-fillon-ministre-des-affaires-sociales-du-tr).

[12] Initially, these incentives were set up to encourage firms to decrease their workweek before the mandatory dates, while they later applied to all firms who implemented RWT agreements. The first set of incentives were job level lump-sum subsidies that were made available by Aubry 1 to firms that reduced their workweek before January 1, 2000, and to firms with at most 20 employees that decreased their workweek before January 1, 2002 (loi n° 98-461 of June 13, 1998). These firms were also required to avoid lay-offs and even increase by at least 6% their workforce that would be affected by the RWT agreement (although this aspect was difficult to enforce in the long term). Secondly, from 2000 onwards, the previous subsidies were replaced with a reduction in firms' required social security contributions for all jobs paying less than 1.8 times the minimum wage (loi n° 2000-37 of January 19, 2000 and loi n° 2000-1257 of December 23, 2000). This second set of incentives are commonly referred to as the Aubry 2 tax cuts. Finally, in 2003 the Aubry 2 tax cuts were replaced by the "Fillon" tax cuts (loi n° 2003-47 of January 17, 2003), which applied to all jobs paying less than 1.6 times the minimum wage. It is important to note that these subsidy rules were identical for all GMRs between 2003 and 2006. The only source of heterogeneity in total labor costs lies in the differential dynamics of the GMRs themselves.

this attenuation, Panel (B) illustrates that the heterogeneity of costs between GMRs remains significant in 2002, and that these costs began to converge in 2003 as a consequence of the Fillon law and reached full convergence in 2006.[13]

The remainder of our empirical analysis thus focuses on the 2003 to 2006 sub-period and examines the outcomes of firms that had previously signed collective bargaining agreements reducing their legal workweek to 35 hours, to exploit plausibly exogenous variation in the minimum wage. We restrict ourselves to this period for the following reasons. First, since the firms in our sample signed collective bargaining agreements to reduce their workweek in previous years, it is reasonable to expect that the effects of the RWT agreements should have taken place by the year 2003. In other words, it is unlikely that our estimates will be biased from confounding factors from the reduction in the legal workweek. Second, even though firms were aware that the different minimum wages would converge by the year 2005, the uncertainty surrounding the implementation of this policy implied that they could not anticipate how this would take place (Aeberhardt et al., 2016). Put differently, while firms knew that the minimum wage would adjust over the period, they could not predict its future level in 2006. Taken together, these points imply that in our analysis, annual changes in the GMRs can be treated as at least partly exogenous - an aspect that we discuss in full details and address in Section 3.1, together with the other endogeneity issues that we need to address. Panel (C) of Figure 1 provides the precise numbers that are involved. In nominal terms, total minimum wage labor costs rose by 4% to 13.5% between 2003 and 2006, depending on the GMR. Adjusting for inflation over this period of time (which was estimated by the French statistical to add up to 5.6%), we obtain that labor costs for GMR4 and GMR5 actually decreased by 0.3% and 1.6% in real terms, respectively, while they increased by 1.4% to 7.9% across GMR1 to 3.[14]

3 Empirical Evidence

3.1 Estimation Strategy

In this section, we draw on the institutional framework described above to shed light on the empirical relationship between firms' organizations and the magnitude of the minimum wage "constraint" they face. As previously discussed, we concentrate on a period in the French economy where (i) firms faced different costs of employing a minimum wage worker, and (ii) a policy to converge these costs was enacted in the year 2003. To estimate causal effects, a standard approach is to use a difference-in-differences empirical strategy, comparing changes in the outcomes of a treated and a control group of firms, before and after a change in the minimum wage (see, for example, Card and Krueger 1994, Draca et al. 2011, and Drucker et al. 2021). Such an empirical strategy cannot be directly implemented in our context, since the companies included in our estimation sample all have a RWT agreement, so as to ensure that they all apply the same working hours (see above). They are therefore all exposed to changes in their respective minimum wages over the period of study. However, a slightly adjusted specification can be estimated, allowing us to estimate our parameters of interest based on the sample heterogeneity of the costs of employing a minimum-wage worker. To fix ideas,

[13] Notice in contrast that the total labor costs of firms operating under the SMIC (but none of the GMR) do not converge with the GMRs in the year 2006 (and remains heterogeneous in this class of firms), because the Fillon law allows these firms to maintain their workweek at 39 hours. To be precise, the Fillon law allowed firms to remain *de facto* at 39 hours by introducing the possibility to generalize overtime, although with a wage premium. The rate of this premium was set at 25% for the first 8 hours (per week), and at 50% for the hours exceeding this threshold. For smaller firms (below 20 workers), the rate was reduced to 10% until December 2005. Our data do not allow us to observe which firms in this class maintained a 39 hour norm, and which switched to 35 hours. To avoid comparing the outcomes of firms with different legal workweeks, we therefore restrict our analysis to firms that signed a formal RWT agreement before 2002 (while conversely, firms who were subject to the SMIC are all excluded from our sample).

[14] Our theoretical model in Section 4 is specified in real terms such that it is this last set of numbers which is relevant to quantify the shock experienced by French firms during this period of time.

let Y_{jt} denote an outcome of firm j operating in industry $s(j)$ and in area $a(j)$, in year t. To conduct our analysis, we estimate the following difference equation, over the 2003 to 2006 period:

$$\Delta_{03/06} Y_{jt} = \beta \Delta_{03/06} \ln TC_{jt}^{MW} + \theta_{s(j)} + \psi_{a(j)} + \Delta_{03/06} \varepsilon_{jt}, \quad (3)$$

where $\Delta_{03/06}$ denotes changes between the years 2003 and 2006, TC_{jt}^{MW} denotes the total cost of employing a minimum wage worker in firm j at time t, and $\theta_{s(j)}$ and $\psi_{a(j)}$ denote industry and area fixed effects. TC_{jt}^{MW} varies across firms (see Section 2, specifically Panels (B) and (C) of Figure 1), because firms were subject to different minimum wages and associated total labor costs. Furthermore, because it is in differences, the specification in equation (3) accounts for any time-invariant firm characteristics. Lastly, it controls for differential trends in firm industry and area. It is a generalization of the standard difference-in-differences setting to the extent that the main parameter of interest in equation (3), β, is identified from firms that are *differentially* exposed to increases in the minimum wage rather than exposed *vs.* non exposed. It identifies the arguably causal effect changes in the total cost of employing a minimum wage worker have on the outcome of firms under the identifying assumption that GMRs were subject to the same residual trends during the period of 2003-2006, after controlling for aggregate industry and area-level trends.

However, since the total cost of employing a minimum wage worker includes several potentially endogenous subsidies received by firms, there remains the concern that our estimates of β may be biased. More specifically, since several of the subsidies provided by the Aubry laws varied with the size of firms, it may be the case that firms manipulated their size during the pre-regression period (1998 to 2002), thus being able to endogenously lower their costs of employing a minimum wage worker in the subsequent years. To deal with this concern, we instrument for $\Delta_{03/06} \ln TC_{jt}^{MW}$, firms' total labor cost at the minimum wage, using the change in the minimum wage actually paid out to workers, ie. the minimum wage concept of panel (A) of Figure 1. This variable is net of the suspicious component, namely the potentially endogenous, firm-level varying, payroll taxes. The rationale of this strategy is that first, the instrumental variable is likely to have power, since workers' "net" wages are the largest component determining their total cost to firms. Second, as discussed in Section 2, the sub-period surrounding the convergence of the different GMRs led to unanticipated and arguably exogenous changes in the different "net" minimum wages in the French economy. The results of our IV specifications can be found in Appendix D.

Lastly, another concern with the specification in equation (3) is that firms in the different GMRs may have different trends over the period 2003-2006. To deal with this concern, we report in Appendix D results from specifications that exploit variation across firms with different *shares* of workers affected by an increase in the minimum wage (Draca et al., 2011; Riley and Rosazza Bondibene, 2017; Harasztosi and Lindner, 2019). This triple-difference strategy is motivated by the following reasoning: to the extent that firms self-selected into the different GMRs, it is reasonable to expect that absent an increase in the minimum wage, firms under the same GMR would exhibit similar trends over the period 2003-2006. However, because of differences in their wage structure, firms under the same GMR but featuring *ex ante* (in 2002) different shares of workers paid below the minimum wage of 2006 were differentially exposed to increases in the minimum wage.[15]

[15] Appendix D thus reports results from the following specification:

$$\begin{aligned} \Delta_{03/06} Y_{jt} &= \beta \Delta_{03/06} \ln TC_{jt}^{MW} \times SH_j^{MW} + \alpha \Delta_{03/06} \ln TC_{jt}^{MW} + \eta SH_j^{MW} \\ &\quad + \theta_{s(j)} + \psi_{a(j)} + \Delta_{03/06} \varepsilon_{jt}, \end{aligned} \quad (4)$$

where SH_j^{MW} measures the share of low-wage workers in firm j in 2002, ie. earning in 2002 less than the 2006 minimum wage. Again, since $\Delta_{03/06} \ln TC_{jt}^{MW}$ may be endogenous in Equation 4, it is also possible to implement the same instrumental variable strategy as previously. The corresponding results are also reported in Appendix D.

3.2 Data and Measurement

Data sources. Our dataset is constructed from administrative files covering the entire population of French companies and their employees; it is an augmented version of the files used in Caliendo et al. (2015).[16] Our main statistical source is the Annual Social Data Declarations (DADS) in their most detailed version, which is at the job ("poste") level. These files are available over our entire period of interest and contain information about each salaried worker's occupation (Profession and Socio-Professional Category), hours of work, and wage (net of so-called employer level payroll taxes).

A first auxiliary database constructed by the services of the Ministry of Employment and Solidarity at the time of the Aubry laws and entitled "Allègement des cotisations sociales" (Reduction in Social Security Contributions) contains information on the date of the working time reduction agreements (RWT), which determine the GMR to which firms are subject.[17] It also tracks the subsidies that firms received under the Aubry 1 and 2 schemes, which have a strong impact on the evolution of the cost of labor of these firms during our main period of interest (2003 to 2006).

A second auxiliary database, the "2483 Declarations", contains the legal declaration of vocational training efforts (to the Ministry of Labor) by companies with more than 10 employees. They correspond to the "train or pay" system introduced by the law of July 16, 1971, which requires companies with more than 10 employees to devote at least 1.5% of their payroll to vocational training. If their training expenditure falls below this legal threshold, they pay the balance to public training bodies (OPAC) or pay a tax. Subsequently, in September 2003, trade unions and company representatives signed an agreement creating an "individual right to training" (DIF), which was fully incorporated into the 2004 law on "professional training and social dialogue". Under the terms of this agreement, employees on open-ended contracts or, under certain conditions, fixed-term contracts, are entitled to 20 (optional) hours of training per year, formally in a field of their choice - but subject to their employer's agreement, implying that firms have in practice a margin to direct them in their best interest.[18] This institutional setting implies that training is mainly considered by firms as a component of labor cost, and from the perspective of workers, as a non-pecuniary component of their overall compensation, that is not systematically activated. The associated statistical files that were produced by the Ministry of Labor during our period of interest are unfortunately not exhaustive, but cover just over 30% of the firms of interest.[19] They contain extremely useful indicators about the number of beneficiaries of training programs (and the associated hours), overall and by occupation.

Lastly, these files are matched with the exhaustive tax files (Bénéfices industriels et commerciaux - régime normal BIC-RN, Fichier de comptabilité unifié SUSE -FICUS) containing the balance sheet data and profit and losses accounts of firms. The latter allow us to compute various indicators of firms' sizes beyond employ-

[16] In particular, we have been able to integrate granular information on labor costs (at the job level) and worker training (i.e. skills enhancement) from administrative files that have recently been made available to researchers.

[17] The original agreements were negotiated and reported at the plant level. However, there is little variation across plants in multi-establishment firms, such that we aggregate this information at the firm level by allocating all plants in a given firm to the earliest agreement that was signed. In unreported robustness checks, we allocate all plants to the agreement of the largest employer establishment of the considered firm, with no impact on any of our results. Note that this aggregation, which seems harmless overall, simplifies our analyses since the accounting data are available only at the firm level (but not at the establishment level), while the breakdown of employment across establishments in the DADS is noisy for our main years of interest (the plants responsible for paying wages may not be the actual employers).

[18] They can also accumulate these 20 hours over a period of less than six years. In return, companies can request that this training take place outside working hours.
It is important to acknowledge that this reform of vocational training affected all French firms during our estimation period. Our estimation strategy captures differences in training strategies across firms that are exposed to differential increases of their minimum wages (GMR) and thus naturally controls for economy-wide shocks, which are captured by the $\theta_{s(j)}$ terms in Equation 3.

[19] Unfortunately for us, the sample corresponding to the year 2003 is of small size. To circumvent this problem, we estimate the firm-level evolution in hours of training between 2003 and 2006 from different time periods, depending on their availability: 2003/06 if available, and otherwise from either: 2003/07, 2003/05, 2002/06, 2002/07, 2002/05, 2004/06 or 2004/07 (and adjusting for the different number of years in the cases where it is required).

ment: sales and value-added. They also allow us to inspect factor shares and various proxies of productivity.

Firms' internal organizations: hierarchical layers. To document the impact of minimum wage constraints on firms' internal human resource strategies, we borrow from Caliendo and Rossi-Hansberg (2012) and Caliendo et al. (2015) and rely on their theoretical and empirical concepts of hierarchical layers (which will also be at the heart of our theoretical set-up in Section 4). In theory, a layer corresponds to a group of employees, with similar characteristics (especially in terms of knowledge i.e. education or experience), and performing similar tasks in the organization. These layers are hierarchical in the sense that higher layers of management are constituted of increasingly knowledgeable workers, who have the authority to guide the work of their subordinates (typically, by solving operational problems the latter are unable to solve autonomously).

The French classification of occupations (PCS-ESE) provides indications about the hierarchical layer of each worker. As in Caliendo et al. (2015), we map the following five occupational categories, identified below by their 1-digit code, to four distinct hierarchical layers:

2. Salaried company managers, including craftsmen and traders who are managers and owners of their companies;
3. Senior staff or top management positions, including chief financial officers, chief technical officers, heads of human resources, and logistics and purchasing managers;
4. Employees at the supervisor level, including quality control technicians, technical, accounting, and sales supervisors;
5. Qualified and non-qualified clerical employees: secretaries, human resources or accounting employees, telephone operators, and sales employees;
6. Blue-collar qualified and non-qualified workers: welders, assemblers, machine operators, and maintenance workers.

Each class is considered as a distinct hierarchical layer, except PCS5 and PCS6, which are merged into a unified layer of the least skilled production workers.[20] Lastly, firms are classified as 0, 1, 2 or 3-layer firms depending on the number of (managerial) layers they exhibit:[21] firms featuring only one of the above categories are not considered as hierarchical and are thus classified as 0- (managerial) layer firms, as in Caliendo and Rossi-Hansberg (2012). Secondly, firms featuring two layers, be they contiguous in the above classification or not, are classified as 1-(managerial) layer firms. Similarly, firms exhibiting three or four layers will be classified as 2-layer or 3-layer firms, respectively.

Measurement of labor costs. As previously discussed in Section 2, a limitation of the DADS data is that they do not contain the total labor costs associated with each worker. To be precise, Caliendo

[20] Caliendo et al. (2015) document in detail that the distribution of wages of workers in these two classes is extremely similar, indicating similar levels of knowledge (ie. similar training decisions on the part of firms). This finding is confirmed over our period of analysis.

[21] Our labeling convention is similar to Caliendo and Rossi-Hansberg (2012), but different from Caliendo et al. (2015) (if we interpret their papers correctly!), but without impact on any aspect of the analysis: a $L-$ (managerial) layer firm in our paper and in Caliendo and Rossi-Hansberg (2012) simply corresponds to a $L+1-$ (production or managerial) layer firm in Caliendo et al. (2015). In addition, the mapping of occupations to management layers that is achieved with the French classification of occupations is very similar to the mapping achieved with the Portuguese classification (Caliendo et al., 2020), but seems to differ from what is measurable in the US data. For example, Forsythe (2023) is only able to separately identify "management", "supervisors" within 4 different functional categories ("professional", "clerical and sales", "production" and "service"), and production workers in those categories. This only captures at most 2 layers of management. Moreover, in Forsythe (2023), management layers are identified within establishments, but not firms. However, despite the attenuation bias generated by this limitation of US data, several empirical results reported in the paper seem to support our conclusions that internal hierarchies play an important role in the mechanism by which minimum wages affect the economy: in particular, the author argues that "a key channel by which minimum wages spread through an establishment is due to wage spillovers within supervisory relationships".

et al. (2015) for example only rely on data based on the concept of "salaires bruts", which are gross of so-called "worker-level" payroll taxes, but (unfortunately) net of so-called "employer-level" payroll taxes.[22] This concept is approximate but sufficient to their purpose, since they only aim to capture workers' ability or knowledge. In our context, however, this is inadequate because "salaires bruts" do not capture the total labor costs associated with each worker included in the firm's cost minimization problem on which we focus. In particular, the missing amount of "employer-level" payroll taxes depends on firm-level (industry, worktime reduction agreement, etc) as well as worker-level (mainly, wage level) characteristics.

Fortunately however, the DADS files contain enough information for us to apply the schemes described in the different laws, as the Human Resources department of any firm would do. We are thus able to accurately simulate the labor costs associated with each job within each firm (Cottet et al., 2012). In particular, we implement in full details the subsidies associated with the Aubry 1, 2 and Fillon laws, which all have a first-order impact on labor costs over our period of interest. We also implement the earlier "ARTT-loi de Robien" (1996) and the revised version of the "réduction bas salaires" introduced in 1996 by the Juppé government, which still impacted labor costs until 2005 and 2003, respectively.[23]

It is important to acknowledge that our data contains direct firm-level information on the actual use of subsidies for the Aubry 1 and 2 programs, which fortunately make up the bulk of payroll tax reductions over the period of interest. For all other programs however, our data do not allow us to observe whether companies actually applied for subsidies to which they were legally entitled - or, on the contrary, whether they cheated by applying for and receiving subsidies without being entitled to them. However, given the amounts involved (and the penalties for misconduct), we expect our simulations of labor costs to be good approximations of actual labor costs at firm (and employee within firms) levels.[24]

3.3 Descriptive Statistics

Our regression sample corresponds to all French firms that: (1) were continuously active between 2000, the year preceding the change in legal working hours, and 2006, when the convergence of GMR is achieved, and (2) that had already signed an agreement and thus had implemented the legal reduction in working time by the end of the year 2002. As explained above, this double criterion insures that all sample firms are included in a well-defined GMR and subject to the compulsory 35-hour workweek. We also discard firms having fewer than 5 jobs and firms reporting negative value added. Appendix B contains comprehensive descriptive statistics about this population of firms, while we focus in this section on our main dimension of investigation: the interaction between labor costs and firms' organizational strategies.

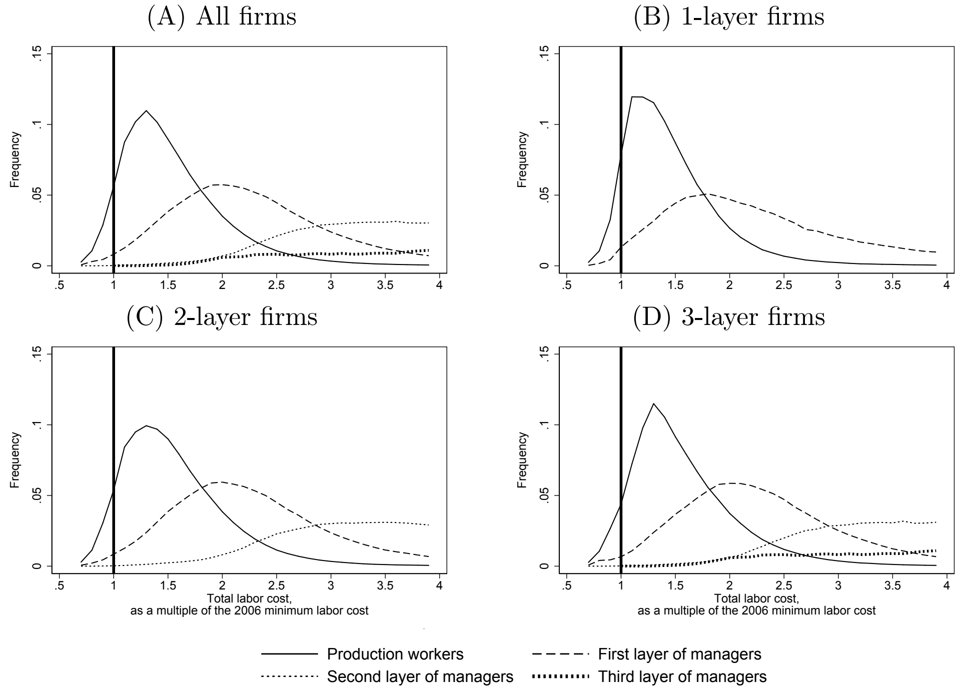
A key dimension to document is the extent to which the empirical methodology proposed by Caliendo et al. (2015) to describe "production hierarchies" remains valid in our data. The answer to this question is positive, which comes as no surprise since we rely on the same data sources, over (almost) the same time period, although Caliendo et al. (2015) focus on manufacturing industries only. Appendix C.1 replicate some of their main results in our sample: in our data as in theirs, firms form well-behaved hierarchies to the extent that the number of hours or jobs in the bottom, production layer (layer 0) is usually larger than in layer 1, which is itself larger than in layer 2, etc. Furthermore, the hierarchy also holds in terms of wages or labor

[22] Of course, these purely legal wordings do not reflect the actual economic incidence of the corresponding taxes!

[23] Refer for example to the following online report published by the French Senate for an overview of the different schemes in the pre-Aubry laws period: https://www.senat.fr/rap/l99-030/l99-03093.html. The details of the Aubry Laws are available online on an official website maintained by the French government: https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000398162. Sébastien Roux of the French statistical institute Insee kindly shared a version of the programs developed by Cottet et al. (2012), which contained a first approximation of the payroll tax subsidies during our period of interest.

[24] Notice that the measurement error that is introduced by our procedure is likely to introduce an endogeneity problem (since this variable corresponds to our indicator of main interest). Fortunately, the instrumental variable strategy presented in Section 3.1 addresses this concern.

Figure 2: Distributions of Hourly Labor Costs Across Layers in Organizations



Notes: The graphs correspond to histograms with bins of width 0.1 obtained for the distribution of hourly labor costs measured in 2002, but normalized by the total labor cost at the minimum wage in 2006. This norm is materialized by a vertical solid line in each panel.

costs, the latter concept being original in our paper. We document this feature in full detail in Figure 2, which reports histograms (by bins of width 0.1) of total labor costs as of 2002, normalized by the total labor cost at the minimum wage in 2006.[25] Our analysis is performed for our entire sample in Panel (A), as well as separately for each category of firms featuring at least one managerial layer in Panels (B) to (D). In each case, we find that the support of the distributions of labor costs overlap across layers, but their mode is sequentially shifted to the right as one climbs hierarchical layers. In more technical terms, the distribution of wages at a given hierarchical layer has strong first order stochastic dominance over distributions of wages at lower layers: in the vast majority of firms, almost all managers earn more than production workers, and those at the top of the hierarchy earn more than those at intermediate layers. As a consequence, virtually all workers who are hit by the minimum wage constraint over our period of study, ie. those earning in 2002 less than the minimum wage in 2006,[26] belong to the bottom layer (Layer 0).

3.4 Organizational and Productivity Responses to Minimum Wages: Regression Results

We now turn to the detailed depiction of how firms' organizations adjusted to *increases* in the minimum wage during the 2003 to 2006 period. The results reported in this section correspond to simple OLS estimation of the long difference specification in equation 3, while Appendix D checks the robustness of the results by replicating all tables with the instrumental variable or triple difference specifications described in Section 3.1. We show

[25] Non-parametrically estimated densities appear to be less stable and more difficult to show on the same support for labor costs in the second and third layers of management, but results are available upon request.

[26] These workers correspond to the mass located to the left of the thick vertical line.

there that the main results, in particular those relating to firms' internal organization and productivity, are preserved, both qualitatively and quantitatively.

Table 1: Impact of Minimum Wage Increases on Wages Across Firms' Organizations

	(1) All layers	(2) Layer 0	(3) Layer 1	(4) Layer 2	(5) Layer 3
	(A) Δ ln total hourly labor cost, all firms				
Mean, nominal:	0.095	0.110	0.097	0.073	0.106
Mean, deflated:	0.039	0.054	0.041	0.017	0.050
Δ ln hourly cost at GMR	0.352*** (0.021)	0.286*** (0.024)	0.196*** (0.046)	0.405*** (0.060)	0.221 (0.150)
# observations	55,344	55,344	49,479	33,202	8,484
R^2	0.045	0.036	0.024	0.035	0.103
Indicator:	(B) Δ ln total hourly labor cost, 1-layer firms				
Mean, nominal:	0.097	0.112	0.109		
Mean, deflated:	0.041	0.056	0.053		
Δ ln hourly cost at GMR	0.299*** (0.047)	0.221*** (0.056)	0.001 (0.130)		
# observations	16,195	16,195	16,195		
R^2	0.080	0.075	0.060		
Indicator:	(C) Δ ln total hourly labor cost, 2-layer firms				
Mean, nominal:	0.093	0.108	0.093	0.073	
Mean, deflated:	0.037	0.052	0.037	0.017	
Δ ln hourly cost at GMR	0.374*** (0.030)	0.312*** (0.034)	0.306*** (0.055)	0.402*** (0.076)	
# observations	24,612	24,612	24,612	24,612	
R^2	0.065	0.053	0.043	0.043	
Indicator:	(D) Δ ln total hourly labor cost, 3-layer firms				
Mean, nominal:	0.033	0.018	0.005	0.006	0.008
Mean, deflated:	-0.023	-0.038	-0.051	-0.050	-0.048
Δ ln hourly cost at GMR	0.440*** (0.048)	0.362*** (0.053)	0.335*** (0.069)	0.350*** (0.094)	0.221 (0.150)
# observations	8,484	8,484	8,484	8,484	8,484
R^2	0.126	0.112	0.101	0.101	0.103

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the 1-digit sector-commuting zone level in parentheses. Regression results from equation (3), estimated by OLS. All specifications contain 4-digit industry fixed effects and commuting zone fixed effects. In Panel (A), the composition of the estimation sample evolves across columns: in the first two columns, all sample firms are included. In columns (3) to (5), only firms exhibiting at least 2 (Layers 0 and 1), 3 or 4 layers respectively, both in 2003 and 2006, are included.

In Table 1, we investigate how the increase in the cost of labor at the level of the legally constrained GMR affected labor costs across layers of firms' hierarchical organizations. We estimate this first set of specifications in the pooled sample in Panel (A), where the reported elasticities thus correspond to averages across all types of hierarchical organizations, and by type of hierarchical organizations in Panels (B) to (D), where the elasticities are thus specific to each population of firms. Column (1) of Panel (A) first reports the population-wide average impact across all layers: we find a positive and highly significant elasticity of 0.352. Since Panel (C) of Figure 1 showed that changes in real labor costs at the GMR between 2003 and 2006 ranged between -1.6% (GMR5) and 7.9% (GMR1), our estimates thus imply that firms in GMR1 to 3 experienced on average (and all else equal) increases in total labor costs reaching up to 2.8% in real terms, while firms in GMR4 and 5 experienced a small relative decrease in their overall labor costs, by up to 0.6%.[27]

[27]At this stage, the quantifications neglect the general equilibrium impact on aggregate labor supply – which will be fully

Columns (2) to (5) propose a breakdown of this pattern by hierarchical level. We find that the impact is strongly positive up to the second managerial stratum (layer 2), but insignificant at the highest managerial layer. In the pooled sample, the estimated values of cost elasticities decrease from one layer to the next, with the exception of the second management layer. Panels (B) to (D) of Table 1 show however that this is due to a composition effect. Indeed, replicating the analysis by organization type, we obtain, firstly, that elasticities are monotonically decreasing from one layer to the next (or that the difference between layers is not statistically significant), but secondly, that all elasticities tend to be higher in more complex organizations. This implies that in Panel (A), elasticities in the upper layers (columns (3) to (5)) correspond to averages computed on sub-samples of companies with higher cost elasticities overall than the full population considered in the first two columns, thus producing the non-linear pattern in the aggregate.

Table 2 further describes how firms' organizations responded to increases in the minimum wage over our period of interest. Panel (A) starts by investigating the response in terms of the size of the different hierarchical layers, both in the pooled sample and by hierarchical form. Our estimates show that a firm subject to a 1% increase in its minimum wage reduced its workforce (as measured in hours) by 0.15% compared to a firm not subject to such an increase. Again, applying this estimate to the shock experienced by firms between 2003 and 2006 implies that firms in GMR4 and 5, which experienced a decrease in real labor costs at the GMR level, grew slightly, by up to 0.2%, while those in GMR1 to 3 contracted, by up to 1.2%. Overall, this result implies that a higher minimum wage (applying globally to all firms in the economy) leads to a structure of the productive fabric consisting of smaller firms. Columns (2) to (5) show that the firm-level contraction mainly occurs at the bottom layer (Layer 0), while the size of upper layers remained relatively stable.[28] This finding is confirmed in Panels (A1) to (A2), which shows furthermore that most of the adjustments take place in the largest firms, ie those featuring more than two managerial layers.

In addition to adjusting the relative sizes of their hierarchical layers, firms also responded to increases in minimum wages by decreasing their number of hierarchical layers. Panel (B) of Table 2 investigates this second organizational channel in greater detail. Column (1) first shows that firms tended to adopt significantly less complex organizations in order to absorb GMR increases, as compared with virtually unaffected or less affected firms. Our estimates imply that a 1% increase in the minimum wage is, all else equal, associated with a decrease in the number of hierarchical layers by $0.2 \times 0.01 \approx 0.002$, on average. The 2003 to 2006 shock therefore led to a flattening of firms in GMR1 to 3, by up to 0.016 hierarchical layers. Columns (2) to (7) show that this pattern is mainly driven by a lower probability of increasing the number of layers, as compared with otherwise similar firms, and in particular to a lower probability of adopting the most complex forms of organization, i.e. those reaching (the maximum possible value of) three management layers.

Table 3 focuses on the third type of organizational strategy we are able to track in our data, relating to training across layers. A specificity of our setting is that the eligibility criteria for some of the subsidies discussed in Sections 2 and 3.1 included a "no redundancy" condition, implying that it was impossible for them to rely on the extensive margin, ie. to fire workers and rely on new hires to adjust their skill mix. Rather, the only possible strategy for them was to rely on internal training, which we are able to capture well in our data. Their only limitation is their partial coverage of our population of interest, which implies that the estimation sample size is divided by around three in Table 3, and even by eight in the case of 1-layer

considered and discussed in Section 4. We actually show there that the price adjustments are limited for the range of increases in the minimum wages that correspond to GMR1 to 5, such that partial and general equilibrium results almost coincide numerically over this range.

[28] As a consequence, the proportion of managers increases, as in Forsythe (2023). This happens despite the fact that the number of layers decreases, as documented in Panel (B) of Table 2. Our theoretical model of Section 4 is able to rationalize these patterns, in particular in regimes where it becomes optimal to select firm organizations which do not exhaust the managerial time constraint.

Table 2: Impact of Minimum Wage Increases on Firms' Organizations

(A) Impact on the Relative Size of Hierarchical Layers

	(1) Total	(2) Layer 0	(3) Layer 1	(4) Layer 2	(5) Layer 3
Indicator:	(A0) Δ ln hours, all firms				
Mean:	0.014	0.016	-0.007	0.041	-0.043
Δ ln hourly cost at GMR	-0.154***	-0.395***	-0.0449	-0.116	0.104
	(0.0502)	(0.0850)	(0.115)	(0.137)	(0.277)
# observations	55,344	55,344	49,479	33,202	8,484
R^2	0.051	0.032	0.028	0.035	0.099
Indicator:	(A1) Δ ln hours, 1-layer firms				
Mean:	0.006	0.017	-0.019		
Δ ln hourly cost at GMR	-0.113	-0.031	-0.107		
	(0.102)	(0.186)	(0.264)		
# observations	16,195	16,195	16,195		
R^2	0.090	0.062	0.059		
Indicator:	(A2) Δ ln hours, 2-layer firms				
Mean:	0.018	0.017	-0.007	0.026	
Δ ln hourly cost at GMR	-0.242***	-0.499***	-0.037	-0.245	
	(0.074)	(0.127)	(0.156)	(0.171)	
# observations	24,612	24,612	24,612	24,612	
R^2	0.073	0.051	0.049	0.042	
Indicator:	(A3) Δ ln hours, 3-layer firms				
Mean:	0.030	0.019	0.016	0.086	-0.043
Δ ln hourly cost at GMR	-0.212*	-0.593***	-0.149	0.074	0.104
	(0.113)	(0.205)	(0.214)	(0.236)	(0.277)
# observations	8,484	8,484	8,484	8,484	8,484
R^2	0.125	0.102	0.109	0.112	0.099

(B) Impact on the Number of Hierarchical Layers

Indicator:	(1) Δ number hierarchical layers	(2) Probability of increase	(3) Probability of decrease	(4) Probability of stability	(5) Δ Proba 3 layers	(6) Δ Proba at least 2 layers	(7) Δ Proba at least 1 layer
Mean :	-0.036	0.165	0.193	0.642	-0.015	-0.013	-0.008
Δ ln hourly cost at GMR	-0.205**	-0.307***	-0.0761	0.383***	-0.211***	0.015	-0.009
	(0.103)	(0.057)	(0.064)	(0.077)	(0.068)	(0.060)	(0.035)
# observations	55,344	55,344	55,344	55,344	55,344	55,344	55,344
R^2	0.022	0.025	0.025	0.028	0.025	0.018	0.019

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the 1-digit sector-commuting zone level in parentheses. Regression results from equation (3), estimated by OLS. All specifications contain 4-digit industry fixed effects and commuting zone fixed effects. In Panel (A0), the composition of the estimation sample evolves across columns. In the first two columns, all sample firms are included. In column (3), only firms exhibiting at least 2 layers (Layer 0 and Layer 1) in 2003 and 2006 are included in the estimation sample. Similarly, in columns (4) to (5), only firms exhibiting at least 3 or 4 layers respectively, both in 2003 and 2006, are included in the estimation sample.

firms.[29] Despite the resulting loss of statistical power for our regressions, we obtain that workers employed in firms that are subject to increases in their minimum wages tend to enrol more often in training programs (Arulampalam et al., 2004).[30] This applies to both production workers and managers. The estimated (semi-)elasticities for these two categories of workers are similar, and our estimates are particularly strong in the subsample of 3-layer firms. Overall, these results imply that a 1% increase in the minimum wage is, all else equal, associated with an increase in the probability to train either type of workers by 0.57 percentage point, on average. The 2003 to 2006 shock was thus associated with an increase in the frequency of training programs by up to 5 percentage points among firms belonging to GMR1 to 3. This impact is large, since it corresponds to approximately one sixth of the baseline frequency of training (around 30% in either case, as shown in Table B1).[31]

Table 3: Impact of Minimum Wage Increases on Training

Indicator:	Δ Proba. training at layer L_0			
	All firms	1-layer firms	2-layer firms	3-layer firms
Shares -1/+1:	0.07/0.25	0.08/0.20	0.07/0.25	0.06/0.28
Δ ln hourly cost at GMR	0.571***	0.582	0.314	0.860***
	(0.148)	(0.473)	(0.205)	(0.313)
# observations	17,406	2,494	9,688	4,569
R^2	0.083	0.228	0.114	0.181
Indicator:	Δ Proba. training at layers L_1 to L_3			
	All firms	1-layer firms	2-layer firms	3-layer firms
Shares -1/+1:	0.07/0.24	0.08/0.16	0.07/0.25	0.06/0.27
Δ ln hourly cost at GMR	0.560***	0.372	0.391*	0.489*
	(0.143)	(0.451)	(0.200)	(0.290)
# observations	17,406	2,494	9,688	4,569
R^2	0.080	0.246	0.105	0.194

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the 1-digit sector-commuting zone level in parentheses. The results correspond to equation (3), estimated by OLS (linear probability model). The dependent variable is the change in the probability of reporting positive hours of training, either for blue collar workers or employees (layer L_0) or for other workers (layers L_1 to L_3). All specifications contain 4-digit industry fixed effects and commuting zone fixed effects.

Lastly, Table 4 investigates the impact of increases in minimum wages on different proxies of firm size and productivity. In Panel (A), we obtain that the impact on the number of jobs is negative and slightly larger in magnitude than the estimated impact on hours. This reveals that firms responded to the minimum wage increases by not only simplifying their overall hierarchical organization, but also by reducing the total number of jobs to be managed. This necessarily implied discarding part-time jobs and increasing the resulting ratio of hours per job. Second, the estimated impact on value added and sales is close to zero, and insignificant. As a consequence, we estimate a positive impact on labor productivity in Panel (B). Columns (3) to (6) of Panel (B) in Table 4 check that this positive productivity impact still holds when controlling for the other production inputs – most importantly, capital. To that end, we compute four different indices of "revenue" TFP[32]

[29] This is due to the fact that firms with fewer than ten workers are not covered at all by the training data - see Section 3.2.

[30] This result contrasts with Neumark and Wascher (2001), but the difference is probably in large part due to the aforementioned freeze on layoffs, which severely limited firms' strategies for upgrading the skills of their workforce.

[31] As previously explained in Section 3.2, in the French institutional setting in terms of vocational training, these results have to be interpreted as evidence that firms exposed to tighter minimum wage constraints chose more often to "train" rather than "pay" (the penalty for insufficient training). Similarly, workers in these firms tended to make a more frequent use of the new "right to training" that was introduced during the estimation period, in 2004, and funded by their employers. These two mechanisms actually both imply that the corresponding training programs were funded by firms, and considered as part of their workers' compensation.

[32] Unfortunately, for the period of interest, the French information system does not contain firm-level price information. This prevents us from computing physical TFP indices.

according to alternative methods and re-estimate our specifications of interest. The obtained correlations are remarkably similar, with the estimated elasticities ranging between 0.07 and 0.08: taking account of all production factors (ie. capital and intermediate inputs, as well as labor) divides the simpler labor productivity estimates by three.[33] This implies that the 2003 to 2006 shock translated into TFP gains by up to 0.6 percentage points for the most affected firms.

Table 4: Impact of Minimum Wage Increases on Firms' Size and Productivity

(A) Impact on Firms' Size

Indicator:	(1) Δ ln hours	(2) Δ ln # jobs	(3) Δ ln VA	(4) Δ ln Sales
Mean, nominal:	0.014	0.009	0.084	0.106
Mean, real:	0.014	0.009	0.028	0.050
Δ ln hourly cost at GMR	-0.154***	-0.218***	0.052	0.006
	(0.050)	(0.052)	(0.065)	(0.054)
# observations	55,344	55,344	55,344	55,344
R^2	0.051	0.047	0.061	0.068

(B) Impact on Productivity

	(1)	(2)	(3)	(4)	(5)	(6)
Indicator:	Δ ln Labor		Δ ln TFP			
Measurement of TFP:	productivity		Accounting		Lev. Pet.	
Measurement of labor:	hours	jobs	hours	jobs	hours	jobs
Mean (deflated):	0.070	0.075	0.023	0.025	0.023	0.022
Δ ln hourly cost at GMR	0.206***	0.271***	0.068***	0.078***	0.074***	0.074***
	(0.056)	(0.062)	(0.022)	(0.023)	(0.023)	(0.024)
# observations	55,344	55,344	53,436	53,436	53,436	53,436
R^2	0.043	0.040	0.080	0.073	0.159	0.184

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the 1-digit sector-commuting zone level in parentheses. The results correspond to equation (3), estimated by OLS. All specifications contain 4-digit industry fixed effects and commuting zone fixed effects. In columns (3) and (4) of Panel (B), the productivity indices have been constructed using the accounting method, with the elasticities of value added to labor and capital being evaluated by their share of compensation in value added. In columns (5) and (6), the Levinsohn and Petrin (2003) estimation method is implemented using Wooldridge (2009) and Petrin and Levinsohn (2012). Labor is measured in hours in columns (1), (3), and (5), and in jobs in columns (2), (4), and (6).

4 Theoretical Framework

The previous regression results presented in Section 3.4 showed that French firms responded to increases in minimum wages by simultaneously adjusting their hierarchical organizations and their training decisions, while simultaneously experiencing significant increases in productivity. In this section, we propose a structural model of firms' organizations to assess whether the organizational adjustments that we observe in the data are likely to qualitatively and quantitatively rationalize the simultaneous firm-level productivity increases. It

[33]This result suggests that labor contributes to one third of the evolution of TFP, with the complement being driven by alternative production factors. Table C2 in Appendix C describes the evolution of firms' demand for capital and intermediate inputs. Despite the relative increase in the cost of labor compared to the cost of other factors of production, firms do not seem to significantly adjust their demand for any other production factors: if anything, this reallocation strategy is really weak, such that the obtained estimates are negative but not statistically significant. As a consequence, the shares of intermediate inputs (Table C2, column 7) and capital (Table C2, column 8) remain broadly stable. This means that firms do not appear to have outsourced more of their production in response to minimum wage increases, and nor have they significantly altered their overall capital intensity.

also allows us to investigate further general equilibrium implications of minimum wages, most importantly in terms of aggregate productivity.[34] Our theoretical set-up is derived from the heterogeneous firm framework proposed by Caliendo and Rossi-Hansberg (2012), where we further assume that knowledge is cumulative in order to fit the French data parsimoniously.[35] The implications of this deviation[36] and, above all, the introduction of a minimum wage constraint are better understood in a self-contained exposition of the model, which we provide in this section.

4.1 Optimal Firms' Organizations when Knowledge is Costly and Cumulative

Demand. We consider an economy that is populated by a measure N of homogeneous agents,[37] which we normalize to 1 without loss of generality. On the demand side, these agents correspond to consumers with CES utility function:

$$U(x) = \left(\int_{\Omega} \alpha^{\frac{1}{\sigma}} x(\alpha)^{\frac{\sigma-1}{\sigma}} M \mu(\alpha) \mathrm{d}\alpha \right)^{\frac{\sigma}{\sigma-1}}, \tag{5}$$

where $\alpha \in \Omega$ denotes a variety which is valued more by consumers whenever α takes a high value, $\sigma > 1$ is the elasticity of substitution across varieties, x denotes consumption, M corresponds to the mass of products while μ is the probability density function over the set Ω of available varieties. In this expression, M and μ are endogenous objects which are derived later in the model. This CES utility function generates the following (inverse) demand schedule at the variety level:

$$p(\alpha) = q(\alpha)^{-\frac{1}{\sigma}} (\alpha I)^{\frac{1}{\sigma}} P^{\frac{\sigma-1}{\sigma}}, \tag{6}$$

where $q(\alpha)$ corresponds to aggregate demand, ie. the summation of $x(\alpha)$ across all agents, and I corresponds to aggregate income, which is also endogenously determined in the model. In what follows, we normalize the CES price index $P = \left(\int_{\Omega} \alpha p(\alpha)^{1-\sigma} M \mu(\alpha) \mathrm{d}\alpha \right)^{\frac{1}{1-\sigma}}$ to 1 without loss of generality.

Supply: occupation choices. On the supply side, to keep the model as parsimonious as possible, only one production factor - labor - is taken into account. All costs are therefore delineated in terms of labor. Agents are each endowed with one unit of time that they supply inelastically. However, they can choose between three different types of occupations: entrepreneurs of firms (be they employers of other workers or simply self-employed entrepreneurs), paid workers directly contributing to production within firms (at the bottom production level, or as managers), or teachers.

Agents deciding to become entrepreneurs pay a fixed cost of entry $k f_E$, where k is the price of a unit of time, which is endogenously determined in equilibrium, and f_E is the number of workers required to perform the fixed-cost tasks. In the model, the latter essentially enable agents to get a draw α from an

[34] Notice that a "standard" Melitz (2003) model could not rationalize the observed within-firm productivity increases. As shown below, our set-up is able to rationalize this margin, but also introduces the possibility of selection effects as in Melitz (2003) (delineated in terms of the demand shifter, α, that we introduce in Equation 5). The set-up thus allows for a comparison of the magnitude of these two channels for the global impact of minimum wages on aggregate productivity.

[35] This variation ensures simultaneously that (1) wages are increasing across layers and, as a consequence, that (2) the legal minimum wage only bites at the bottom layer of firms' hierarchical organizations (Section 4.2). Both are features of the French data (Section 3.3) and the second characteristic eases numerical simulations considerably.

[36] As explained in full details below, the parametric restriction introduced by Caliendo and Rossi-Hansberg (2012), which implied that production workers always acquire a strictly positive level of knowledge, is rejected in the French data. This has three implications: first, it widens the set of organizational strategies available to firms. Second, it complicates the proofs of some of the central results (which however still hold). Lastly, it also reintroduces a fair amount of numerical challenge.

[37] In our empirical and simulation analyses, we focus on firm-level outcomes: firm size, number of layers, productivity, etc. As a result, modeling firm heterogeneity is very important, which motivates our choice of the Caliendo and Rossi-Hansberg (2012) framework as a baseline for our theoretical model. However, such models cannot incorporate ex-ante worker heterogeneity without making the matching problem exceptionally difficult, and as a result we have to impose ex-ante homogeneity of agents in our model.

exogeneously given cumulative density function $G(\alpha)$ (with associated probability density function $g(\alpha)$). This draw determines the demand conditions faced by the candidate firms (according to Equation 5), and thus their level of profits. Depending on the value of the idiosyncratic α draw, the agent decides whether to actually become an entrepreneur and launch production, or not. If production actually starts, the entrepreneur has to incur an additional fixed cost kf.

Production consists in solving problems of varying "difficulty",[38] which is indexed by $d \in \mathbb{R}^+$ and is drawn from a cumulative density function $F(d) = 1 - e^{-\lambda d}$. By default, agents are not able to solve production problems autonomously ($z = 0$). However, it is possible for them to increase their skills to a higher level $z > 0$ subject to a training cost kcz. This level of ability enables them to solve all problems with difficulty $d \leq z$. The associated time cost is their entire time endowment (ie., 1) when they have to solve the production problem autonomously, but only $h < 1$ whenever the problem has been pre-processed by another agent with skill $z' < d$, who was not able to solve the problem autonomously. Notice that agents featuring the minimal skill level $z' = 0$ are useful in production precisely because of their ability to pre-process production problems, which allows their managers to solve them at a reduced time cost ($h < 1$). In contrast to the formulation of the model in Caliendo and Rossi-Hansberg (2012), we assume that knowledge is cumulative: achieving knowledge level z requires mastering all intermediate levels between 0 and z.[39] The labor market is competitive, and in a mixed-strategy equilibrium workers must be compensated for their skill investment, such that workers earn the following gross wage:[40]

$$w(z) = k(cz + 1) \tag{7}$$

As a result, after paying for their training, all workers end up with a net wage of k, which is the price of a unit of time which is endogenously determined in equilibrium.

Actual production within hierarchical organizations requires knowledge (aka the ability to solve problems) and labor (aka time). Bottom production workers (at level $L = 0$) spend their entire endowment of time studying the problem they pick, and produce A units of output when their level of knowledge allows them to solve it. When they do not know the solution, they ask a manager in the next layer above, at level $L = 1$. The latter spends h units of time studying the problem and solves it if d is less than the manager's skill level (thus producing A units of output). In the opposite case, she asks a manager in the layer above (ie., layer $L = 2$, if there is one) and the process repeats. Some workers are also assigned to perform the fixed-cost tasks embodied in f_E and f, which does not require acquiring knowledge ($z = 0$).[41]

Lastly, teachers educate workers: learning one unit of knowledge requires c units of teachers' time, who earn k per unit of time. The education sector is competitive such that education services are charged at marginal cost, ie. kcz in order to achieve level z.

[38] More precisely, production problems are heterogeneous in the sense that solving problems indexed by different d require different types of knowledge. The notion of "difficulty" is not completely appropriate, but is a convenient (yet approximate) label to describe the setting. In fact, the specification of $F()$ involves sorting problems according to their (decreasing) frequency (rather than according to their "difficulty"): $d = 0$ thus simply corresponds to the most frequent problems.

[39] In the non-overlapping skill specification from Caliendo and Rossi-Hansberg (2012), managers can generically be allocated a smaller interval of skill than the workers in the layers below them. Such a situation would imply that they should receive a lower wage. This would also imply that the minimum wage could impact managers before production workers or lower-layer managers (and incidentally, complicate numerical simulations by a large amount). Figure 2 shows that in the data, there is a clear hierarchy of wages across layers and that minimum wages only bind at bottom layers. In practice, the cumulative knowledge assumption thus provides a better approximation of the French data.

[40] In the specification of Equation 7, training costs are incorporated in workers' compensation, which corresponds to the French context, as discussed in Section 3. In fact, Equation 7 corresponds to the most parsimonious wage specification possible given our focus on the productivity implications of minimum wages (Caliendo and Rossi-Hansberg, 2012), rather than their impact on worker-level outcomes. Introducing richer mechanisms of wage formation, notably incorporating dimensions of *ex ante* worker-level heterogeneity as in Garicano and Rossi-Hansberg (2006) and capturing more sources of heterogeneity/inequality, would involve a (truly) significant increase in complexity and is beyond the scope of this paper. We leave this important task for future research.

[41] This assumption is rather innocuous since it only affects the breakdown of workers between fixed costs and the educational sector, which is not the main focus of our article.

Supply: firms' organizations. Each firm is set up by an entrepreneur who is located at the top of the hierarchy (as a CEO), and who decides:

- The total number of layers $L \in \{0, 1, 2, 3, ...\}$ of her firm. The case $L = 0$ corresponds to self-employed managers.
- The number of workers n_L^l to be hired at each layer $l \in \{0, 1, ..., L\}$
- The skill level z_L^l for each layer.

We derive the optimal firm structure in two steps: first, a cost-minimization problem conditional on the level of output q that has to be attained, and second, a profit-maximization step allowing to determine the optimal level of output q conditional on the demand draw α.

Supply: cost minimization. First, the production conditions described above imply straightforwardly that the levels of skill must be increasing across layers - otherwise, there is no benefit to passing an unsolved problem to agents that are less competent than the considered worker. Second, it is necessary to hire enough workers at a given layer l in order to screen the total amount of problems that are transmitted at that hierarchical level. This implies the following inequalities:

$$\forall l \in \{0, 1, ..., L-1\}, \qquad \underbrace{h}_{\substack{\text{Time per}\\ \text{problem}}} \underbrace{n_L^0 e^{-\lambda z_L^l}}_{\substack{\#\text{ problems}\\ \text{transmitted}\\ \text{to layer } l+1}} \leq n_L^{l+1} \tag{8}$$

$$n_L^L = 1 \tag{9}$$

$$0 \leq z_L^0 \tag{10}$$

$$q \leq n_L^0 A(1 - e^{-\lambda z_L^L}) \tag{11}$$

The variable cost of producing a level q of output given wages k is:

$$C(q, k) = \min_{L \geq 0, (n_L^l)_l, (z_L^l)_l} C_L(q, k) = k \left(\sum_{l=0}^{L} n_L^l (c z_L^l + 1) \right) \tag{12}$$

where $C_L(q, k)$ corresponds to the cost of producing q with a firm organized into L layers. The resolution of each of these sub-programs is detailed in Appendix E. The proofs differ from those in Caliendo and Rossi-Hansberg (2012) because of two main deviations:

- First, as previously mentioned, we introduce an assumption of cumulative (or "overlapping") knowledge imposing that any agent who is able to solve a problem of difficulty z is also able to solve any easier problem, which ensures that wages increase monotonically from the bottom to the top of the hierarchies.[42]
- Second and more importantly, we do not impose their parametric restriction ensuring that bottom, production workers always acquire a strictly positive level of knowledge, $z_L^0 > 0$, because we later obtain from the model calibration that this restriction – which corresponds to $\frac{c}{\lambda} \leq h$ in our model – is rejected in the French data. This indicates that in France, given the prevalence of difficult problems

[42] However, this feature comes at the cost of a lower tractability, and in particular non-monotonic marginal costs (see Proposition 5 in Appendix E), depending on the number of layers and on whether firms find it optimal to train production workers or not. This in turn is determined by their respective demand draws, α.

captured by parameter λ, it is particularly expensive to train workers (c is high) as compared with transferring problems upwards to skilled managers, given the relatively low managerial costs (h is low):

$$\frac{c}{\lambda} > h \tag{13}$$

Technically, this feature complicates some of the proofs of Appendix E. Economically, it implies that in the firms' optimization problem described in Equation 12, it is relatively more cost-effective for them to rely on verticalized hierarchical organizations than to train workers within layers to achieve a given output target.[43] Skills are so expensive that $L = 1$-layer firms sometimes do not find it optimal to train production workers at the bottom ($z_1^0 = 0$) and even further reduce labor costs at the bottom ($l = 0$) layer by compressing its size below their manager's capacity. In such cases, the manager (at layer $l = 1$) does not exhaust his or her management time constraint.[44]

However, several important characteristics of the baseline Caliendo and Rossi-Hansberg (2012) framework are preserved despite these two deviations. In Appendix E, we show in particular that the firms' cost function remains sufficiently well behaved: it remains increasing in q, and homogeneous of degree one in the baseline labor compensation k, such that the cost-minimizing firm organization for a given output target q is insensitive to k. Importantly, this implies that the $Q-$productivity index proposed in Caliendo and Rossi-Hansberg (2012) remains valid, which is obviously key to our research question. It simply corresponds to the inverse of average variable costs (without considering the fixed costs of production), normalizing k to one so as to keep factor prices constant across the different general equilibrium scenarios that we will compare in our simulations:

$$a(q) = \frac{q}{C(q,1)} = \frac{kq}{C(q,k)} \tag{14}$$

Figures 3 and 4 illustrate graphically these different characteristics of our set-up. Panel (A) of Figure 3 constructs the curves describing average costs of production in firms featuring different numbers of layers, from $L = 0$ (self-employed agents) to $L = 3$ (one bottom layer of production workers, and three managerial layers). It allows us to determine the optimal number of hierarchical layers, L, depending on the production target, q, which corresponds to the value associated to the lowest possible average cost: the overall cost function thus corresponds to the lower envelope of the curves of Figure 3. Increasing the number of hierarchical layers deteriorates units costs for low scales of production, but becomes cost-effective as the production target becomes larger: adding a hierarchical layer is thus similar to incurring a higher fixed cost of production in order to achieve lower marginal costs of production for large scales of production, as shown in Panel (B) of Figure 3.

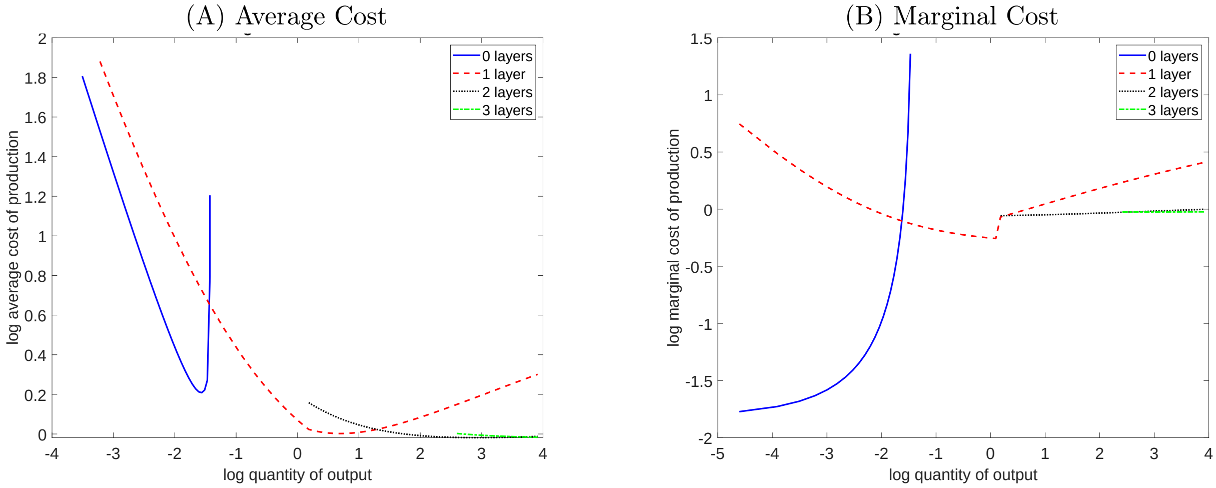
More technically, Panel (A) of Figure 3 also shows that average costs are functions that are continuous, but not continuously differentiable. In particular, 1-layer firms feature discontinuous marginal costs of production, and the average cost function features a point of non-differentiability, at the point where they start finding it optimal to train bottom workers, ie. when z_1^0 switches from zero to a strictly positive value.[45] Meanwhile, employment at the bottom layer (n_1^0) increases from values which do not saturate the time constraint of the manager, to values where the latter works at full capacity (given z_1^0 and thus the total amount of problems

[43] This feature, which captures an important characteristic of the French economy (see Section 5), will also render the minimum wage constraint introduced in Section 4.2 even more binding. Overall, this parametric condition also implies that our quantifications consist of a rather upper bound about the aggregate cost of minimum wages since, as we show later, mitigating their cost at the firm level implies relying less on hierarchies and more on training, ie. less on the cheap technology, and more on the relatively costly one.

[44] In larger firms ($L > 1$), on the other hand, the time constraints of managers and entrepreneurs are always exhausted.

[45] As long as $L \geq 1$ and as shown in Panel (B) of Figure 3 (eg. blue and red curves), marginal production costs are not always increasing in the production target, q. Note that this constitutes a difference with Caliendo and Rossi-Hansberg (2012) (see Appendix E).

Figure 3: Firms' Cost Functions



Notes: This figure is constructed from the calibrated version of the model presented in Section 5, in a version which simulates the French economy in the absence of a minimum wage. The parameters are calibrated to the following values: $\lambda = 1$, $c = 0.220$, $h = 0.194$, $A = 0.240$.

that are transmitted to her, $hn_1^0 e^{-\lambda z_1^0} = 1$).

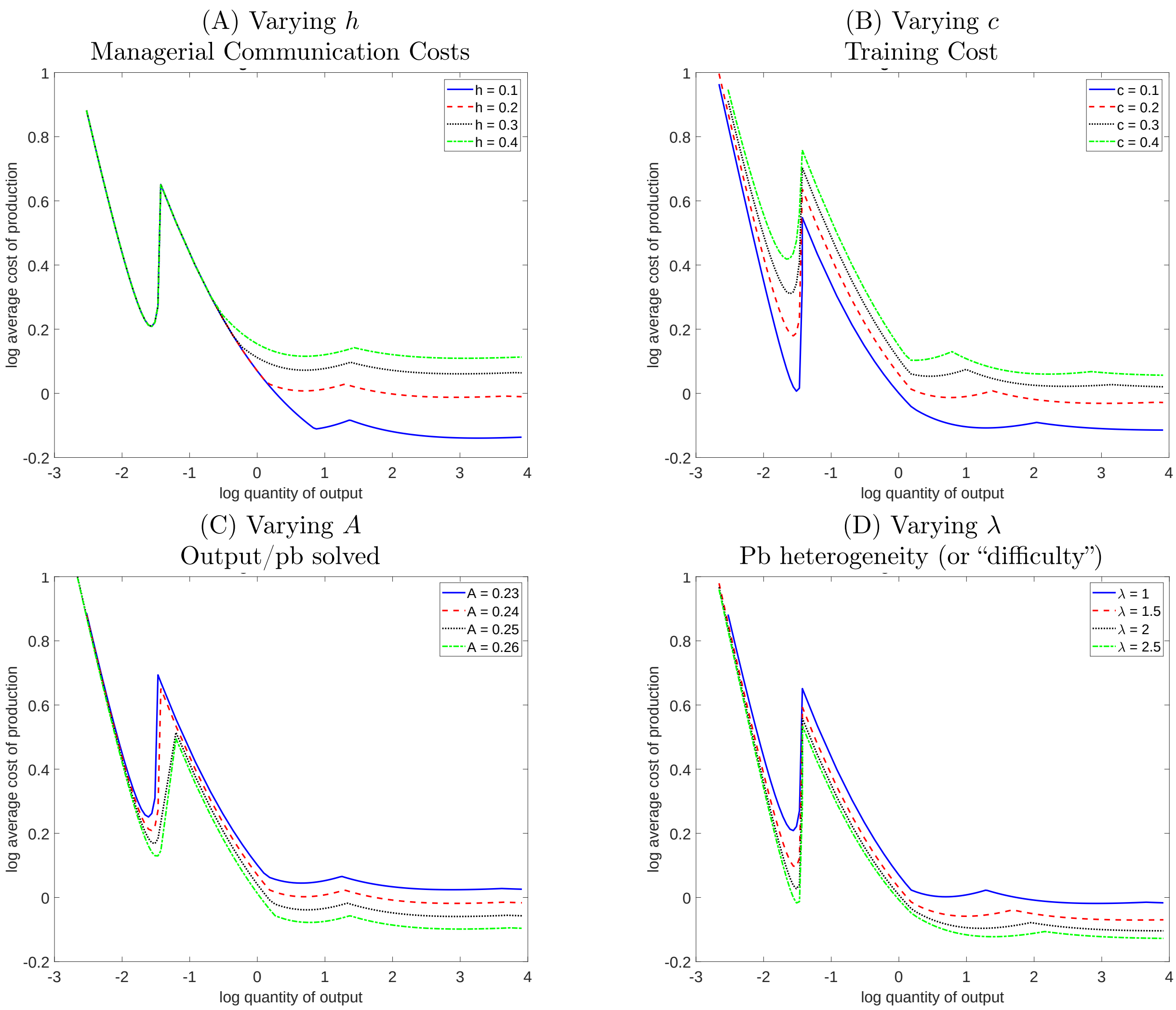
Figure 4 documents how average costs are shaped by the main technological parameters of the model. In Panel (A), we first consider the impact of an increase in the cost h of transmitting problems to managers. Unsurprisingly, production costs are unaffected when $L = 0$ (self-employment). They are also unaffected in firms with $L = 1$ hierarchical layer, when the time (capacity) constraint of the manager is not saturated, ie. $n_1^0 < \frac{\exp(\lambda z_1^0)}{h}$. In all other firms, costs curves are shifted upwards. As h becomes larger, Equation 13 no longer holds and the regime where the time constraint of managers in 1-layer firms is not saturated disappears: it is never optimal to waste managerial time when management costs are high relative to training costs. Panel (B) precisely considers the impact of the latter (parameter c). They affect all organizational forms and shift all curves upwards. As c becomes large, Equation 13 generates a wider range of optimality for organizations which do not exhaust managers' time constraint. In contrast in Panel (D), higher values of λ, which corresponds to more favorable (more concentrated) distribution of production problems, affects all organizational forms but relaxes Equation 13, thus reducing the range of optimality for organizations which do not exhaust managers' time constraint. Lastly in Panel (C), the general efficiency parameter, A, affects all organizational forms but is fully neutral in terms of the range of optimality for the different organizational forms.

Supply: profit maximization. Ultimately, all entrepreneurs solve the following profit maximization program:

$$\pi(\alpha, k) = \max_{q \geq 0} \underbrace{p(\alpha)q}_{q^{\frac{\sigma-1}{\sigma}}(\alpha k)^{\frac{1}{\sigma}}} - C(q, k) - kf, \tag{15}$$

which allows them to select the optimal organizational form $\left(L, (n_L^l)_l, (z_L^l)_l\right)$ as well as the optimal level of production q depending on their idiosyncratic demand conditions (captured by α). In such a CES framework

Figure 4: Firms' Average Cost Function: Comparative Statics wrt Technologies



Notes: This figure is constructed from the model presented in Section 4 and calibrated in Section 5, in a version which simulates the French economy in the absence of a minimum wage. In Panel (A), the figure represents the lower envelope of the layer-specific average cost curves, calculated for 4 different values of h, while $c = 0.220$, $\lambda = 1$, $A = 0.240$. In Panel (B), this exercise is replicated for 4 different values of c, while h is set to 0.194, $\lambda = 1$, $A = 0.240$. In Panel (C), this exercise is replicated for 4 different values of A, while $h = 0.194$, $\lambda = 1$ and $c = 0.220$. Lastly, the exercise is also replicated in Panel (D) for 4 different values of λ, while $h = 0.194$, $A = 0.240$, and $c = 0.220$.

with monopolistic competition, optimal pricing is a constant markup over marginal cost:

$$p(\alpha, k) = \frac{\sigma}{\sigma - 1} MC\left(q(\alpha); k\right) \tag{16}$$

such that the optimal quantity to be produced satisfies:

$$q(\alpha, k) = \alpha k \left(\frac{\sigma}{\sigma - 1} MC\left(q(\alpha); k\right)\right)^{-\sigma}. \tag{17}$$

General equilibrium. To close the model in general equilibrium, we assume an exogenous firm death rate of δ per period. The equilibrium is defined by the three following conditions:

1. Conditional on entry (and once the demand conditions α are revealed), entrepreneurs only produce if their profit in Equation 15 (which is net of entry fixed costs) is positive. Since profits are increasing in α, this implies that they only produce if their actual α draw is larger than a cutoff $\overline{\alpha}$ defined by the following equation:
$$\pi(\overline{\alpha}, k) = 0. \tag{18}$$

2. Entrepreneurs only enter (ie. pay the fixed entry cost, f_E) if their expected profit (when α is not yet revealed) is positive:
$$\int_{\overline{\alpha}}^{+\infty} \frac{\pi(\alpha, k)}{\delta} g(\alpha) d\alpha = k f_E. \tag{19}$$

3. In a steady state, in each period, the $M\delta$ exiting firms are exactly replaced by the subset of entrants who remain in activity, $M^E\left(1 - G(\overline{\alpha})\right)$ (while all entrants, M^E, pay the fixed entry cost, f_E).
The labor market clearing equation is thus:[46]
$$1 = \frac{M}{1 - G(\overline{\alpha})}\left[\delta f_E + \int_{\overline{\alpha}}^{+\infty} \left(C(q(\alpha); 1) + f\right) g(\alpha) d\alpha\right]. \tag{20}$$
The term on the left in Equation 20 corresponds to the supply of labor that is readily available to be hired as production workers or managers (which coincide with the mass of agents and is normalized to one), while the term on the right corresponds to total labor demand (with associated training time requirement directed to teachers, fixed costs of production, and fixed cost of entry in each period) by entrepreneurs who actually launch production.

This system of three equations determines the equilibrium values of the productivity entry threshold, $\overline{\alpha}$, the measure of active firms, M, and the equilibrium price of a unit of time, k.[47]

4.2 Introducing the Legal Minimum Wage Constraint

We now introduce a minimum wage constraint in the previous setup. This constraint takes the form of a lower bound, denoted by $\overline{w}$, that is imposed on all (real) wages that are paid out by firms featuring $L \geq 1$. Specifically, we assume that the minimum wage constraint does not apply to self-employed workers (corresponding to $L = 0$), which is meant to reflect real-world characteristics, since self-employed agents can easily circumvent the minimum wage constraint by declaring their own compensation as dividends or other types

[46] This expression relies on the fact that the cost function C is homogeneous of degree 1 in k: see the proof in Appendix E.

[47] Notice that these equations also determine the endogenous objects that we introduced earlier. In particular, the probability density function of varieties that are available on the market relates to g and $\overline{\alpha}$ via the following formula: $\mu(\alpha) = \frac{g(\alpha)}{1-G(\overline{\alpha})}$ for all $\alpha \geq \overline{\alpha}$. Secondly, good market clearing requires: $I = kN = k$, where I was introduced in Equation 6 and where the last equality follows from the fact that N was normalized to 1. This implies that Equation 6 can be rewritten as: $p(\alpha) = q(\alpha)^{-\frac{1}{\sigma}} (\alpha k)^{\frac{1}{\sigma}} P^{\frac{\sigma-1}{\sigma}}$.

of capital income rather than wages. Secondly, we also assume that the wage k of teachers is not subject to labor laws, and therefore unaffected by the introduction of the minimum wage constraint. This second hypothesis aims to account for the French institutional framework, where teachers are almost all civil servants while general labor laws (and the national minimum wage in particular) do not apply to civil servants.[48]

Such a minimum wage constraint on real wages paid out by firms can equivalently be rewritten as a constraint on the minimum level of knowledge to be attained at the lower level of firms' hierarchies:

$$w(z_L^0) = k(cz_L^0 + 1) \geq \overline{w} \quad \iff \quad z_L^0 \geq \underbrace{\frac{\overline{w} - k}{ck}}_{\overline{z}_L^0} \tag{21}$$

This condition reflects the fact that hiring a worker with a skill level below $\overline{z}_L^0$ is always dominated by hiring one at $\overline{z}_L^0$ for a given value $\overline{w}$ of the minimum wage. The minimum wage constraint thus simply alters the cost minimization program of firms with $L \geq 1$, by imposing a tighter constraint on z_L^0 in our previous Equation 10. The proof strategies are globally unaltered, as shown in Appendix E.3.1, but this addition to the set-up has non-trivial implications that we explain below.

Implications in partial equilibrium. Appendix E.3.2 first provides some important comparative statics results in partial equilibrium, ie. holding k fixed, but endogeneizing the optimal level of output, q.

First, as illustrated in Figure 5, the minimum wage constraint leads to a general shift of all cost curves both upwards and to the right. This means that all else equal, in partial equilibrium, production costs increase as a consequence of the introduction of a binding minimum wage constraint, and the range of optimality of the different organizational forms is shifted to the right: for a given production target, q, optimal organizational forms tend to be (slightly) flatter. In addition, it increases the range of production targets for which it is optimal for 1-layer firms to remain "small" relative to the time capacity of their managers. This property is illustrated in Figure 5, where the range of production over which such organizational forms are optimal unambiguously increases as $\overline{w}$ increases.

Proposition 1 formalizes these first results by considering firms who are "marginal" between organizational forms and investigating their responses to tighter minimum wage constraints:[49]

Proposition 1. *If the minimum wage is binding on firms with M layers but not L layers, then any firm that is on the margin between choosing M or L will shift to production with L layers.*

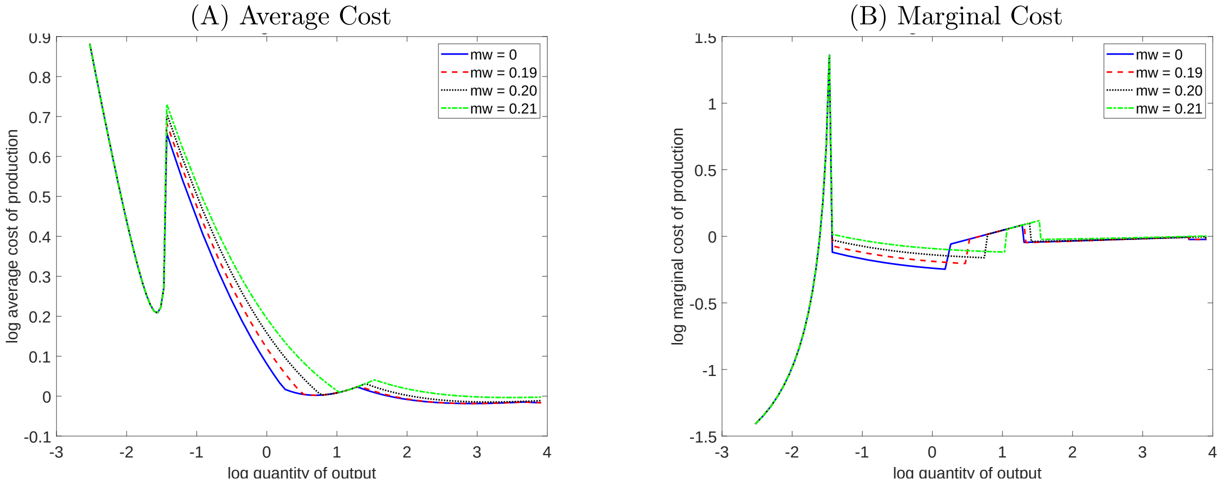
Panel (A) in Figure 5 shows that average costs always rise when firms are faced with stricter minimum wage constraints, while Panel (B) describes that the same is not true for marginal costs, which sometimes decrease depending on the range of production that is considered. This is mainly a consequence of Proposition 1, because marginal costs are both non-monotonic and mainly determined by the number of layers (and, in 1-layer firms, by whether the time constraint of the manager is slack or tight): since minimum wages alter the range of optimality of these different regimes, they have non-monotonic implications in terms of marginal production costs (and hence, in terms of firms' pricing strategies).

Of course, higher minimum wages not only change the optimal number of hierarchical levels, but also the optimal level of knowledge across hierarchical layers since bottom, production workers become more

[48]None of these assumptions is critical, and our framework could easily be modified to describe a different institutional context. We also assume that workers performing fixed-cost tasks are not subject to the minimum wage.

[49]The proof of this proposition is reported in Appendix E.3.2. It is impossible to prove analytically that firms monotonically increase their number of layers with q, or with α in general equilibrium. However, such a monotonic pattern is true in all of our simulations. In such situations, and whenever $z_{L+1}^0 < z_L^0$ at the point of indifference, then Proposition 1 implies that the minimum wage causes marginal firms to remove a hierarchical layer.

Figure 5: Firms' Cost Function: Comparative Statics wrt Minimum Wages, $\overline{w}$



Notes: This figure is constructed from the calibrated version model presented in Section 5, in versions which simulate the French economy with different levels of the minimum wage, $\overline{w}$ (measured in hundreds of € in both panels). The parameters are calibrated to the following values: $\lambda = 1$, $c = 0.220$, $h = 0.194$, $A = 0.240$.

knowledgeable (Equation 21). Appendix E.3.1 shows that whether it is optimal to increase or decrease skills across managerial layers depends in particular on the number of layers. For example, when the manager does not reach her time constraint in 1-layer firms, it is optimal to *decrease* her level of knowledge when production workers become more skilled because of a higher minimum wage, ie. the skills of the manager and of the production workers are substitutes in this configuration, which limits the increase in marginal costs of production generated by minimum wages as much as possible. In contrast, when the manager exhaust her time constraint in 1-layer firms, it is optimal to *increase* her level of knowledge, which means that her skills are complements to those of production workers. This strategy allows maintaining low marginal costs for sufficiently large scales of production: over the corresponding range (which is narrow!), average and marginal costs are unaffected by minimum wages in Figure 5.[50]

Proposition 2 formalizes these results by considering the impact of tighter minimum wage constraints for firms which are in a regime (ie. a range of output targets, q) where it is not optimal to alter their hierarchical organizations as a response to these shifts. It exhibits some well-defined cases[51] where clear theoretical predictions arise in terms of the impact of minimum wages on marginal costs of production, ϕ, the size of the bottom production layer, n_L^0, and the skills of their managers, z_L^l. However, as previously explained, they appear to be quite heterogeneous, which motivates the simulation techniques that are introduced in Section 5:

Proposition 2. *Given a number L of layers, a (marginal) increase in a binding minimum wage has the following impact on firms, depending on their optimal hierarchical organization:*

- *For $L = 1$, if n_0^1 is below the level which would saturate the time constraint of the manager,*[52] *z_1^1 and n_1^0 decrease and the marginal cost of production, which we denote by ϕ, increases with the minimum wage if and only if $z_1^1 > \frac{\sigma-1}{\lambda}$. Intuitively, firms in this regime need to charge higher prices and contract production to limit the costs generated by higher minimum wages.*

[50] This non-monotonicity of the impact of minimum wages also implies that their impact on the prices charged by firms is not monotonic either, as a consequence of Equation 16.

[51] While the condition which arises in Proposition 2, $z_L^L > \frac{\sigma-1}{\lambda}$, depends on an endogenous outcome of the model (z_L^L), we should note that, in our simulations, this corresponds to a weak assumption. As a consequence, many firms seem to fall into these different cases in our later simulations.

[52] Technically, this implies that the Lagrange multiplier θ introduced in Appendix E.3.2 is constrained to 0.

- *For $L = 1$, if n_0^1 saturates the time constraint of the manager, then z_1^1 and n_1^0 increase and ϕ decreases with the minimum wage. This means that the optimal strategy in this regime is the opposite of the previous: firms need to lower prices and increase their volume of production in order to absorb the minimum wage contraint.*

- *For $L \geq 2$, if $z_L^L > \frac{\sigma-1}{\lambda}$ and $h \leq 0.25$, then n_L^0 and all z_L^l increase while ϕ decreases with a marginally-binding minimum wage.*
  However, as the minimum wage continues to increase and becomes more than marginally binding on z_L^0, the results become ambiguous. Indeed, the increase in production costs tends to lower the optimal output target q by an amount that becomes sufficient to offset the initial tendency to increase n_L^0. This means that the optimal strategy in this regime is first to increase volumes in order to achieve lower marginal costs and decrease prices, up to a point where the converse holds true (until it becomes optimal to remove a hierarchical layer to save on fixed production costs, see Proposition 1).
  In addition, for $L \geq 2$, the effect on the number of workers in higher layers of the firm is always ambiguous, as is the effect on total firm size.

Further implications in general equilibrium. In general equilibrium, an increase in the minimum wage lowers the profits of at least some firms such that the price of labor, k, must decrease to restore balance in the labor market. This in turn alters the partial equilibrium results: a reduction in k tends to lower marginal cost, but it also lowers q due to a reduction in income. The combined effect of these forces is ambiguous on firms' organizations and thus on the productivity implications of minimum wages. Quantifying them requires numerical simulation techniques, which are detailed in Section 5.

5 Calibration and Simulations

Using our model from Section 4, we now proceed to a careful calibration exercise, with three different objectives in mind. Firstly, we check that the proposed set-up is compatible with the cross-sectional structure of the French firm population, which is captured by our (static) calibration targets. Secondly, we check that our model is further "validated" by the French data, by verifying that the organizational mechanisms at the heart of our model are capable of rationalizing, both qualitatively and quantitatively, the responses of French firms to the tightening of minimum wage constraints that we documented in Section 3. In technical terms, these correspond to "dynamic" moments in the data that are not used as targets in the calibration procedure: for this reason, this verification is a powerful test of the framework's empirical relevance. Thirdly and finally, we use our calibrated framework to describe and quantify the impact of minimum wage constraints of different magnitudes on firms' strategies and on the resulting aggregate productivity. We extend the analysis to different counterfactual technological scenarios, with the aim to describe the possible implications of the current waves of technologies.

5.1 Baseline Calibration

Data calibration targets. We calibrate our model to a set of moments measured from our dataset in 2006, which corresponds to the endpoint of our estimation period in Section 3 and to a reasonable approximation to a period of stable equilibrium in the French economy.[53] At this date, all GMRs had converged, thus all firms

[53]See Section 3.2 for the detailed description of our estimation sample. It implies in particular that we focus solely on the population of relatively stable firms, which have been consistently active during the estimation period of Section 3, ie. over a period of 7 years (2000 to 2006), and which are part of GMRs. Appendix F complements the analysis by checking the robustness of our findings to four alternative experiments, each relying on different, more comprehensive (up to virtually quasi-exhaustive)

faced the same level of legal minimum wage, as in our theoretical set-up of Section 4. Panel (A) of Table 5 describes the precise moments of the data which we use as calibration targets. Since our mechanism of main interest involves firms' optimal organizations, we introduce a rich set of indicators describing the firm size distribution and its articulation with production hierarchies: first, the breakdown of the firm population by classes of hierarchical organizations (except for the two bottom classes, which are more difficult to identify separately in the DADS),[54] and secondly, firms' average size in each of these classes. These five moments contribute jointly to the identification of parameters $\frac{c}{\lambda}$, h, A as well as the distribution of α, which all determine firms' sizes and organizations (see firms' optimization programs in Section 4 and the characterization of optimal choices in Appendix E). The average hourly wage mainly contributes to identifying $\frac{c}{\lambda}$ and k, while the share of workers bound by the minimum wage mainly identifies $\overline{w}$.[55] Parameter δ is directly measured in the data from the percentage of firms in our estimation sample that were created between 1993 and 1999. This follows from the fact that a "time period" in our empirical setting of Section 3 corresponds to 7 consecutive years, since all sample firms were consistently active between 2000 and 2006. The share of workers in these "new" firms contributes mainly to identifying the fixed costs f_E and f as well as the distribution of α (but also c, h and A).

Parameters that are calibrated to values from previous literature. We set the values of the other parameters of our model to values from previous literature. Specifically, the elasticity of substitution σ is set to 3.8 as in Caliendo and Rossi-Hansberg (2012), and as in the latter paper, we normalize λ to 1. In addition, the share of workers who are teachers could not be accurately estimated in the DADS, because education is largely in the public sector in France and is only poorly captured in the 2006 version of the DADS files. This is however a critical moment to identify the fixed costs (f_E, f) together with the distribution of α, such that we thus rely on external estimates made available by the French statistical institute.[56]

Specification of the distribution of α. The distribution of α embodies the only source of firm-level heterogeneity that is introduced in the model. It thus plays a critical role in terms of fit with the data. For this reason, we adopt a flexible specification for the distribution of α: specifically, we set a wide support in terms of its distribution, by allowing values ranging between 1 and 1000. This implies that a firm earning a draw of 1000 can sell 1000 more quantity at any price than a firm earning a draw of 1: this wide range allows for possibly highly heterogeneous demand conditions. We also approximate non-parametrically the density $g(\alpha)$ by taking the exponential function of a cubic spline across four nodes, located at $\alpha = \{1, 50, 500, 1000\}$, where the values of the spline function at the four nodes $\{g_1, g_2, g_3, g_4\}$ are parameters set in our calibration –

datasets.

[54] A difficulty with the DADS data is that is it not straightforward to distinguish firms with $L = 0$ from firms with $L = 1$. To be precise, we observe many firms in these files, which are registered as proper corporations, but which only feature one production layer, and no managerial layer. The latter are likely to either correspond to partnerships (which should probably classified in the $L = 0$ class) or to small 1-layer firms with managers that are not paid in the form of wages. This is allowed in the smallest and simplest structures (smallest "SARL", the equivalent of US LLCs), but not in standard corporations (large SARL, "SA" and "SAS", the equivalent of US joint-stock companies), such that in practice this problem only affects the measurement of 1-layer firms. To mitigate this concern, the two classes ($L = 0$ and $L = 1$) are considered jointly in our main calibration. Appendix F also checks that our results are not sensitive to the fact the 2006 version of the matched DADS and fiscal files might underestimate the share of self-employed entrepreneurs ($L = 0$).

[55] In our calibration, we follow the practice of the Ministry of Labor (Ananian and Calavrezo, 2011), which relies on a rather restrictive definition to assess this share: it amounts to focusing on workers in our sample earning less than 1.05 times the minimum wage. With such a definition, we obtain that around 5% of workers are bound by the minimum wage, as in Ananian and Calavrezo (2011). This share appears to almost coincide with the share of UK workers which were "hit" by the minimum wage introduced in 1999: Dickens and Manning (2004) estimate it to be around 6 to 7%.

[56] INSEE's Tableaux de l'Economie Française (2020 edition, data for 2018). The URL for these data is the following: https://www.insee.fr/fr/statistiques/4277675?sommaire=4318291. We excluded all sectors not covered in the DADS (administration, social action) and rescaled the proportion of teachers accordingly, thus attaining a value that is quite close to the US figure used in Caliendo and Rossi-Hansberg (2012). We were unfortunately not able to find data for 2006 and thus had to assume that the proportion of teachers is broadly stable over time.

Table 5: Baseline Calibration to French Data

(A) Empirical Targets

Moment	Data	Model
Share of agents who are teachers	0.098	0.098
Share of workers in new firms	0.203	0.222
Share of 2-layer firms	0.446	0.448
Share of 3-layer firms	0.228	0.228
Average firm size	93.48	92.32
Average size of 2-layer firms	87.23	87.07
Average size of 3-layer firms	214.00	215.54
Average (hourly) wage (€)	20.21	20.25
Share of workers (in firms) bound by minimum wage	0.047	0.047

(B) Calibrated Values of Parameters

Parameter	Value
Cost of knowledge acquisition, c	0.220
Time cost of problem communication, h	0.194
Output per problem solved, A ($\sim 10^2$ €/hour)	0.240
Fixed cost of entry, f_E	5.718
Fixed cost of production, f	5.698
Minimum wage $\bar{w}$, relative to average wage	0.914
2nd-node of α distribution ($\alpha = 50$)	$g_2 = -12.14,\ e^{g_2} = 5.341E-06$
3rd-node of α distribution ($\alpha = 500$)	$g_3 = -15.51,\ e^{g_3} = 1.837E-07$
4th-node of α distribution ($\alpha = 1000$)	$g_4 = -24.39,\ e^{g_4} = 2.556E-11$

Notes: In Panel (A), all moments are computed from the DADS, except the share of teachers, which is sourced from INSEE's Tableaux de l'Economie Française (2020 edition, data for 2018). e^{g_2}, e^{g_3} and e^{g_4} correspond to values of the density function $g(\alpha)$ for $\alpha = 50$, 500 and 1000, respectively.

except for g_1, which is fully determined conditional on the other three by the normalization condition ensuring that the integral of the distribution is 1.[57]

In the actual implementation of our optimization algorithm, we then solve for firm organization, output and profit across a grid of 250 values of $\alpha \in [1; 1000]$. We augment this grid with over 130000 values of $\alpha \in [1; 1000]$ but only approximate firms' outcomes (organization, output and profit) with a linear interpolation of the 250 previous values: bypassing actual numerical resolution increases computational speed. Overall, this strategy amounts to solving a discretized but highly accurate version of our initially continuous model.

Calibration criterion. The full list of parameters that we identify from the above list of nine data moments is the following: $\{c, h, A, f_E, f, \bar{w}, g_2, g_3, g_4\}$. There are as many moments as parameters, such that our procedure is just identified. We solve for the vector of parameter values that minimizes the following distance between the target moments and their values in the simulated model: :

$$Q = \sqrt{\sum_{j=1}^{9} \left(\frac{\hat{m}_j - m_j}{\min(\hat{m}_j, m_j)} \right)^2}$$

where m is the vector of target values for the moments, $\hat{m}$ their values in our numerical simulation, and j indexes the nine different moments used. For each vector of parameter values, we iterate over k (the

[57]In their quantitative exercise, Caliendo and Rossi-Hansberg (2012) only introduce one parameter, corresponding to the Pareto coefficient of the size distribution of firms (footnote 25, p. 1422). As we show below, this is too coarse to fit the French data in a detailed way, and it is also useful to relax the Pareto assumption in order to improve the fit with the calibration targets.

endogenous price of labor) to find the value that satisfies general equilibrium.

Calibration results. The results of the calibration are reported in Panel (B) of Table 5 while the values of the calibrated moments can be found in the final column of Panel (A). Note that, to ease comparison with the parameter values in Caliendo and Rossi-Hansberg (2012), all monetary values in the model correspond to 100s of Euros per hour. We first note that the model achieves an excellent fit for the data moments that are introduced in the calibration procedure: all of them are matched with a deviation that is inferior to 1% except for the share of workers in new firms. This indicates that the period of seven years that we use might be slightly too short (!) for the size distribution of new firms to converge to the firm size distribution of incumbents, but the fit remains very good since the distance between the model and the target is less than 10%.

In the French economy, training costs c are estimated at a relatively high value relative to h, such that $\frac{c}{\lambda} > h$. As explained in Section 4.1, this implies that some firms in the economy might find it optimal to adopt a "limit" behavior whereby they set the skill level of bottom, production workers at the lowest possible value. This behavior could emerge in any firm, and the smallest (1-layer) firms might in addition find it optimal to compress the size of their layer of production workers below the value that would saturate the manager's time constraint.[58] The value (0.24) obtained for the quantity of output produced per problem solved, A, is slightly superior to the value of average wage (0.20) that is used as a calibration target: this allows firms to earn a positive profit margin and cover their fixed costs of entry and production. The latter (f_E and f) are both estimated to be roughly equivalent to the wage bill of 5.7 agents. The results obtained for the distribution of α imply that the probability of a draw of 500 is only 3% of the probability of a draw of 50, while the probability of a draw of 1000 is only 0.0005% of the probability of a draw of 50. These values imply that the distribution of α has thinner tails than a Pareto distribution, the standard benchmark in similar settings. Lastly, the binding minimum wage is estimated to be 8.6% lower than the average wage paid out by firms. This parameter is mostly identified from the share of workers bound by the minimum wage and is quite sensitive to the value of the corresponding data moment.

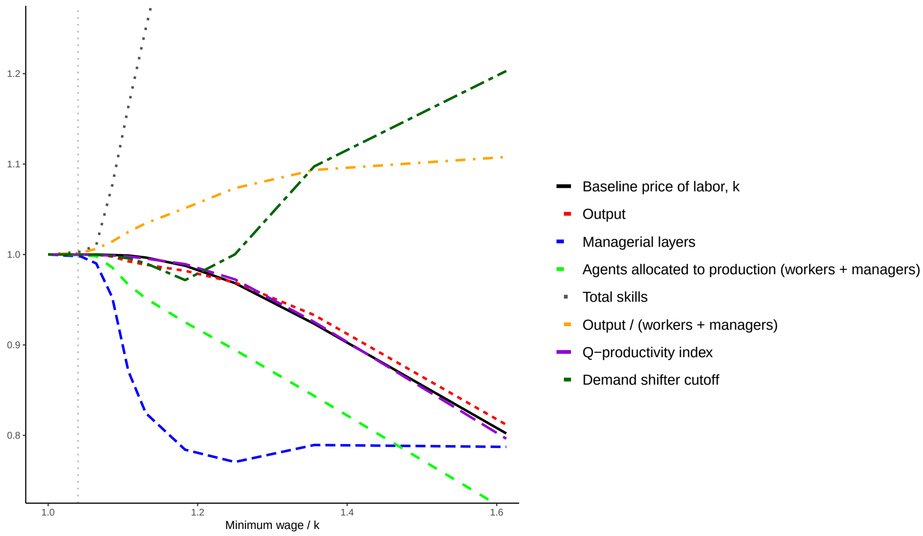
5.2 Simulation of Alternative Minimum Wage Scenarios

Considered scenarios. Having calibrated our model, we can simulate it to investigate the various dimensions in which minimum wages impact the economy in general equilibrium. To that end, we report in Figure 6 and Table 6 general equilibrium results obtained in six different scenarios: (1) the baseline value, corresponding to the initial calibration to the French data described in Section 5.1; (2) a no minimum wage scenario, which appears to numerically coincide with the case where the minimum wage is decreased by 4% as compared with the baseline scenario; (3) a scenario where the minimum wage is 4% higher than the baseline, which corresponds to the real increase experienced by GMR2 between 2003 and 2006 (a "medium" case scenario); (4) a scenario where the minimum wage is 8% higher than the baseline, which corresponds to the real increase experienced by GMR1 between 2003 and 2006; and lastly (5) and (6), two scenarios where the minimum wage is 16% and 24% higher than the baseline, respectively. These last two values are more extreme scenarios and correspond respectively to two and three times the real increase experienced by GMR1 between 2003 and 2006 (see Figure 1), but they remain in the range of empirically observed shocks. In particular, Bailey et al. (2021) analyze the effects of the 1966 Fair Labor Standards Act (FLSA), which implied an increase in the minimum wage by 28% in nominal terms, and by 23% in real terms. Harasztosi

[58] When minimum wage constraints become more stringent, as in our simulation exercises in Section 5.2 (see Table 6, Column (6)), such "limit" behaviors become increasingly likely because minimum wages implicitly set minimum constraints on the skills of production workers (Equation 21).

and Lindner (2019) document the impact of an increase in the Hungarian real minimum wage by around 60% in real terms.

Figure 6: General Equilibrium Margins of Adjustment to Minimum Wages



Notes: This figure presents simulations results for alternative scenarios where the minimum wage is either removed or set at increasingly high levels (relative to k, the baseline price of labor in general equilibrium). Parameter values are indicated in Table 5.

Model performance on non-targeted data moments. Scenarios (1) to (3) are symmetric in the sense that they characterize economies with minimum wages that are either 4% larger (scenario (2), where the minimum wage does not bind) or 4% higher (scenario (3)) than the baseline calibration (scenario (1)). This configuration is useful, as it allows us to calculate the average (both upwards and downwards) impact of a 4% *variation* in the minimum wage,[59] which can be compared with the regression results and thus provide an assessment of the prediction performance on non-targeted data moments. Indeed, the elasticities estimated in Section 3.4 also capture average effects of changes in minimum wages because they are identified on firms which experience both (differential) increases (GMR 1 to 3) and decreases (GMR 4 and 5) in their minimum wages. One point of caution is that in our setting, the regression results relate more to partial equilibrium than to general equilibrium comparative statics, although the different GMR groups are large enough to have the potential to affect general equilibrium prices (k), but our simulations show that price effects are in fact negligible across scenarios (1) to (3), so that in this range, partial and general equilibrium quantifications almost coincide numerically.

We first consider the impact of minimum wages on firm sizes, which corresponds to the outcome that is most robustly measured in the data. The estimated elasticity reached -0.154 in Table 2, which implies that firms exposed to a real increase in their minimum wage by 4% contracted their size (as measured by hours worked) by 0.6% on average, all else equal. This compares with our simulation results, which show that average firm size goes down from 92.32 workers in the baseline scenario to 90.36 workers in the "+4%" scenario, which is 2% smaller. Symmetrically, a decrease in the minimum wage by 4% (down to the "no minimum wage scenario") relative to the baseline would generate a negligible impact on firm size, such that on average the impact is around 1%. This number is remarkably similar to regression results given the simplicity (low dimensionality) of our framework.

[59] As a reminder, in contrast, our calibration procedure only relies on static data moments (Table 5).

Table 6: General Equilibrium Implications of Minimum Wages

(A) Aggregate Implications

Minimum wage, relative to k	(1) Baseline 1.04	(2) No MW 1.00	(3) + 4% 1.09	(4) + 8% 1.13	(5) + 16% 1.25	(6) + 24% 1.61
Net (real) wage, k (€/hour)	17.74	17.74	17.73	17.68	17.18	14.23
Av. labor cost/wage per worker, $k(cz+1)$	20.25	20.25	20.48	21.02	21.92	22.97
Share of firms bound by MW	0.19	0.00	0.41	0.57	0.93	1.00
Share of workers bound by MW	0.05	0.00	0.46	0.85	0.93	0.99
(as shares of total number of agents)	*Occupational choices and total skills:*					
Fixed-cost workers	0.26	0.26	0.26	0.27	0.26	0.25
Workers/managers	0.63	0.63	0.62	0.60	0.57	0.45
Firms (entrepreneurs), M	0.01	0.01	0.01	0.01	0.01	0.01
Teachers ($\propto$ total skills in economy)	0.10	0.10	0.11	0.12	0.17	0.29
	Production and productivity:					
Total output ($\sim$ €/hour)	13.28	13.28	13.25	13.13	12.87	10.78
Profits per firm ($\sim$ €/hour)	5.68	5.68	5.66	5.57	5.50	5.96
Average firm size (in jobs)	92.32	92.38	90.36	84.94	82.93	86.18
Average revenue per worker (€/hour)	28.37	28.32	28.85	29.63	29.90	30.42
Q-productivity index, $q/C_L(q,1)$ (€/hour)	17.82	17.82	17.80	17.74	17.33	14.19
Production cutoff, $\overline{\alpha}$	51.97	51.98	51.89	51.47	51.98	62.52

(B) Firms' Organizations

Minimum wage, relative to k	(1) Baseline 1.04	(2) No MW 1.00	(3) + 4% 1.09	(4) + 8% 1.13	(5) + 16% 1.25	(6) + 24% 1.61
(as shares of total number of firms)	*Structure of firm population:*					
1-layer firms	0.32	0.32	0.36	0.43	0.53	0.50
1-layer firms with $n_1^0 < e^{\lambda \overline{z}_0}/h$	0.00	0.00	0.00	0.00	0.00	0.45
2-layer firms	0.45	0.45	0.46	0.57	0.47	0.50
3-layer firms	0.23	0.23	0.18	0.00	0.00	0.00
(Share of agents)	*Hierarchical organizations:*					
Fraction of agents in $l = 0$	0.55	0.55	0.55	0.55	0.54	0.45
Fraction of agents in $l = 1$	0.07	0.07	0.07	0.05	0.03	0.00
Fraction of agents in $l = 2$	0.01	0.01	0.00	0.00	0.00	0.00
(Share bound by MW)	*Minimum Wage constraint at the firm level:*					
Firms with $L = 1$	0.00	0.00	0.00	0.00	0.88	1.00
Firms with $L = 2$	0.43	0.00	0.50	1.00	1.00	1.00
Firms with $L = 3$	0.00	0.00	1.00	1.00	-	-
(Share bound by MW)	*Minimum Wage constraint at the worker level:*					
Firms with $L = 1$	0.00	0.00	0.00	0.00	0.09	0.16
Firms with $L = 2$	0.05	0.00	0.08	0.84	0.84	0.83
Firms with $L = 3$	0.00	0.00	0.38	0.01	0.00	0.00
(Relative to k)	*Wages of entrepreneurs*					
In 1-layer firms	1.95	1.95	1.95	1.96	2.02	2.12
In 2-layer firms	2.27	2.27	2.31	2.41	2.49	2.53
In 3-layer firms	2.54	2.54	2.55	2.58	-	-
(Relative to k)	*Wages of workers and managers:*					
In $l = 0$	1.08	1.08	1.09	1.13	1.25	1.61
In $l = 1$	1.58	1.57	1.62	1.73	1.75	1.73
In $l = 2$	1.79	1.79	1.79	1.83	-	-

Notes: This table presents simulations of alternative scenarios where the minimum wage is either removed (Column (2)), or increased by 4%, 8%, 16% or 24%, respectively, in Columns (3) to (6), relative to the baseline calibration presented in Column (1). The baseline scenario is described in detail in Section 5.1. Among considered indicators, total output can be interpreted as per agent and per hour, in €: $P.\int M\mu(\alpha)q(\alpha)\mathrm{d}\alpha$, since P is normalized to 1. Average output per worker (and per hour) corresponds to $\frac{P.\int M\mu(\alpha)q(\alpha)\mathrm{d}\alpha}{\#\text{ Workers} + \#\text{ Managers}}$. Average revenue per worker (and per hour) corresponds to $\frac{p(\alpha).\int M\mu(\alpha)q(\alpha)\mathrm{d}\alpha}{\#\text{ Workers} + \#\text{ Managers}}$.

Turning to indices of productivity, which correspond to our dimension of main interest, our regression results in Table 4 imply that a 4% real increase in the minimum wage is associated with an increase in labor productivity by roughly 0.8 to 1.1% depending on the way the latter is measured empirically. This number is to be compared to the simulated values: the latter predict that a similar shock, measured from either the "no minimum wage" scenario or from the baseline scenario, will induce an increase in output per worker by 0.7% and in revenue per worker by 0.9%, which is extremely close to the regression results.

Finally, it is also possible to evaluate the model's predictions in terms of different organizational dimensions. The simulation results show that the impact of a 4% shock to the minimum wage, from either the "no minimum wage" scenario or from the baseline scenario, is associated with an average 2.4% decrease in the number of hierarchical levels, a 4.6% increase in skills and a 0.6% increase in wages. For a shock of the same magnitude, the regression results show a 0.4% decrease in the number of hierarchical levels, despite the limitations of our empirical indicator (which is limited to just three managerial levels), a 6.6% increase in the probability of training and a 1.4% increase in wages.

Overall, all signs and orders of magnitudes are preserved, and the performance of the model is even extremely good for the main indicator of interest: productivity.

General equilibrium impact of Minimum Wages on aggregate output. We now turn to the characterization of the general equilibrium impact of minimum wages. Panel (A) of Table 6 first provides an overview of their impact on aggregate outcomes. The first line shows how the baseline price (of labor), k, adjusts to increasingly stringent minimum wage constraints. Our simulations show that its downwards adjustment is extremely limited in scenarios (1) to (4), and that its decrease only becomes steeper in scenarios (5) and (6), as also shown graphically in Figure 6.

Similarly, despite the fact that as many as 19% of firms and 5% of workers are directly bound by the minimum wage constraint in our baseline scenario (1), the aggregate output loss relative to the "no minimum wage" scenario (2) is negligible up to the second digit. As minimum wages further increase, we obtain that the output loss remains negligible in scenario (3), reaches 1% in scenario (4), 3% in scenario (5), and ultimately 19% in scenario (6). Figure 6 confirms that the output cost of minimum wage constraints remains extremely limited for values attaining up to 1.2 times the baseline price of labor, k.

Organizational adjustments. While the impact of a moderate tightening of the minimum wage constraint on output is limited, the impact on firms' organizations is not. In fact, it is precisely these organizational adjustments that make it possible to limit the fall in output, thanks to the productivity increases they induce. Panel (B) of Table 6 and Figure 6 document the organizational adjustments in full detail. A first channel consists in the massive increase in skills, as shown in Figure 6. This adjustment necessitates significant reallocations of agents from production or managerial occupations to the education sector, which, in our model, captures the total amount of human capital embodied in workers. Wages increase accordingly at all layers, as documented in Panel (B) of Table 6. The second channel lies in the fast flattening of firms. The share of complex firms decreases quickly as minimum wage constraints become tighter. In the last scenario of Table 6, most 1-layer firms even find it optimal to contract their layer of production workers to a level which does not even saturate the time constraint of their manager. The overall decrease in average firm size results from composition effects: firms decrease their number of layers, while average firm sizes within classes defined in terms of number of hierarchical levels may increase or decrease.

Figure 6 shows that a particularly striking feature of the simulation results is that the previous firm-level organizational decisions adjust extremely quickly, even in scenarios where the baseline cost of labor, k, only experiences small changes. For example, k only adjusts by less than 0.1% in the "+4%" scenario relative to the baseline (Panel (A) of Table 6), but the shares of 1, 2 or 3-layer firms adjust by 12, 2 and 21%, respectively (Panel (B) of Table 6). This is driven by the fact, first, that the minimum wage regulatory constraint itself (Equation 21) directly affects organizations and skills at the bottom, and second, that the internal training decisions (in terms of z) have multiplier effects across hierarchies, thus amplifying the impact of small changes at the bottom, including in terms of which organizational form ultimately becomes optimal for a given demand draw, α.

Implications for aggregate productivity. Our simulations show that firms' large organizational responses have a significant productivity impact. It is precisely this mechanism that mitigates the impact of (moderate) minimum wages on aggregate output. To be precise, in our set-up, aggregate productivity can be affected by three different channels. Firstly, the "pure" selection mechanism implies that firms benefiting from more favorable idiosyncratic demand conditions (captured by α) are more likely to operate, as a consequence of Equations 18 and 19. This mechanism is identical to that of Melitz (2003).[60] Secondly, among active firms, firms benefiting from more favorable demand conditions tend to capture larger market shares: this "reallocation" channel affects aggregate productivity by reweighting the contribution of each individual firm to the aggregate. Thirdly, the endogenous response of firms' organizations affects production costs and hence firm-level productivity. This third channel may counteract the first two. Our simulations allow us to assess the comparative importance of all of these mechanisms in response to minimum wage constraints.

Panel (A) of Table 6 provided first results in this respect. The magnitude of the "pure" selection channel is captured by the general equilibrium adjustment of $\overline{\alpha}$, the production cut-off. In scenarios (1) to (5), this parameter is almost stable, implying that the selection margin plays only a very limited role. In addition, the impact of minimum wages appears to be non-monotonic: scenarios with moderate minimum wage constraints (scenarios (1) to (4)) actually show an excess of entries, while the selection effect only emerges in regimes with very stringent minimum wage constraints (scenarios (5) and (6)). On the other hand, the magnitude of the organizational channel can be assessed using various indicators of average productivity that are also reported in Table 6. The $Q-$productivity index, which is a pure indicator of productive efficiency, always deteriorates since minimum wages always prevent companies from implementing their best strategies. Here again, however, the decline is very limited for moderate minimum wage constraints, but worsen as they become more severe. In contrast, indicators of revenue productivity, such as average revenue per worker, rise steadily across scenarios with increasingly stringent minimum wage constraints. We obtain that average increases in revenue per worker reaches 1.9% in scenario (3) ("+4%"), 4.5% in scenario (4) ("+8%") and 5.4% in scenario (5) ("+16%"), and ultimately 7.2% in the "+24%" scenario.

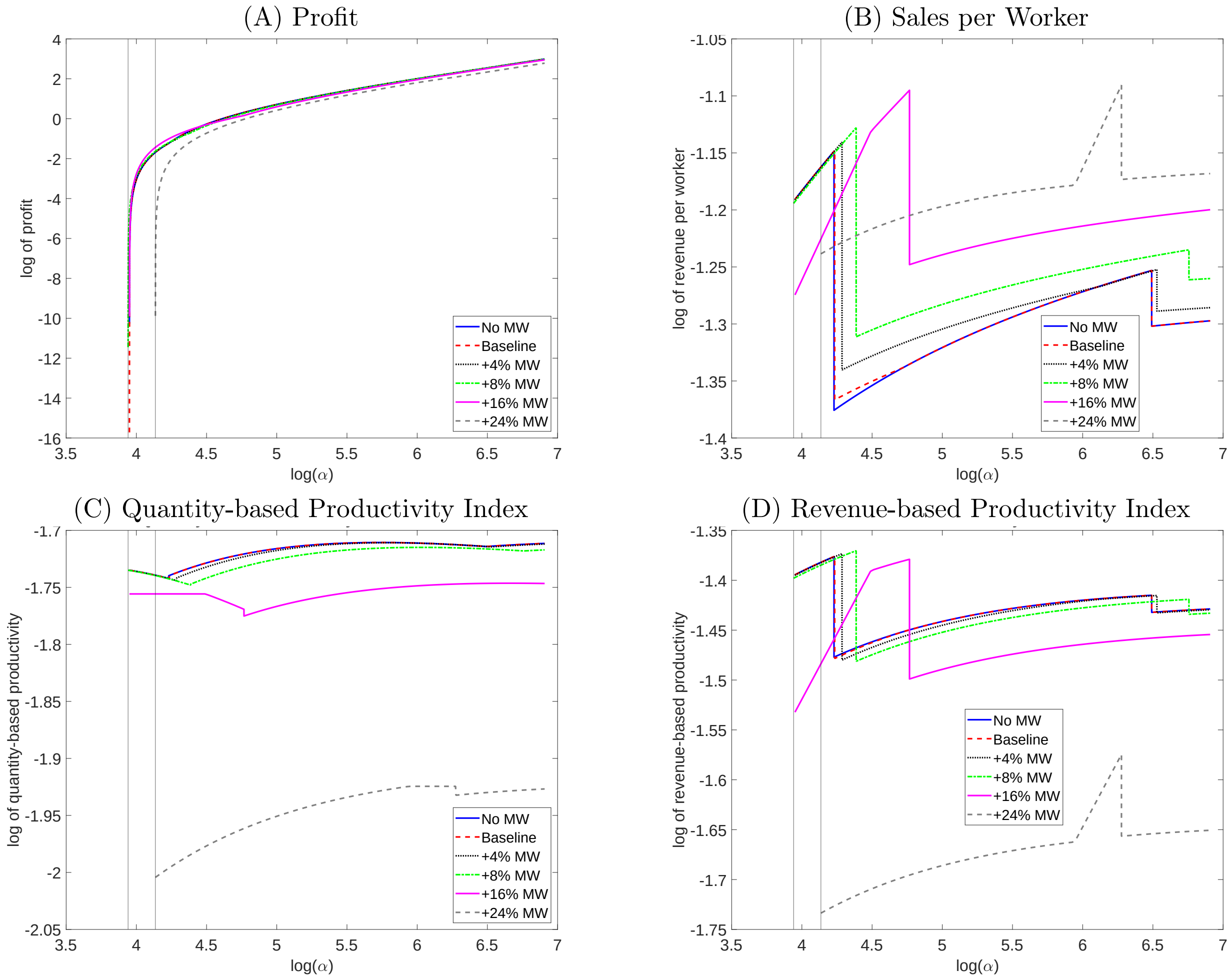
To better understand the firm-level adjustments that underlie the above aggregate patterns, it is interesting to analyze the relationships between α, the idiosyncratic demand shifter, and the different firm-level indicators of performance and productivity. Figure 7 plots profits, sales per worker as well as the Quantity- and Revenue-productivity indices as a function of α.[61] In terms of profits (Panel (A)), unsurprisingly, the impact of minimum wage constraints is always negative whatever α, as in Machin et al. (2003), but our simulations suggest that this adverse impact is limited in magnitude for moderate minimum wage constraints. Furthermore, the firms facing the worst market conditions in terms of α only start being significantly selected out of the market for extremely stringent minimum wage constraints ("+24%", scenario (6)). The same analysis applies to the Q-productivity index in Panel (C), but the situation begins to deteriorate significantly

[60] We formulate exogenous heterogeneity at the firm level in terms of demand heterogeneity rather than marginal cost, in order to make it clear that everything referred to as "productivity" in our paper corresponds only to the endogenous component that is determined by firms' organizational strategies (and not to ex ante exogenous differences), but this modeling choice is fully equivalent to the standard Melitz (2003) formulation.

[61] The revenue-productivity index simply corresponds to to the $Q-$ productivity index multiplied by firm-level prices (Caliendo and Rossi-Hansberg, 2012).

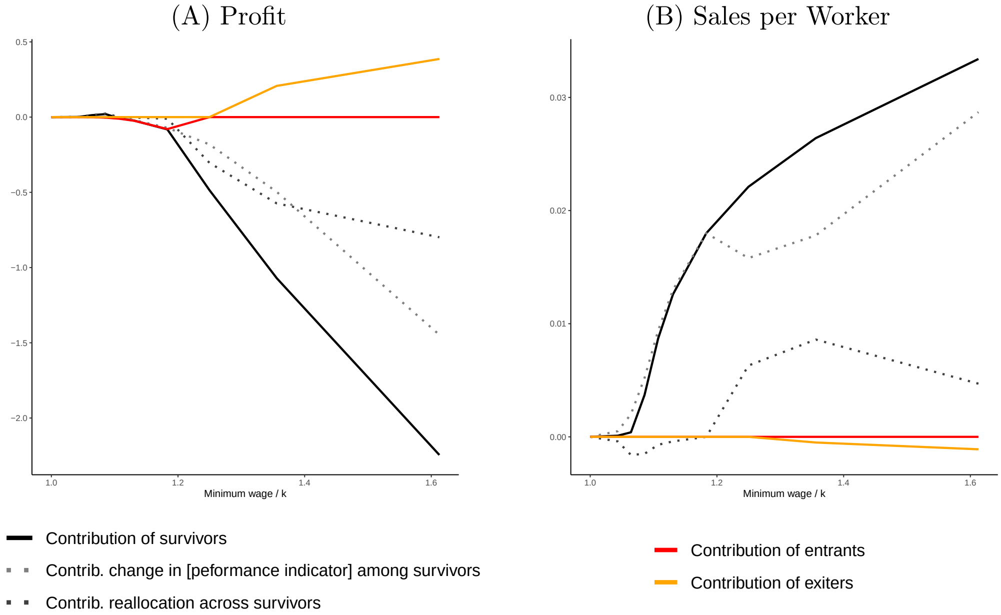
earlier in terms of this indicator, as early as scenario (5) ("+16%" scenario). In terms of sales per worker or *R*evenue productivity, the impact of minimum wages is in contrast often positive for low values of α (which corresponds to the most populated section of the support of the distribution), although this impact is strongly heterogeneous and non-monotonic. Table 6 shows that the resulting overall impact is positive in all scenarios in the French context, but the complexity of the underlying patterns might explain why previous empirical estimates in the literature have led to heterogeneous results depending on the context and magnitude of the minimum wage shock considered (Rizov et al., 2016; Riley and Rosazza Bondibene, 2017). The obtained patterns in terms of these indicators are heavily affected by the endogenous responses of firms' organizations, which tend to counteract unfavorable demand conditions (low α). They inherit the non-monotonicities and non-differentiabilities of the cost function plotted in Panel (B) of Figure 5. In particular, and in contrast to Melitz (2003), firms facing the least favorable demand conditions are not always the least productive, because they try and counterbalance them by remaining small and endogenously adjusting their organizations in order to achieve higher productivity, although the cost of this strategy (in terms of skill upgrading) ultimately becomes too large and affects profits heavily.

Figure 7: Firms' Selection and Performance as a Function of α



Notes: This figure is derived from simulations of the calibrated model, in different scenarios in terms of the minimum wage. The parameters are calibrated to the following values: $\lambda = 1$, $c = 0.220$, $h = 0.194$, $A = 0.240$. In Panel (C), the index of "Quantity-based Productivity" corresponds to $q/C_L(q, 1)$, as in Equation 14. In Panel (D), the index of "Revenue-based Productivity" corresponds to $pq/C_L(q, 1)$.

Figure 8: Decomposition of the Impact of Minimum Wages on Average Profits & Labor Productivity
(relative to the "No Minimum Wage" economy)



Notes: In each Panel, the decomposition (Equation 22) relies on output weights, but virtually identical results are obtained with employment weights. Panel (A) plots the contribution of the different terms to average profits in economies characterized by different tightness of the minimum wage constraints. Panel (B) replicates the same exercise for average sales per worker.

Lastly, Figure 8 proposes a more comprehensive quantification of the three main channels that determine aggregate productivity: pure selection, reallocation of market shares, and the endogenous organizational channel. To that end, we simply rely on the decomposition initially proposed by Melitz and Polanec (2015)[62], which allows us to break-down the changes in average profit or average revenue per worker in economies with minimum wages, relative to the "no minimum wage" scenario:

$$\Delta\Phi = \Delta\tilde{\phi}_S + \Delta COV_S + s_E\left(\Phi_E - \Phi_S\right) + s_X\left(\Phi_S^{no\ MW} - \Phi_X^{no\ MW}\right) \tag{22}$$

In this equation, Φ denotes the aggregate indicator of performance that is considered: in Figure 8, the latter corresponds alternatively to average profits (Panel (A)) or to average revenue per worker (Panel (B)). $\tilde{\phi}_S$ is the unweighted average of this same indicator across firms which are active under both the "no minimum wage" and the considered scenario, ie. "survivors", S. The difference between such averages computed in scenarios with *versus* without minimum wages thus captures the shift in the firm-level distribution of the considered indicator: when Φ measures productivity, this term thus quantifies the endogenous organizational channel. ΔCOV_S corresponds to the change in the covariance between market share and performance among firms which are consistently active under both scenarios and captures market share reallocations.[63] Lastly, the pure selection channel is captured by the last two terms in Equation 22. To be precise, $s_E\left(\Phi_E - \Phi_S\right)$ captures the contribution of entrants (E), in scenarios where the $\overline{\alpha}$ threshold is lower than in the "no minimum wage" benchmark: this contribution is positive whenever the average performance of entrants is higher than that of survivors. Lastly, the term $s_X\left(\Phi_S^{no\ MW} - \Phi_X^{no\ MW}\right)$ captures the contribution of exiters (X), in the opposite case where the $\overline{\alpha}$ threshold increases: the contribution of exits is positive whenever the average performance of exiters is lower than that of survivors.

[62] See Autor et al. (2020) for another important (and recent) application of the same decomposition methodology.
[63] Following Dustmann et al. (2021), we expect this term to be positive and economically significant.

Figure 8 displays the decomposition that is obtained for aggregate profits in Panel (A). They monotonically decrease when minimum wage constraints become more stringent, and the main contributions to this pattern are the within-firm decreases and the reallocation across firms that are consistently active across scenarios ("survivors"),[64] with the first becoming dominant for the highest levels of minimum wage constraints. Excess entry also contributes to depressing average profit for moderate minimum wage constraints, while exits tend to dampen aggregate profit losses for higher minimum wage constraints.

The patterns that are obtained in Panel (B) in terms of the indicator of (revenue) productivity, ie. sales per worker, are quite different. First, sales per worker monotonically increase when minimum wage constraints become more stringent, and the main contribution to aggregate productivity increases is, by far, the within-firm increase generated by firms' organizational strategies. The reallocation effects become significant for tighter minimum wage constraints, but the magnitude of this second channel never exceeds half of the previous. Lastly, the contribution the pure selection channel (entries and exits) is only marginal.[65]

Comparison with a model specification without organizations ("unproductive managers"). The previous developments suggest that firms' productivity responses to increases in the minimum wages are largely driven by their endogenous organizational responses. In Appendix H, we propose an even more direct strategy for assessing the importance of this channel – and more precisely within this channel, of the contribution of the responses of hierarchical organizations. There, we specify a version of the model in which workers' skills remain useful to production, but (hierarchical) organizations do not. More precisely, we assume there that managers are still useful to "bundle" teams of workers, but that this task exhausts their entire time constraint such that they are unable to contribute to actual problem solving. In such a setting, no firm finds it optimal to increase the number of managerial layers beyond $L = 1$, and managers do not need skills, but all firms still have to optimize their size in terms of number of production workers, as well as the level of skills of the latter. Appendix H proposes a calibration of such a framework to the French data: the comparison of its quantitative predictions to our main results provides a precise quantification of the mitigation impact of firms' organizations to minimum wage constraints.

We obtain that firms find it optimal to train their bottom production workers (and pay them accordingly) even in absence of a minimum wage. As a consequence, the minimum wage only starts to be binding for values above 1.12 k, while it was binding for at least some workers for values slightly exceeding k in our main calibration. From this point, skills adjust even more quickly in the set-up without organizations than in the baseline set-up, since this is the only possible mitigation channel, but this only translates into mediocre increases in output per worker. As a consequence, the $Q-$productivity index and aggregate output both decrease four times faster than in the baseline set-up as the minimum wage constraint ($\overline{w}/k$) becomes more stringent. In other terms, the pure plasticity of firms' organizations dampens the aggregate output cost of minimum wage constraints by dividing it by a factor of four, compared to a set-up where organizations are restricted to be fully rigid.

Further robustness checks. In the main specification of our model, the labor market is always in equilibrium. However, we investigate in Appendix G whether the employment cost of minimum wages could be amplified in a specification where we introduce job search costs, such that unemployment can emerge. This augmented version of the model thus corresponds to an economy with a more rigid labor market, a feature often attributed to the French case. In the calibration procedure, we target an unemployment rate as an

[64] These results are in particular consistent with Dustmann et al. (2021).

[65] In addition, we documented above that firms facing the least favorable demand conditions counteracted them proactively via their organizations, which implies that they are not the least productive in terms of revenue per worker (Panel B of Figure 8). This implies in particular that "exits" do not always have a positive contribution to aggregate productivity in our set-up.

additional moment while otherwise using the same baseline moments as in our central calibration. The main take-away of this experiment is that introducing a richer set of occupational "options" (ie. an unemployment state) available to workers does not significantly alter the quantitative predictions of the model. In particular, the unemployment margin absorbs very little of the minimum wage constraint, while productivity remains the main channel of adjustment. As a consequence, the overall employment impact is simulated to be very limited.[66]

5.3 Simulation of Alternative Technological Shock Scenarios

Interaction between technologies and Minimum Wages. We now examine the extent to which the impact of the minimum wage described in the previous section is modulated by different types of technologies. Our framework is particularly suitable and rich, since four different parameters capture the technological environment of the economy. Firstly, parameter A captures output per problem solved and thus corresponds to an overall worker productivity parameter. Secondly, studying a problem takes up all the time allocated to production workers, but "only" h extra units of time for managers when they are called upon. Thus, the parameter h captures a second technological dimension, directly determining "managerial costs": Garicano and Rossi-Hansberg (2006) relate this dimension to firms' internal communication technologies. This dimension is of particular relevance, since during the Covid-19 pandemic most firms were forced to adopt or develop such IT technologies supporting remote work in order to remain operational during lockdowns. Quantifying the long term implications of this unique technological shock remains to date largely an open question (Barrero et al., 2021; Gibbs et al., 2023). Thirdly, parameter c represents the cost at which production workers and managers achieve higher levels of problem-solving ability. In practice, this parameter is likely to be determined by information technologies (Garicano and Rossi-Hansberg, 2006; Garicano et al., 2016). Finally, parameter λ determines the overall distribution of production problem difficulty. Our intuition is that some AI-based technologies, such as LLM, might well affect this dimension of production.[67]

In the simulations which follow, we consider economy-wide technology shocks that are of the same type as those considered in Garicano and Rossi-Hansberg (2006) or Caliendo and Rossi-Hansberg (2012), i.e. that affect the technology parameters introduced in the model (A, c, h and λ) of all firms homogeneously and

[66] This specification of the model delivers an aggregate employment elasticity which can be compared with previous results obtained in the empirical literature. It appears to be highly non-linear and in particular, close to 0 for limited increases in the minimum wage, since aggregate employment is barely affected over this range in our simulations. This result is consistent with many estimates of the empirical literature, from Card and Krueger (1994) to Bailey et al. (2021). For larger increases in the minimum wage, the employment elasticity is less precisely estimated (probably because of our numerical approximations). Contrasting the "No MW" with the "+24%" scenarios, we estimate that this elasticity lies in the $[-2.3; -0.9]$ interval. This compares to the synthesis of the literature's estimates displayed in Figure A6 of Harasztosi and Lindner (2019) and Figure 6 of Bailey et al. (2021), which are most often negative and range roughly in the $[-2; +2]$ interval.

[67] A number of recent contributions discuss the potential impact of these "AI" technologies, but no definitive answer has yet been found, as it may well be uncertain, heterogeneous and endogenous to innovators' decisions (Acemoglu and Johnson, 2023; Rotman, 2023). The first case studies however suggest that the upcoming wave of AI based technologies might well augment the capabilities of some workers (Mullainathan and Obermeyer, 2021) by correcting their errors, which in our model would be captured by a drop in c (like the standard information technologies that were considered in Garicano et al., 2006) or an increase in λ. The experiment conducted in Kleinberg et al. (2017), where ML tools are found to have the potential to help judges to reduce crime by up to 24.7% with no change in jailing rates, or reduce jailing rates by up to 41.9% with no increase in crime rates, could also be interpreted as an increase in λ, as judges seem to make fewer errors in their jailing decisions. On the other hand, Noy and Zhang (2023) interpret the results of their experiments as "ChatGPT substantially [raising] average productivity", which in our set-up could be driven either by an increase in A or λ. Results and interpretations are similar in Brynjolfsson et al. (2023) and Peng et al. (2023). Note that one additional concern raised in this literature is the amount of labor replacement that would be implied by these technologies if they are used as automation devices (Eloundou et al., 2023; Rotman, 2023; Svanberg et al., 2024). Recent research by Acemoglu (2024) suggests that the figures may be lower than originally thought, with average labor cost savings of "only" 15% (over 10 years) and economy-wide TFP gains below 1%, also over 10 years. This aspect remains out of the scope of our paper: in particular, we leave for now the challenge of introducing additional production factors (capital or intermediate inputs) and a task-based approach (Aaronson and Phelan, 2019) into the framework for future research and focus here solely on the interaction between technologies and firms' endogenous work organization responses (Garicano and Rossi-Hansberg, 2015).

simultaneously.[68] In terms of the magnitude of the shocks that are to be considered, the literature offers little guidance so far. Ideally, our experiment would require information about the (social) costs associated with variations of the four technology parameters but unfortunately, to our knowledge, this information is not available in the literature. We thus propose simply to consider the figures that have been estimated for recent, empirically relevant episodes of technological change. Bloom et al. (2012) show for example that US multinationals have on average a productivity premium of 3.2 percent over non-US multinationals, which the authors associate with greater (overall) ICT intensity and associated optimal organizational adjustments. Accordingly, we successively consider technology shocks in terms of each of our four model parameters and calibrate their values in order to generate an economy-wide production premium of 3.2 per cent (i.e. a 3.2 per cent increase in aggregate output) in the "no minimum wage" scenario:

- Baseline calibration: $\lambda = 1,\ h = 0.194,\ c = 0.220,\ A = 0.240$.
- Scenario 1 [varying λ, production problem difficulty]: $\lambda = 1.319,\ h = 0.194,\ c = 0.220,\ A = 0.240$.
- Scenario 2 [varying h, communication technologies]: $\lambda = 1,\ h = 0.167,\ c = 0.220,\ A = 0.240$.
- Scenario 3 [varying c, information technologies]: $\lambda = 1,\ h = 0.194,\ c = 0.167,\ A = 0.240$.
- Scenario 4 [varying A, output per problem solved]: $\lambda = 1,\ h = 0.194,\ c = 0.220,\ A = 0.248$.

In each of these scenarios, we replicate the same analyses as in Section 5.2 and consider different magnitudes of the minimum wage constraint. All results are reported in Figure 9, where the x-axis normalizes the value of the minimum wage relative to k, the baseline price of labor.

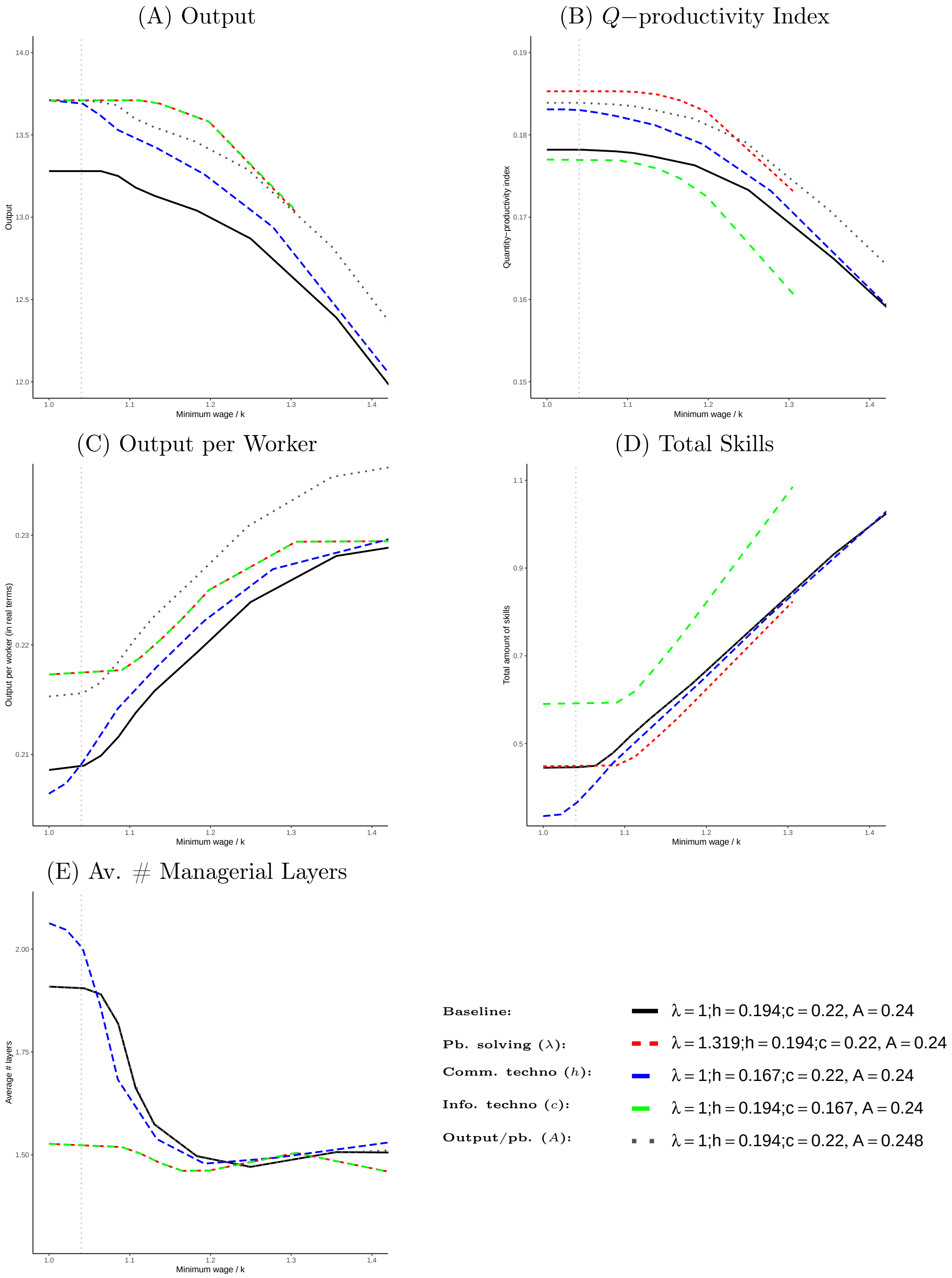
Panel (A) first documents the impact of the minimum wage constraint in terms of the broadest, economy-wide performance indicator: aggregate output. It shows strikingly that the impact of minimum wages is highly differentiated across technological environments. Although output is systematically higher than in the reference scenario in all cases, the extent of the decline when minimum wages are introduced and become increasingly stringent varies greatly from one technological environment to another.

Firstly, in the scenario where the aggregate productivity parameter A is increased, the resulting curve is translated upwards but the overall profile is identical (homothetic) to the base scenario. As a result, the relative impact of minimum wages is unaltered. This "neutrality" result is due to the fact that this parameter affects all agents homogeneously, such that the structure of the economy remains fundamentally identical to the baseline scenario, whatever the magnitude of the minimum wage constraint. In particular, the curves describing optimal firms' organizations and total skills are unaltered in Panels (D) and (E). As previously, we obtain that firms' optimal response to minimum wages involves hiring fewer but better trained workers (Panel (D)) organized in flatter organizations to save on managerial "fixed" costs (Panel (E)), such that output per worker increases (Panel (C)), by a larger amount than in the baseline scenario thanks to the shock on A. More precisely, the curves depicting output per worker in Panel (C) and the Q-productivity index in Panel (B) are simply translated from the baseline scenario, reflecting the homogeneous productivity premium affecting the entire economy.

Turning to the cases of communication technologies (affecting parameter h), we show that the situation is much less favorable. In Panel (A) of Figure 9, the adverse impact of minimum wages on aggregate output is indeed amplified relative to the baseline, as the slope of the corresponding curve is steeper. Total skills adjust relatively more than in the reference scenario (Panel (D)), while business organizations flatten more quickly (Panel (E)). As we showed earlier, it is this organizational margin that is very effective in dampening

[68] This type of comparative statics exercise is relevant, in particular, for thinking about the potential impact of new AI technologies such as ChatGPT (as an example of generative AI model), which have been introduced recently and at low cost (at least for now, and for the current level of service), and about whether their impact might be modulated by labor market institutions such as the minimum wage.

Figure 9: Impact of Minimum Wage Across Different Technological Scenarios



Notes: Across all Panels of this Figure, we consider different scenarios where the minimum wage ranges between k and $1.4k$ in general equilibrium. Thus, each point on the X-axis corresponds to a distinct general equilibrium computation. All measurement units are identical to those indicated in Table 6. Five different technological scenarios are considered, with the black curve corresponding to the baseline calibration of Sections 5.1 and 5.2 ($\lambda = 1,\ h = 0.194,\ c = 0.220,\ A = 0.24$). All other scenarios correspond to more efficient technologies, which are normalized so as to generate an economy-wide production premium of 3.2% in the "no Minimum Wage" case: a higher λ (1.319), corresponding to distributions of production problems which are easier to solve, a lower h (0.167), corresponding to lower managerial costs, a lower c (0.167), corresponding to lower training costs, and a higher A (0.248), corresponding to higher overall productivity for all workers. Panel (A) corresponds to total aggregate output, while all other Panels correspond to averages across firms. In Panel (B), the $Q-$productivity index corresponds to $q/C_L(q,1)$, as in Equation 14 and Table 6. In Panel (D), "Total skills" correspond to the total amount of knowledge acquired by workers and managers, $(\sum n_L^l z_L^l)$. They also match the total flow of knowledge created by the mass of teachers who are active in equilibrium, $\frac{\#\text{ Teachers}}{c}$.

the costs associated with minimum wage regulatory constraints. However, in this scenario, this strategy is socially more costly, since it unfortunately prevents the economy from reaping the gains associated with a lower h. Indeed, the gains associated with more efficient communication technologies (lower h) require sufficiently verticalized business organizations to be effective. Minimum wages induce firms to flatten quickly, such that the actual potential of efficiency gains associated with the lower value of parameter h becomes smaller. This is the reason why the minimum wage constraint is particularly costly in terms of its impact on the Q-productivity index in Panel (B).

Lastly, Figure 9 shows that the cases of economies benefiting from cheaper knowledge (lower c) or less dispersed production problems (higher λ) are much more favorable, as the relative cost of minimum wages tends to be dampened.[69] In terms of aggregate output (Panel (A)), the profile of the curves corresponding to each of these scenarios actually coincide. These economies are also equivalent in terms of output per worker (Panel (C)) and firms' organizations (Panel (E)),[70] but they differ in terms of the Q-productivity index (Panel (B)) and in terms of training decisions (Panel (D)).[71] Unsurprisingly, the optimal investment in knowledge is lower in economies where production problems are easier to solve (higher λ) thus requiring fewer skills. Conversely, these investments are higher when knowledge is cheaper to acquire. In these scenarios, the lower cost of the minimum wage constraint is simply driven by the fact that more efficient technologies in terms of c or λ make it easier to render bottom, production workers more productive - and meet the minimum wage Condition (21) of firms' cost minimization programs.

Varying the magnitude of technological shocks. In a last experiment in Figure 10, we check how the previous results evolve for different magnitudes of the technological shocks, focusing on impacts on aggregate output. Panel (A) simply replicates Figure 9 for comparison. In Panel (B), the considered shock is doubled and leads to an economy-wide productivity premium of 6.4% (rather than 3.2%) in the case of "no minimum wage", while in Panel (C), the considered shock is tripled (economy-wide productivity premium of 9.6%). The precise values of the technological parameters that are required to achieve these two targets are indicated in the notes to Figure 10.

As previously, when the aggregate productivity parameter A is increased, the resulting curves are translated upwards but remain homothetic to the base scenario in all Panels. Thus, as previously, the relative cost of the minimum wage constraint for the economy remains unaltered relative to our base scenario of Section 5.1, whatever the magnitude of the technological shock in terms of A that is considered. All economies characterized by curves steeper than this benchmark should be interpreted as presenting a higher relative cost of the minimum wage constraint, while conversely, those associated with flatter curves correspond to lower relative costs.

With this reference in mind, we obtain that the relative cost of minimum wages tends to be higher in economies characterized by efficient communication technologies (low h) than in economies featuring higher overall productivity (A), and that this disadvantage is amplified when the magnitude of the considered technological shocks is larger. For instance, a minimum wage constraint set at 20% above the baseline price of labor generates an output drop by 2.2% in economies with higher A than the baseline, whatever the magnitude of the technological shock. In contrast, in economies benefiting from lower h, we obtain that output drops by 3.6% relative to the "no minimum-wage" economy in the case of the small technological shock

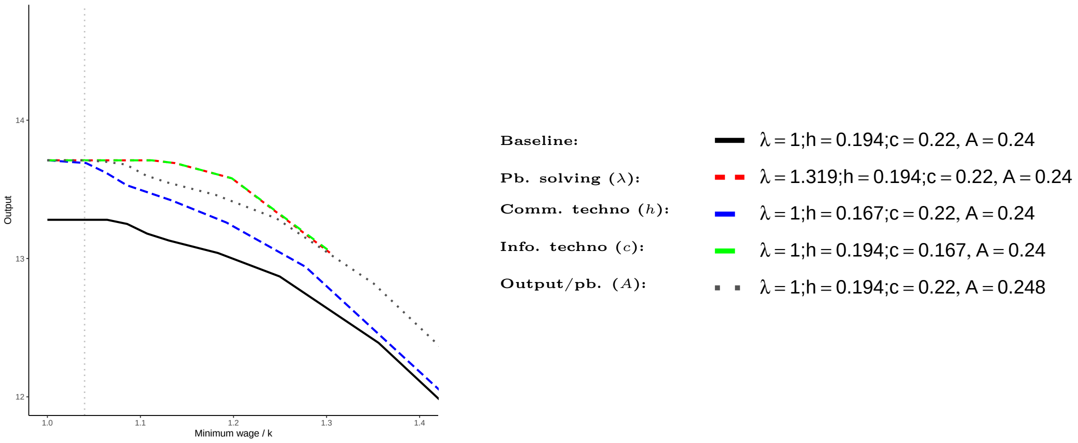
[69] This is partly driven by the fact that the economy switches from a technological environment where $\frac{c}{\lambda} > h$ to an environment where $\frac{c}{\lambda} < h$. As discussed in Section 4, the second regime prevents the emergence of inefficient regimes in which managerial time in 1-layer firms is (optimally) wasted. This aspect contributes to mitigating the minimum wage constraint.

[70] This result was already present in Caliendo and Rossi-Hansberg (2012), who present their comparative static exercises of interest in terms of $\frac{c}{\lambda}$ directly rather than separately in terms of c and λ.

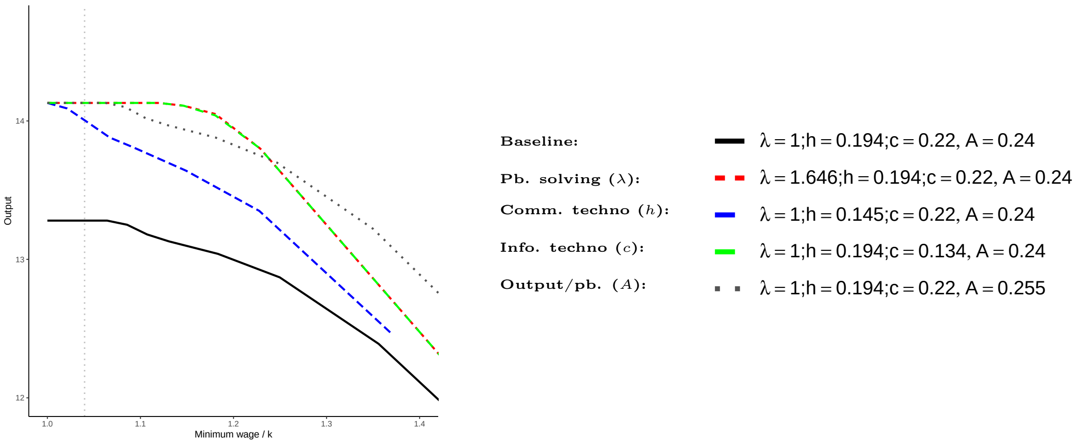
[71] Note that the curves are simply translated, with the difference in skill responses exactly off-setting the difference in terms of Q-productivity.

Figure 10: Impact of Minimum Wage For Higher Technological Shocks

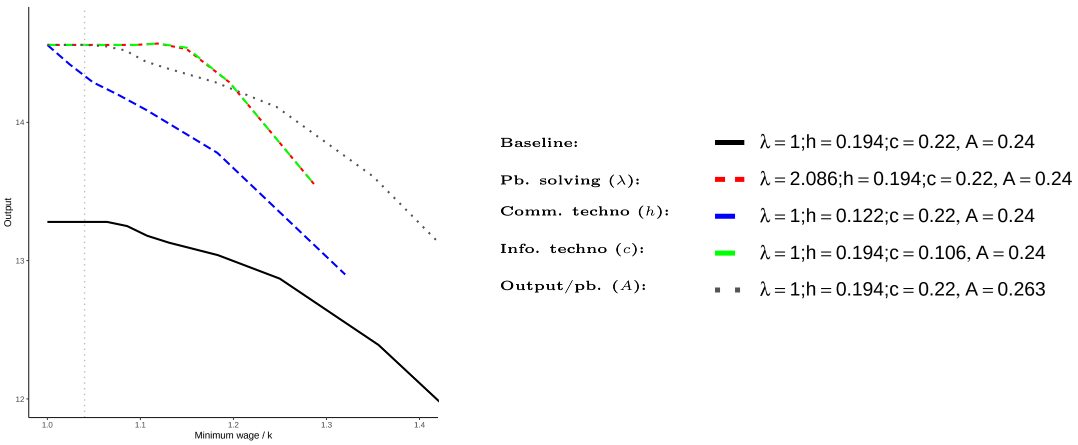
(A) Small Technological Shock: +3.2% Aggregate Output



(B) Moderate Technological Shock: +6.4% Aggregate Output



(C) Large Technological Shock: +9.6% Aggregate Output



Notes: this Figure replicates Panel (A) of Figure 9, and additionally considers different magnitudes of the underlying technological shocks in Panels (B) and (C). In Panel (B), these shocks are normalized so as to generate an economy-wide productivity premium of 6.4% in the "no minimum wage" case. As in Figure 9 and taking the "no minimum wage" scenario of Section 5.3 as a baseline, this is reached by, either: increasing λ (from 1 to 1.646), decreasing h (from 0.194 to 0.145), decreasing c (from 0.22 to 0.134) or increasing A (from 0.24 to 0.255). In Panel (C), these shocks are normalized so as to generate an economy-wide productivity premium of 9.6% in the "no minimum wage" case. Taking the "no minimum wage" scenario of Section 5.3 as a baseline, this is reached by, either: increasing λ (from 1 to 2.086, decreasing h (from 0.194 to 0.122), decreasing c (from 0.22 to 0.106) or increasing A (from 0.24 to 0.263).

of Panel (A), but by 4.9% with the medium technological shock of Panel (B) and by 6.3% with the large technological shock that is considered in Panel (C). This amplification result is intuitive, since minimum wages reduce the advantages for firms of organizing themselves into hierarchies. The benefits lost are particularly significant when hierarchies are "technologically" efficient, ie. when h is lowest.

In economies characterized by low c or large λ, the relative cost of moderate minimum wage constraints remains lower than in the A benchmark, but this advantage decreases when technological shocks become larger. For instance, for a minimum wage constraint set at 20% above the baseline price of labor, output drops by 1.0% (relative to a no minimum wage economy) in the case of the small technological shock of Panel (A), but by 1.3% with the medium technological shock of Panel (B) and by 2.1% with the large technological shock considered in Panel (C). Importantly, the situation reverses when minimum wage constraints become more stringent: in the upper range, large shocks in terms of c and λ tend to amplify the aggregate output cost of minimum wage constraints, to a point that the situation becomes worse than in the benchmark (A) scenario. For example, for a minimum wage constraint at $1.3k$, output drops by 4.8% (relative to the no minimum wage economy) in the case of the small technological shock of Panel (A), but by 6.4% with the medium technological shock of Panel (B) and by 7.5% with the large technological shock of Panel (C). In contrast, in the benchmark scenario where the technological shock is driven by parameter A rather than c or λ, the drop is output is 5.0% in all three cases. The intuition behind this result is as follows: for strict minimum wage constraints, firms flatten regardless of technological conditions. As a result, the organizational channel disappears and becomes ineffective in mitigating the minimum wage constraint. Ultimately, it prevents the economy from reaping the gains associated with any technology that would generate higher gains within a hierarchical organization: c, λ, or h, which all three are "leveraged" by hierarchical organizations - while general labor augmenting technologies (A) are not.

More generally, the curves obtained for different magnitudes of technological shocks are not homothetic, implying that major technological shocks of the type of the AI revolution ahead of us will most likely require a thorough re-examination of the costs and benefits of labor market institutions and regulations.

6 Conclusion

In this paper, we take advantage of a unique institutional setting, and of a unique information system in order to document how minimum wages impact firms' strategies and performances. In our empirical analyses, we obtain that the firm-level productivity channel is critical in mitigating the cost of minimum wage constraints. As in Caliendo and Rossi-Hansberg (2012), this productivity channel takes the form of organizational adjustments in terms of hierarchical layers, ie. a flattening of firms' organizations, and increases in the frequency of worker-level training programs. To rationalize these results, we propose a quantitative version of a structural model of optimal hierarchies which documents qualitatively as well as quantitatively how firms' internal organizations respond to increases in the minimum wage, allowing us to examine the implications of the model for workers' skills, wages and firms' productivity. We calibrate the model to salient moments of the French data and verify that it delivers predictions that are compatible with our (untargeted) estimation results. Then, we rely on this model to perform simulations which show that firms' training and endogenous organizational strategies are both key to mitigating the impact of (moderate) minimum wage constraints via their implications for aggregate productivity. In contrast, while reallocations across firms are significant, the contribution of the "pure" selection channel (Melitz, 2003) is quantitatively weak.

Further simulations show that the relative aggregate cost of minimum wage constraints is shaped by technologies: it is amplified in economies featuring more efficient communication technologies, while it tends to be dampened in economies featuring more efficient information or problem-solving technologies. However, the converse holds true when technological shocks and minimum wage constraints are both large. Overall, these results suggest that the current waves of AI-based technologies as well as the massive investment in communication technologies during the recent Covid-19 pandemic are likely to require a thorough re-examination of cost-benefit analyses of labor market institutions and regulations. In particular, our framework shows that the precise quantification of the shocks involved and their precise technological content are key dimensions to be further documented in order to draw truly robust implications for the labor market.

We leave several important aspects for future research. These include, notably, the analysis of the sensitivity of our quantifications to deviations from assumptions of monopolistic competition (on output markets) or perfect competition (on input/labor markets). These assumptions were directly taken from benchmark quantitative models of productivity analysis (Melitz, 2003) in order to ease the interpretation of our results. However, the recent literature documents that firms in at least some sectors might well have more power both on the output market (see e.g. De Loecker et al., 2020) and the labor market[72] than implied by these modeling assumptions. We expect, however, that firms first fully exhaust the organizational/productivity channel, as evidenced by the organizational changes in the data, in order to preserve as much profit as possible in the face of minimum wage constraints, before compressing mark-ups or wage markdowns. Relatedly, a second challenge[73] lies in a richer modeling of wages, incorporating a wider set of dimensions (especially, *ex ante* worker heterogeneity as in Garicano and Rossi-Hansberg, 2006) which could interact with minimum wages and deliver a further analysis of the welfare implications of minimum wages. Finally, detailed firm-level data on technology use, if it became available, would make it possible to endogenize (AI) technology adoption behavior and determine whether firms' organizational optimization problems are an important driver of technological change.

References

Aaronson, D., French, E., and MacDonald, J. (2008). The Minimum Wage, Restaurant Prices, and Labor Market Structure. *Journal of Human Resources*, 43(3):688–720.

Aaronson, D. and Phelan, B. J. (2019). Wage Shocks and the Technological Substitution of Low-wage Jobs. *The Economic Journal*, 129(617):1–34.

Acemoglu, D. (2001). Good Jobs versus Bad Jobs. *Journal of Labor Economics*, 19(1):1–21.

Acemoglu, D. (2024). The Simple Macroeconomics of AI. *Economic Policy (79th Economic Policy Panel, April 2024)*.

Acemoglu, D. and Johnson, S. (February 6, 2023). What's Wrong with ChatGPT? Technical report, Project Syndicate.

Aeberhardt, R., Givord, P., and Marbot, C. (2016). Spillover Effect of the Minimum Wage in France: An Unconditional Quantile Regression. Working Paper 2016-05, Center for Research in Economics and Statistics.

Ananian, S. and Calavrezo, O. (2011). Les trajectoires salariales des individus payés au voisinage du Smic dans le secteur privé. Une analyse empirique sur données françaises entre 1995 et 2007. *Économie et Statistique (DARES)*, 448(1):49–78.

Angelici, M. and Profeta, P. (2024). Smart working: Work flexibility without constraints. *Management Science*, 70(3):1680–1705.

[72] See for example Azar et al., 2023 or Berger et al., 2022 for recent analyses of the implications of monopsony power in presence of minimum wages. Separately quantifying the contribution of the productivity vs. market power channels would in particular require detailed information about firm-level output prices, which were not available in 2003/2006.

[73] The technical complexity of such extensions will be significant: to our knowledge, the theoretical set-up of Garicano and Rossi-Hansberg (2006) has never been calibrated to real-world data. The simplified version of the model in Antràs et al. (2006) has been calibrated to US data in an optimal taxation analysis in Lawson (2019).

Antràs, P., Garicano, L., and Rossi-Hansberg, E. (2006). Offshoring in a Knowledge Economy. *Quarterly Journal of Economics*, 121(1):31–77.

Arulampalam, W., Booth, A. L., and Bryan, M. L. (2004). Training and the New Minimum Wage. *Economic Journal*, 114(494):C87–C94.

Ashenfelter, O. and Jurajda, v. (2022). Minimum Wages, Wages, and Price Pass-Through: The Case of McDonald's Restaurants. *Journal of Labor Economics*, 40(S1):S179–S201.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The Fall of the Labor Share and the Rise of Superstar Firms. *Quarterly Journal of Economics*, 135(2):645–709.

Azar, J., Huet-Vaughn, E., Marinescu, I., Taska, B., and von Wachter, T. (2023). Minimum Wage Employment Effects and Labour Market Concentration. *Review of Economic Studies*, 91(4):1843–1883.

Bailey, M. J., DiNardo, J., and Stuart, B. A. (2021). The Economic Impact of a High National Minimum Wage: Evidence from the 1966 Fair Labor Standards Act. *Journal of Labor Economics*, 39(S2):S329–S367.

Barrero, J. M., Bloom, N., and Davis, S. J. (2021). Why Working from Home Will Stick. Working Paper 28731, National Bureau of Economic Research.

Bello, S. L. and Pesaresi, L. (2024). Equilibrium Effects of the Minimum Wage: The Role of Labor and Product Market Power. Working paper.

Berger, D. W., Herkenhoff, K. F., and Mongey, S. (2022). Minimum Wages, Efficiency and Welfare. Working Paper 29662, National Bureau of Economic Research.

Bloom, N., Han, R., and Liang, J. (2022). How Hybrid Working From Home Works Out. Working Paper 30292, National Bureau of Economic Research.

Bloom, N., Sadun, R., and Van Reenen, J. (2012). Americans Do IT Better: US Multinationals and the Productivity Miracle. *American Economic Review*, 102(1):167–201.

Brynjolfsson, E., Li, D., and Raymond, L. R. (2023). Generative AI at Work. Working Paper 31161, National Bureau of Economic Research.

Caliendo, L., Mion, G., Opromolla, L. D., and Rossi-Hansberg, E. (2020). Productivity and Organization in Portuguese Firms. *Journal of Political Economy*, 128(11):4211–4257.

Caliendo, L., Monte, F., and Rossi-Hansberg, E. (2015). The Anatomy of French Production Hierarchies. *Journal of Political Economy*, 123(4):809–852.

Caliendo, L. and Rossi-Hansberg, E. (2012). The Impact of Trade on Organization and Productivity. *Quarterly Journal of Economics*, 127(3):1393–1467.

Card, D., Katz, L. F., and Krueger, A. B. (1993). An Evaluation of Recent Evidence on the Employment Effects of Minimum and Subminimum Wages. Working Paper 4528, National Bureau of Economic Research.

Card, D. and Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *American Economic Review*, 84(4):772–793.

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The Effect of Minimum Wages on Low-Wage Jobs. *Quarterly Journal of Economics*, 134(3):1405–1454.

Collard-Wexler, A. and De Loecker, J. (2020). Productivity and Capital Measurement Error. Working Paper 22437, National Bureau of Economic Research.

Cottet, V., Quantin, Q., and Régnier, V. (2012). Coût du travail et allègements de charges: une estimation au niveau établissement de 1996 à 2008. *Document de travail de l'Insee*.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, 135(2):561–644.

Derenoncourt, E. and Montialoux, C. (2020). Minimum Wages and Racial Inequality. *Quarterly Journal of Economics*, 136(1):169–228.

Dickens, R., Machin, S., and Manning, A. (1999). The Effects of Minimum Wages on Employment: Theory and Evidence from Britain. *Journal of Labor Economics*, 17(1):1–22.

Dickens, R. and Manning, A. (2004). Has the National Minimum Wage Reduced UK Wage Inequality? *Journal of the Royal Statistical Society Series A*, 167(4):613–626.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica*, 64(5):1001–1044.

Draca, M., Machin, S., and Van Reenen, J. (2011). Minimum Wages and Firm Profitability. *American Economic Journal: Applied Economics*, 3(1):129–51.

Drucker, L., Mazirov, K., and Neumark, D. (2021). Who Pays for and Who Benefits from Minimum Wage Increases? Evidence from Israeli Tax Data on Business Owners and Workers. *Journal of Public Economics*, 199(C):104423.

Dustmann, C., Lindner, A., Schönberg, U., Umkehrer, M., and vom Berge, P. (2021). Reallocation Effects of the Minimum Wage. *Quarterly Journal of Economics*, 137(1):267–328.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2023). GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. Technical report.

Forsythe, E. (2023). The Effect of Minimum Wage Policies on the Wage and Occupational Structure of Establishments. *Journal of Labor Economics*, 41(S1):S291–S324.

Foster, L., Haltiwanger, J., and Syverson, C. (2008). Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *American Economic Review*, 98(1):394–425.

Garicano, L. (2000). Hierarchies and the Organization of Knowledge in Production. *Journal of Political Economy*, 108(5):874–904.

Garicano, L. and Hubbard, T. N. (2007). Managerial Leverage Is Limited by the Extent of the Market: Hierarchies, Specialization, and the Utilization of Lawyers' Human Capital. *Journal of Law and Economics*, 50(1):1–43.

Garicano, L., Lelarge, C., and Van Reenen, J. (2016). Firm Size Distortions and the Productivity Distribution: Evidence from France. *American Economic Review*, 106(11):3439–79.

Garicano, L. and Rossi-Hansberg, E. (2006). Organization and Inequality in a Knowledge Economy. *Quarterly Journal of Economics*, 121(4):1383–1435.

Garicano, L. and Rossi-Hansberg, E. (2015). Knowledge-Based Hierarchies: Using Organizations to Understand the Economy. *Annual Review of Economics*, 7:1–30.

Garicano, L. and Zandt, T. V. (2013). Hierarchies and the Division of Labor. In Gibbons, R. and Roberts, J., editors, *Handbook of Organizational Economics*, pages 604–654. Princeton University Press, Princeton.

Gibbs, M., Mengel, F., and Siemroth, C. (2023). Work from Home and Productivity: Evidence from Personnel and Analytics Data on Information Technology Professionals. *Journal of Political Economy Microeconomics*, 1(1):7–41.

Harasztosi, P. and Lindner, A. (2019). Who Pays for the Minimum Wage? *American Economic Review*, 109(8):2693–2727.

Hau, H., Huang, Y., and Wang, G. (2020). Firm Response to Competitive Shocks: Evidence from China's Minimum Wage Policy. *Review of Economic Studies*, 87(6):2639–2671.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *Quarterly Journal of Economics*, 124(4):1403–1448.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., and Mullainathan, S. (2017). Human Decisions and Machine Predictions. *Quarterly Journal of Economics*, 133(1):237–293.

Lawson, N. (2019). Taxing the Job Creators: Efficient Taxation with Bargaining in Hierarchical Firms. *Labour Economics*, 56:1–25.

Lee, D. S. (1999). Wage Inequality in the United States During the 1980s: Rising Dispersion or Falling Minimum Wage? *Quarterly Journal of Economics*, 114(3):977–1023.

Levinsohn, J. and Petrin, A. (2003). Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies*, 70(2):317–341.

Lydall, H. F. (1959). The Distribution of Employment Incomes. *Econometrica*, 27(1):110–115.

Machin, S., Manning, A., and Rahman, L. (2003). Where the Minimum Wage Bites Hard: Introduction of Minimum Wages to a Low Wage Sector. *Journal of the European Economic Association*, 1(1):154–180.

Manning, A. (2011). Chapter 11 - Imperfect Competition in the Labor Market. volume 4 of *Handbook of Labor Economics*, pages 973–1041. Elsevier.

Marschak, J. and Radner, R. (1972). *Economic Theory of Teams*. Yale University Press.

Mayneris, F., Poncet, S., and Zhang, T. (2018). Improving or disappearing: Firm-level adjustments to minimum wages in China. *Journal of Development Economics*, 135:20–42.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71(6):1695–1725.

Melitz, M. J. and Polanec, S. (2015). Dynamic Olley-Pakes productivity decomposition with entry and exit. *RAND Journal of Economics*, 46(2):362–375.

Metcalf, D. (2008). Why has the British National Minimum Wage had Little or No Impact on Employment? *Journal of Industrial Relations*, 50(3):489–512.

Morikawa, M. (2023). Productivity Dynamics of Remote Work During the COVID-19 Pandemic. *Industrial Relations: A Journal of Economy and Society*, 62(3):317–331.

Mullainathan, S. and Obermeyer, Z. (2021). Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care. *Quarterly Journal of Economics*, 137(2):679–727.

Neumark, D., Salas, J. M. I., and Wascher, W. (2014). Revisiting the Minimum Wage/Employment Debate: Throwing Out the Baby with the Bathwater? *ILR Review*, 67(3):608–648.

Neumark, D. and Wascher, W. (2001). Minimum Wages and Training Revisited. *Journal of Labor Economics*, 19(3):563–595.

Noy, S. and Zhang, W. (2023). Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. Technical report.

Peng, S., Kalliamvakou, E., Cihon, P., and Demirer, M. (2023). The Impact of AI on Developer Productivity: Evidence from GitHub Copilot.

Petrin, A. and Levinsohn, J. (2012). Measuring Aggregate Productivity Growth Using Plant-level Data. *RAND Journal of Economics*, 43(4):705–725.

Renkin, T., Montialoux, C., and Siegenthaler, M. (2022). The Pass-Through of Minimum Wages into U.S. Retail Prices: Evidence from Supermarket Scanner Data. *The Review of Economics and Statistics*, 104(5):890–908.

Riley, R. and Rosazza Bondibene, C. (2017). Raising the Standard: Minimum Wages and Firm Productivity. *Labour Economics*, 44:27–50.

Rizov, M., Croucher, R., and Lange, T. (2016). The UK National Minimum Wage's Impact on Productivity. *British Journal of Management*, 27(4):819–835.

Rosen, S. (1982). Authority, Control, and the Distribution of Earnings. *Bell Journal of Economics*, 13(2):311–323.

Rotman, D. (March 25, 2023). ChatGPT is about to revolutionize the economy. We need to decide what that looks like. https://www.technologyreview.com/2023/03/25/1070275/chatgpt-revolutionize-economy-decide-what-looks-like/, MIT Technology Review.

Sorkin, I. (2015). Are There Long-Run Effects of the Minimum Wage? *Review of Economic Dynamics*, 18(2):306–333.

Stewart, M. B. (2004). The Employment Effects of the National Minimum Wage. *Economic Journal*, 114(494):C110–C116.

Stewart, M. B. (2012). Quantile Estimates of Counterfactual Distribution Shifts and the Effect of Minimum Wage Increases on the Wage Distribution. *Journal of the Royal Statistical Society: Series A*, 175(1):263–287.

Svanberg, M., Li, W., Fleming, M., Goehring, B., and Thompson, N. (2024). Beyond AI Exposure: Which Tasks are Cost-Effective to Automate with Computer Vision? Working paper, Available at SSRN.

Wiltshire, J. C., McPherson, C., Reich, M., and Sosinskiy, D. (2023). Minimum Wage Effects and Monopsony Explanations. Department Discussion Papers 2023-02, Department of Economics, University of Victoria.

Wooldridge, J. M. (2009). On Estimating Firm-level Production Functions using Proxy Variables to Control for Unobservables. *Economics Letters*, 104(3):112–114.

Appendix

A Main Administrative References

- Labor law article (Code du travail) regulating the French Minimum Wage:

  https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000006072050/LEGISCTA000006178024/#LEGISCTA000006178024

- Aubry Laws implementing the GMR:

  https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000398162

- Official indices (SHBO) published by the French Ministry of Labor:

  https://dares.travail-emploi.gouv.fr/dares-etudes-et-statistiques/statistiques-de-a-a-z/article/les-indices-de-salaire-de-base

B Estimation Sample: Detailed Descriptive Statistics

This Appendix complements Section 3.3 by providing a comprehensive description of the regression sample. It corresponds to the population of firms that were created before 2000, that are consistently active until 2006 (at least), that employ at least 5 employees ("postes"), and whose value added is positive over this period. We furthermore exclude firms that had not signed an RWT agreement by 2002, in order to insure that all sample firms were subject to the same legal workweek.[74]

Tables B1 and B2 provide a detailed description of the structure of the sample in terms of the main characteristics of interest, as measured in 2002, the pre-regression year.[75] Table B1 first decomposes the descriptive statistics by GMR group. As explained in Section 2, this breakdown generates the main source of heterogeneity in terms of minimum wages, across firms. Table B1 shows that GMR1 and GMR2 correspond on average to larger firms in terms of jobs, which tend to also exhibit higher indices of productivity and more hierarchical layers. Although their minimum wage levels are lower than those of other firms (both in terms of "net" wages and total labor costs), they tend to pay higher wages on average, both overall and layer by layer. Lastly, these firms train their workers much more often than other firms, both at the (bottom) production layer and in managerial layers. Overall, these characteristics suggest that French firms self-selected across the different GMR groups. This suggests that non-observable characteristics are likely to generate endogeneity problems in our equations of interest, thus motivating the different estimation strategies explained in Section 3.1 (with associated results reported in Section 3.4 and Appendix D).

Table B2 provides a breakdown of the same descriptive statistics by magnitude of exposure to the minimum wage constraint. This dimension is measured by the share of workers who are paid in 2002 a lower amount than the 2006 legal minimum wage: these workers are thus all "hit" by the minimum wage constraint during our regression period (2003 to 2006). Perhaps surprisingly, this indicator is only weakly correlated with GMRs: although firms in GMR1 are 8 percentage points more likely to exhibit a low share (lower than 25%) of their workforce paid at the minimum wage, and firms in GMR5 are 9.5 percentage point less likely to do so, the figures are much more balanced across GMR2 to 4. In addition, as shown in Table B2, all GMRs are represented in each quartile of the variable capturing exposure to the minimum wage constraint. This feature of the data is interesting, since it secures statistical power for the triple-difference strategy proposed in Section 3.1. We obtain that firms' size as measured by jobs or hierarchical layers is much more correlated with their share of workers exposed to the minimum wage constraint than with their GMR affiliation. In contrast, productivity and labor costs are less correlated with the share of workers at the minimum wage than with GMR affiliation.

Lastly, Tables B3 to B5 provide information about the industry composition of our estimation sample, together with its correlation with the two previous variables: GMRs and exposure to minimum wage constraints. All industries are represented in all GMRs, although their respective industry structure exhibits significant heterogeneity. Most notably, extractive and manufacturing industries tend to be over-represented in GMR1 and 2, while hotels and restaurants concentrate in GMR3 to 5. Conversely, manufacturing industries are over-represented in the first quartile of firms least exposed to minimum wages, while hotels and restaurants are over-represented in the two intermediate quartiles and the extractive industries as well as business services are over-represented in the upper quartile of minimum wage exposure.

[74]The Fillon law of 2003 enabled the many companies that were late (in terms of the Aubry law), ie. that had not yet signed a RWT agreement, to remain *de facto* at 39 hours, provided they paid overtime at a higher rate than the "legal" hours. However, this rate depended on the size of the company: only 10% for companies with (strictly speaking) fewer than 20 employees, but 25% for companies with more than 20 employees (and even 50% above 8 hours' overtime). In addition to being subject to different working times, these firms were also able to strategically (and endogenously) manipulate their labor costs and sizes, which would further complicate our empirical estimation. For these two reasons, we exclude them from our regression analyses.

[75]Regression tables in the main text contain the mean evolution of these variables over the regression period, 2003 to 2006.

Table B1: Descriptive Statistics by GMR, 2002

GMR	GMR1	GMR2	GMR3	GMR4	GMR5	All firms
# Observations	3,500	14,810	10,945	25,220	869	55,344
Number of plants	2.889	3.755	2.059	1.317	2.330	2.232
per firm	(21.980)	(21.870)	(5.924)	(2.718)	(20.070)	(13.270)
Jobs	245	178	88	24	41	92
	(5,866)	(1,065)	(281)	(102)	(205)	(1,583)
VA (10^6€)	7.535	8.271	3.957	0.960	1.380	3.932
	(74.326)	(47.283)	(19.137)	(10.365)	(5.494)	(32.853)
VA (10^3€) / hour	0.031	0.033	0.030	0.028	0.026	0.030
	(0.069)	(0.189)	(0.022)	(0.067)	(0.013)	(0.109)
Total factor productivity,	0.843	0.836	0.886	0.841	0.837	0.849
Accounting method	(0.281)	(0.505)	(0.598)	(0.307)	(0.325)	(0.434)
Total factor productivity,	5.637	4.636	4.009	3.113	3.044	3.850
LevPet method	(33.540)	(17.970)	(25.940)	(11.620)	(2.847)	(18.750)
Number of layers	2.196	2.176	2.027	1.562	1.636	1.860
per firm	(0.773)	(0.778)	(0.790)	(0.865)	(0.888)	(0.870)
Share of firms with	0.030	0.030	0.036	0.113	0.113	0.070
0 layer	(0.170)	(0.171)	(0.186)	(0.317)	(0.316)	(0.256)
Share of firms with	0.131	0.139	0.191	0.350	0.304	0.248
1 layer	(0.337)	(0.346)	(0.393)	(0.477)	(0.460)	(0.432)
Share of firms with	0.453	0.455	0.483	0.399	0.418	0.434
2 layers	(0.498)	(0.498)	(0.500)	(0.490)	(0.493)	(0.496)
Share of firms with	0.386	0.376	0.290	0.138	0.166	0.248
3 layers	(0.487)	(0.484)	(0.454)	(0.345)	(0.372)	(0.432)
Layer 0:	0.627	0.670	0.652	0.684	0.689	0.671
Share in hours	(0.258)	(0.241)	(0.249)	(0.251)	(0.248)	(0.249)
Layer 1:	0.241	0.218	0.236	0.225	0.217	0.226
Share in hours	(0.176)	(0.166)	(0.184)	(0.188)	(0.179)	(0.181)
Layer 2:	0.120	0.102	0.102	0.082	0.085	0.094
Share in hours	(0.149)	(0.139)	(0.132)	(0.125)	(0.128)	(0.132)
Layer 3:	0.012	0.010	0.010	0.009	0.009	0.010
Share in hours	(0.026)	(0.021)	(0.024)	(0.029)	(0.027)	(0.026)
Hourly labor cost	7.803	8.342	8.725	8.830	8.520	8.609
at GMR level	(0.171)	(0.147)	(0.131)	(0.126)	(0.132)	(0.322)
Gross hourly wage at GMR level	7.177	7.266	7.391	7.423	6.917	7.352
	(0)	(0)	(0)	(0)	(0)	(0.111)
Share of firms with MW workers representing:						
$[0; \frac{1}{4}[$ of total workforce	0.885	0.817	0.806	0.786	0.710	0.804
$[\frac{1}{4}; \frac{1}{2}[$ of total workforce	0.073	0.100	0.111	0.135	0.155	0.117
$[\frac{1}{2}; \frac{3}{4}[$ of total workforce	0.031	0.068	0.074	0.064	0.104	0.065
$[\frac{3}{4}; 1[$ of total workforce	0.012	0.016	0.008	0.015	0.031	0.014
Share of 2002 workers	0.094	0.132	0.135	0.144	0.194	0.136
below the 2006 min. wage	(0.157)	(0.195)	(0.190)	(0.197)	(0.229)	(0.194)
Hourly labor cost:	18.750	18.490	17.850	17.480	16.810	17.890
Average	(6.165)	(6.653)	(5.949)	(6.228)	(5.622)	(6.297)
Hourly labor cost:	14.020	14.030	13.660	13.550	13.100	13.720
Layer 0	(3.331)	(3.509)	(3.069)	(3.088)	(2.790)	(3.222)
Hourly labor cost:	21.160	21.980	21.070	21.970	20.500	21.710
Layer 1	(6.081)	(9.301)	(8.032)	(9.401)	(7.564)	(8.904)
Hourly labor cost:	37.760	39.740	37.910	37.820	37.100	38.450
Layer 2	(12.610)	(15.560)	(14.060)	(15.940)	(14.410)	(15.170)
Hourly labor cost:	70.250	72.290	67.390	58.590	60.940	67.350
Layer 3	(35.670)	(38.740)	(35.220)	(30.370)	(40.320)	(36.090)
Training programs (Y/N)	0.455	0.443	0.417	0.273	0.392	0.412
Training at layer L_0	0.340	0.317	0.297	0.178	0.236	0.294
Training at layers L_1 to L_3	0.340	0.334	0.309	0.189	0.296	0.308

Source: Estimation sample, constructed from matched DADS, FICUS, BRN, "(Aubry) payroll subsidy" (French Ministry of Labor) and "CERFA 2483" (French Ministry of Labor) files. In each cell of the table, we report averages together with standard deviations (in parentheses). In this table, the "Gross hourly wage at GMR level" has no within-GMR variation, since it corresponds exactly to the legally binding GMR value.

Table B2: Descriptive Statistics by Sub-samples of Varying Shares of Low-Wage Workers, 2002

Share of LW workers:	$[0; \frac{1}{4}[$	$[\frac{1}{4}; \frac{1}{2}[$	$[\frac{1}{2}; \frac{3}{4}[$	$[\frac{3}{4}; 1]$	All firms
# Observations	44,474	6,481	3,617	772	55,344
Number of plants	2.323	2.093	1.570	1.242	2.232
per firm	(14.210)	(10.030)	(5.537)	(1.188)	(13.270)
Jobs	100	58	71	37	92
	(1,758)	(274)	(441)	(66)	(1,583)
VA (10^6€)	4.593	1.316	1.221	561	3.932
	(36.531)	(4.843)	(6.070)	(752)	(32.854)
VA (10^3€) / hour	0.032	0.022	0.019	0.017	0.030
	(0.122)	(0.016)	(0.007)	(0.008)	(0.109)
Total factor productivity,	0.862	0.812	0.772	0.776	0.849
Accounting method	(0.461)	(0.300)	(0.274)	(0.409)	(0.434)
Total factor productivity,	3.827	4.110	3.799	3.178	3.850
LevPet method	(19.510)	(17.490)	(12.320)	(5.249)	(18.750)
Number of layers	1.960	1.538	1.411	0.861	1.860
	(0.836)	(0.857)	(0.892)	(0.845)	(0.870)
Share of firms with	0.052	0.109	0.158	0.389	0.070
0 layer	(0.222)	(0.312)	(0.364)	(0.488)	(0.256)
Share of firms with	0.214	0.376	0.393	0.408	0.248
1 layer	(0.410)	(0.485)	(0.488)	(0.492)	(0.432)
Share of firms with	0.455	0.381	0.330	0.157	0.434
2 layers	(0.498)	(0.486)	(0.470)	(0.364)	(0.496)
Share of firms with	0.279	0.133	0.119	0.047	0.248
3 layers	(0.448)	(0.340)	(0.324)	(0.211)	(0.432)
Layer 0:	0.632	0.795	0.863	0.928	0.671
Share in hours	(0.253)	(0.160)	(0.116)	(0.104)	(0.249)
Layer 1:	0.248	0.162	0.111	0.063	0.226
Share in hours	(0.185)	(0.139)	(0.106)	(0.098)	(0.181)
Layer 2:	0.109	0.039	0.023	0.008	0.094
Share in hours	(0.141)	(0.060)	(0.037)	(0.024)	(0.132)
Layer 3:	0.011	0.004	0.004	0.001	0.010
Share in hours	(0.028)	(0.016)	(0.013)	(0.006)	(0.026)
Hourly labor cost	8.599	8.650	8.653	8.614	8.609
at GMR level	(0.329)	(0.294)	(0.272)	(0.297)	(0.322)
Gross hourly wage at GMR level	7.351	7.358	7.347	7.335	7.352
	(0.111)	(0.113)	(0.114)	(0.125)	(0.111)
Share of firms:					
in GMR1	0.070	0.040	0.030	0.048	0.063
in GMR2	0.272	0.228	0.276	0.302	0.268
in GMR3	0.198	0.187	0.225	0.118	0.198
in GMR4	0.446	0.525	0.444	0.497	0.456
in GMR5	0.014	0.021	0.025	0.035	0.016
Share of 2002 workers	0.053	0.362	0.608	0.836	0.136
below the 2006 min. wage	(0.067)	(0.072)	(0.069)	(0.071)	(0.194)
Hourly labor cost:	19.250	13.140	11.430	9.911	17.890
Average	(6.228)	(2.256)	(1.392)	(1.052)	(6.297)
Hourly labor cost:	14.510	10.940	9.994	9.233	13.720
Layer 0	(3.081)	(1.014)	(0.696)	(0.615)	(3.222)
Hourly labor cost:	22.290	19.280	18.690	19.070	21.710
Layer 1	(8.888)	(8.330)	(8.578)	(9.696)	(8.904)
Hourly labor cost:	38.960	35.490	34.730	35.190	38.450
Layer 2	(14.990)	(16.170)	(15.450)	(16.300)	(15.170)
Hourly labor cost:	68.620	57.080	52.880	51.470	67.350
Layer 3	(36.510)	(28.670)	(31.220)	(22.600)	(36.090)
Training programs (Y/N)	0.424	0.318	0.267	0.214	0.412
Training at layer L_0	0.311	0.161	0.106	0.029	0.294
Training at layers L_1 to L_3	0.324	0.166	0.104	0.053	0.308

Source: Estimation sample, constructed from matched DADS, FICUS, BRN, "(Aubry) payroll subsidy" (French Ministry of Labor) and "CERFA 2483" (French Ministry of Labor) files. The variable defining the table columns corresponds to the "share of low-wage workers", as measured in 2002 by the total number of hours worked by employees earning in 2002 less than the 2006 minimum wage (in nominal terms). In each cell of the table, we report averages together with standard deviations (in parentheses).

Table B3: Industry Structure of the Estimation Sample, by GMR Group

	NACE 1-digit	GMR1 3,500	GMR2 14,810	GMR3 10,945	GMR4 25,220	GMR5 869	All GMR 55,344
Extractive industries, food, textiles, wood	1	0.092	0.086	0.067	0.053	0.062	0.067
Manufacturing industries: paper, chemicals, minerals, metals	2	0.188	0.179	0.184	0.126	0.115	0.155
Manufacturing industries: equipment, machinery, transportation	3	0.085	0.06	0.060	0.044	0.037	0.054
Electricity, gas, water, construction	4	0.113	0.097	0.088	0.146	0.110	0.119
Trade, hotels, restaurants	5	0.207	0.268	0.385	0.400	0.435	0.350
Transport, communications, finance	6	0.037	0.056	0.037	0.022	0.039	0.035
Real estate, business services, administration	7	0.222	0.177	0.110	0.130	0.140	0.145
Education, health, social action	8	0.029	0.056	0.040	0.036	0.046	0.042
Other services (personal, domestic)	9	0.027	0.021	0.028	0.042	0.015	0.032
All industries		1	1	1	1	1	1

Source: DADS files.

Table B4: Industry Structure of the Estimation Sample, by Share of Low-Wage Workers

	NACE 1-digit	$[0; \frac{1}{4}[$ 44,866	$[\frac{1}{4}; \frac{1}{2}[$ 6,293	$[\frac{1}{2}; \frac{3}{4}[$ 3,473	$[\frac{3}{4}; 1]$ 712	All firms 55,344
Extractive industries, food, textiles, wood	1	0.063	0.093	0.062	0.132	0.067
Manufacturing industries: paper, chemicals, minerals, metals	2	0.180	0.074	0.030	0.023	0.155
Manufacturing industries: equipment, machinery, transportation	3	0.060	0.040	0.014	0.021	0.054
Electricity, gas, water, construction	4	0.143	0.029	0.006	0.006	0.119
Trade, hotels, restaurants	5	0.290	0.570	0.671	0.499	0.350
Transport, communications, finance	6	0.041	0.016	0.008	0.006	0.035
Real estate, business services, administration	7	0.152	0.083	0.147	0.216	0.145
Education, health, social action	8	0.042	0.045	0.032	0.038	0.042
Other services (personal, domestic)	9	0.029	0.050	0.032	0.058	0.032
All industries		1	1	1	1	1

Source: DADS files. In this table, low-wage workers are defined as workers earning in 2002 less than the 2006 minimum wage.

Table B5: Share of Low-Wage Workers, by Industry, in the Estimation Sample

	NACE 1-digit	$[0; \frac{1}{4}[$ 44,866	$[\frac{1}{4}; \frac{1}{2}[$ 6,293	$[\frac{1}{2}; \frac{3}{4}[$ 3,473	$[\frac{3}{4}; 1]$ 712	All firms 55,344
Extractive industries, food, textiles, wood	1	0.751	0.161	0.060	0.027	1
Manufacturing industries: paper, chemicals, minerals, metals	2	0.930	0.056	0.012	0.002	1
Manufacturing industries: equipment, machinery, transportation	3	0.890	0.088	0.017	0.005	1
Electricity, gas, water, construction	4	0.968	0.028	0.003	0.001	1
Trade, hotels, restaurants	5	0.664	0.191	0.125	0.020	1
Transport, communications, finance	6	0.931	0.052	0.015	0.003	1
Real estate, business services, administration	7	0.845	0.068	0.067	0.021	1
Education, health, social action	8	0.812	0.127	0.049	0.012	1
Other services (personal, domestic)	9	0.728	0.183	0.064	0.025	1
All industries		0.811	0.114	0.063	0.013	1

Source: DADS files. In this table, low-wage workers are defined as workers earning in 2002 less than the 2006 minimum wage.

C Empirical Results: Robustness Checks and Extensions

C.1 Production Hierarchies in the Estimation Sample

The regressions documenting the impact of minimum wages on firms' organizations in Section 3.4 rely critically on the methodology proposed by Caliendo et al. (2015) for mapping the French classification of occupations with hierarchical layers. This appendix provides additional descriptive statistics on the firm organizations that emerge from the French data using this methodology.

Table C1 shows that the proposed empirical concept of hierarchical layer generates firm-level organizations that are most often well-behaved "pyramids", with the lower levels being larger than the upper ones. This feature holds true in the vast majority of firms when the size of each layer is measured in terms of hours (Panel A), and even more robustly when the size is measured in terms of jobs (Panel B). We obtain that more than 90% of 1 (managerial) layer firms have more production workers than managers, more than 70% of 2 (managerial) layer firms have simultaneously more production workers than middle managers, and more middle managers than top managers. Lastly, almost 66% of 3 (managerial) layer firms also have regular pyramidal forms. Panel (C) complements the picture by showing that labor costs most often increase monotonically from one hierarchical layer to the next. This monotonicity holds true for 94% of 1 layer firms, 90% of 2 layer firms and 82% of 3 layer firms.

These descriptive statistics guide us in our modeling choices in Section 4. In particular, the assumption that knowledge is cumulative guarantees the monotonicity of wages from one layer to the next, which is almost universally observed in French firms.

Table C1: Check of Organizational Hierarchy Conditions

(A) Using Hours (2003)

Condition tested:	$h_L^{\ell} \geq h_L^{\ell+1}$ for all ℓ	$h_L^0 \geq h_L^1$	$h_L^1 \geq h_L^2$	$h_L^2 \geq h_L^3$
1-layer firms	0.901	0.901		
2-layer firms	0.621	0.796	0.793	
3-layer firms	0.537	0.787	0.794	0.877

(B) Using Jobs (2003)

Condition tested:	$N_L^{\ell} \geq N_L^{\ell+1}$ for all ℓ	$N_L^0 \geq N_L^1$	$N_L^1 \geq N_L^2$	$N_L^2 \geq N_L^3$
1-layer firms	0.921	0.921		
2-layer firms	0.722	0.825	0.874	
3-layer firms	0.658	0.809	0.831	0.968

(C) Using Hourly Labor Costs (2003)

Condition tested:	$w_L^{\ell+1} \geq w_L^{\ell}$ for all ℓ	$w_L^1 \geq w_L^0$	$w_L^2 \geq w_L^1$	$w_L^3 \geq w_L^2$
1-layer firms	0.940	0.940		
2-layer firms	0.899	0.947	0.951	
3-layer firms	0.822	0.972	0.962	0.881

Source: The table reports the share of firms that satisfy Caliendo et al. (2015)'s *hierarchy condition*: that lower hierarchical layers are larger. In Panel (A), the size of each layer is measured using the total number of hours. In Panel (B), the size of each layer is measured using the total number of jobs. Lastly, Panel (C) illustrates share of firms that satisfy the hierarchy condition that hourly wages (as measured by total labor costs) are higher in the higher hierarchical layers of firms.

C.2 Production Factors Beyond Labor

Our paper primarily aims to quantify the magnitude of the organizational adjustments of firms when they face tighter minimum wage constraints (and their endogenous productivity implications). By construction, the main text thus focuses on labor as the main production factor. However, by altering the price of labor relative to other production factors, minimum wages are likely to also affect firms' demand for other production factors.

Table C2 relies on the same estimation framework as Section 3.4 to inspect whether these complementary adjustments are large relative to our mechanism of main interest. Columns (1) to (3) focus on the share of expenditures allocated to different production factors, normalized by sales. We obtain that the share of labor increases significantly, and that it is compensated by lower shares of both intermediate inputs and capital. The estimated coefficients are however imprecisely estimated in these two cases, and none of them is significant. In the case of capital, this might be due to measurement error, since capital is typically difficult to measure from accounting data because they only contain historical costs (see for example Collard-Wexler and De Loecker, 2020 for a recent discussion). Column (4) shows that capital intensity, measured by capital stock per hour, has remained stable overall, which is consistent with a decrease in capital expenditure, since the denominator of this ratio, hours worked, has contracted (see Table 4).

In contrast to capital, intermediate inputs are *a priori* well measured in the accounting data, suggesting that firms did not alter their usage of intermediate inputs in response to increases in the minimum wage. In particular, outsourcing would typically be registered as intermediate expenses: column (2) of Table C2 thus shows that outsourcing was not an important channel of adjustment to higher minimum wages during our period of interest.

Table C2: Impact of Minimum Wage Increases on Production Factors Beyond Labor

	(1)	(2)	(3)	(4)
Indicator:	Δ labor costs / sales	Δ intermediate inputs / sales	Δ capital costs / sales	Δ ln capital /hours
Mean:	0.001	0.005	0.004	0.118
Δ ln hourly cost at GMR	0.060***	-0.017	-0.010	-0.059
	(0.013)	(0.013)	(0.016)	(0.076)
# observations	55,002	55,010	55,344	55,344
R^2	0.040	0.039	0.015	0.030

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the 1-digit sector-commuting zone level in parentheses. Regression results from equation (3), estimated by OLS. All specifications contain 4-digit industry fixed effects and commuting zone fixed effects. In Column (3), the user cost of capital is approximated as 20% of the capital stock carried on the balance sheet (as for example in Garicano et al., 2016).

D Regression Results: Alternative Estimation Strategies and Specifications

This Appendix re-estimates the specifications reported in the main text using the three alternative estimation strategies proposed in Section 3.1.

Instrumented differences-in-differences. In Table D1, we first re-estimate the double-difference specifications of the main text, but combine them with an instrumental variable strategy for the main regressor of interest: minimum-wage labor costs (see Section 3.1 for rationale).[76] Overall, the uninstrumented double-difference estimates reported in the main text appear to be rather conservative (i.e. a lower bound) for our mechanisms of interest:

- Labor costs in Panel (A): the positive sign of the main difference-in-differences specification is preserved and remains statistically significant. The magnitude of the coefficients is however one third lower than in the main specification.
- Relative sizes of hierarchical layers in Panel (B): the (negative) sign and its statistical significance are preserved. We obtain that the total number of hours is negatively affected and that this contraction mainly occurs at the bottom of the hierarchy. The magnitude of the estimated coefficient at layer L_0 is approximately two times larger when instrumented (Table D1) than with simple OLS (Table 2).
- Number of hierarchical layers in Panel (C): the instrumented coefficient has the same sign and magnitude as the un-instrumented version of Table 2, but it is imprecisely estimated and thus not statistically significant. The other results are preserved, and if anything amplified in terms of magnitude.
- Probability of training programs in Panel (D): the sign and significance of the estimated coefficients are preserved, and magnitudes are doubled.
- Firms' sizes in Panel (E): the negative, statistically significant impact on hours and jobs is preserved, and slightly amplified in terms of magnitude. As in Table 4, the impact on sales or value added is small in magnitude and insignificant.
- Firms' productivities in Panel (F): the (positive) coefficients for labor productivity and TFP are fully preserved, in terms of both magnitude and statistical significance.

Table D1: Impact of Minimum Wage Increases on Firms' Outcomes: IV Estimations

(A) Impact on Labor Costs

	(1)	(2)	(3)	(4)	(5)
Indicator:	Δ ln total hourly labor cost				
	Total	Layer 0	Layer 1	Layer 2	Layer 3
Mean:	0.095	0.110	0.097	0.073	0.106
Δ ln hourly cost at GMR	0.234*** (0.028)	0.0987*** (0.031)	-0.0217 (0.061)	0.265*** (0.077)	0.0487 (0.184)
# observations	55,344	55,344	49,479	33,202	8,484

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the 1-digit sector-commuting zone level in parentheses. Regression results from equation (3), estimated by 2SLS, using the increase in legal GMRs as an instrumental variable for the evolution of the corresponding total labor costs. Panel (E), Column (5) reports the first stage regression. Value of the Stock-Yogo Cragg-Donald Wald F-statistic: 1.0e+05. All specifications contain 4-digit industry fixed effects and commuting zone fixed effects.

[76] The value of the Stock-Yogo Cragg-Donald Wald F-statistic is 1.0e+05.

Table D1: Impact of Minimum Wage Increases on Firms' Outcomes: IV Estimations, Continued

(B) Impact on the Relative Size of Hierarchical Layers

Indicator:	(1)	(2)	(3) Δ ln hours	(4)	(5)
	Total	Layer 0	Layer 1	Layer 2	Layer 3
Mean:	0.014	0.016	-0.007	0.041	-0.043
Δ ln hourly cost at GMR	-0.322*** (0.068)	-0.760*** (0.113)	-0.007 (0.148)	-0.098 (0.173)	-0.155 (0.321)
# observations	55,344	55,344	49,479	33,202	8,484

(C) Impact on the Number of Hierarchical Layers

Indicator:	(1) Δ number hierarchical Layers	(2) Probability of increase	(3) Probability of decrease	(4) Probability of stability	(5) Δ Proba 3 layers	(6) Δ Proba at least 2 layers	(7) Δ Proba at least 1 layers
Mean :	-0.036	0.165	0.193	0.642	-0.015	-0.013	-0.008
Δ ln hourly cost at GMR	-0.193 (0.133)	-0.516*** (0.074)	-0.297*** (0.081)	0.813*** (0.096)	-0.286*** (0.087)	0.0634 (0.078)	0.0289 (0.043)
# observations	55,344	55,344	55,344	55,344	55,344	55,344	55,344

(D) Impact on Training across Layers

Indicator:	Δ Probability Training Hours in L0			
	All firms	1-layer firms	2-layer firms	3-layer firms
Shares -1/+1:	0.07/0.25	0.08/0.20	0.07/0.25	0.06/0.28
Δ ln hourly cost at GMR	1.098*** (0.183)	0.471 (0.565)	0.802*** (0.255)	1.417*** (0.380)
# observations	17,406	2,494	9,688	4,569
Indicator:	Δ Probability Training Hours in Managerial Layers			
	All firms	1-layer firms	2-layer firms	3-layer firms
Shares -1/+1:	0.07/0.24	0.08/0.16	0.07/0.25	0.06/0.27
Δ ln hourly cost at GMR	1.115*** (0.178)	0.687 (0.512)	0.984*** (0.253)	1.017*** (0.367)
# observations	17,406	2,494	9,688	4,569

(E) Impact on Firms' Size

Indicator:	(1) Δ ln hours	(2) Δ ln # jobs	(3) Δ ln VA	(4) Δ ln Sales	(5) Δ ln hourly cost at GMR
Mean:	0.014	0.009	0.084	0.106	0.072
Δ ln hourly cost at GMR	-0.322*** (0,068)	-0.403*** (0.069)	-0.079 (0.084)	-0.031 (0.069)	
Δ ln GMR					1.961*** (0,009)
# observations	55,344	55,344	55,344	55,344	55,344

(F) Impact on Productivity

Indicator:	(1) Δ ln Labor	(2)	(3)	(4) Δ ln TFP	(5)	(6)
Measurement of TFP:	productivity		Accounting		Lev. Pet.	
Measurement of labor:	hours	jobs	hours	jobs	hours	jobs
Mean:	0.070	0.075	0.023	0.025	0.023	0.022
Δ ln hourly cost at GMR	0.244*** (0.072)	0.324*** (0.079)	0.051* (0.029)	0.063** (0.030)	0.068** (0.030)	0.061* (0.031)
# observations	55,344	55,344	53,436	53,436	53,436	53,436

Notes, continued: All specifications contain 4-digit industry fixed effects and commuting zone fixed effects. In Panel (B), the composition of the estimation sample evolves across columns. In the first two columns, all sample firms are included. In column (3), only firms exhibiting at least 2 layers (Layer 0 and Layer 1) in 2003 and 2006 are included in the estimation sample. Similarly, in columns (4) to (5), only firms exhibiting at least 3 or 4 layers respectively, both in 2003 and 2006, are included in the estimation sample. In Panel (F) Columns (3) and (4), the productivity indices have been constructed using the accounting method, with the elasticities of value added to production factors (labor and capital) being evaluated by their share of compensation in value added. In Columns (5) and (6), the estimation method proposed by Levinsohn and Petrin (2003) is implemented using Wooldridge (2009) and Petrin and Levinsohn (2012). Labor is measured in terms of hours in Columns (1), (3), and (5), and in terms of jobs in columns (2), (4), and (6).

Triple differences. In the specifications with triple differences (Table D2), we introduce an additional interaction between our baseline variable capturing the increase in labor costs at the level of GMR, with the (*ex ante*, 2002) share of low-wage workers in the considered firm. The latter variable is defined as in Table B2, as the share of workers of the bottom (production) layer who earned in 2002 less than the 2006 minimum wage, weighted by hours worked. The interaction becomes the variable of main interest. It captures how the baseline impact of increases in labor costs at the GMR is modulated in firms with ex ante high, or conversely low, proportions of low-wage workers (see Section 3.1 for rationale). As Table B2 shows, this proportion varies between 0 and 1 in our data, with most companies falling between 0 and 0.25, and the sample average being 0.136 with a standard deviation of 0.2. This implies that the estimated coefficients of main interest, ie. those associated with the interaction terms have to be typically divided by 5 in order to capture the standard sample heterogeneity and be compared with our estimates of the main text:

- Labor costs in Panel (A): the main result that labor costs increase at the bottom hierarchical layer is preserved and significant, but the associated magnitude is reduced as compared with Table 1, by an order 5 (taking account that the interaction term has to be normalized, see above). In addition, we obtain in these triple difference specifications that the transmission of increases in labor costs at the bottom layer are attenuated, rather than preserved, in subsequent hierarchical layers. The resulting pooled impact, estimated across all workers in all firms, is not statistically different from zero.
- Relative sizes of hierarchical layers in Panel (B): we obtain that the pooled negative impact on total hours worked is preserved in magnitude, but only weakly significant. The breakdown across layers is not significant.
- Number of hierarchical layers in Panel (C): Organizational adjustments are estimated to be amplified in firms having a high share of minimum wage workers. The estimated interaction terms show that the negative impact of total labor costs at the minimum wage is globally preserved for the total number of layers (taking account of the normalization by 5 that has to be implemented). The channel generating this result is, however, different in the double vs. triple difference specifications: in Table D2, it appears to be determined by a higher probability of decreasing the number of hierarchical levels and a lower probability of selecting complex organizations, rather than by the lower probability of increasing the number of layers (as was the case in Table 2).
- Probability of implementing training programs in Panel (D): the estimated results in the triple difference specifications are unfortunately at most weakly significant, but when this is the case, correctly signed and of the same order of magnitude as in Table 3.
- Firms' sizes in Panel (E): the results from Table 4 are fully preserved, both qualitatively and quantitatively, but only weakly significant.
- Firms' productivities in Panel (F): the obtained results are fully preserved, both quantitatively and in terms of significance, in specifications where productivity is measured by labor productivity or by TFP using an accounting method. However, they lack significance in specifications where productivity is measured by TFP estimated with the method proposed by Petrin and Levinsohn (2012).

Table D2: Impact of Minimum Wage Increases on Firms' Outcomes: Triple Differences

(A) Impact on Labor Costs

	(1)	(2)	(3)	(4)	(5)
Indicator:	Δ ln total hourly labor cost				
	Total	Layer 0	Layer 1	Layer 2	Layer 3
Mean:	0.095	0.110	0.097	0.073	0.106
Δ ln hourly cost	0.376***	0.265***	0.260***	0.520***	0.227
at GMR	(0.025)	(0.029)	(0.054)	(0.071)	(0.178)
Share LW hours	0.121***	0.334***	-0.625*	-1.377**	-0.127
	(0.009)	(0.123)	(0.354)	(0.561)	(1.210)
Interaction	-0.085	0.334***	-0.625*	-1.377**	-0.127
	(0.126)	(0.123)	(0.354)	(0.561)	(1.210)
# observations	55,344	55,344	49,479	33,202	8,484
R^2	0.059	0.046	0.024	0.035	0.103

(B) Impact on the Relative Size of Hierarchical Layers

	(1)	(2)	(3)	(4)	(5)
Indicator:	Δ ln hours				
	Total	Layer 0	Layer 1	Layer 2	Layer 3
Mean:	0.014	0.016	-0.007	0.041	-0.043
Δ ln hourly cost	-0.093	-0.400***	0.025	-0.344**	0.222
at GMR	(0.060)	(0.106)	(0.135)	(0.161)	(0.325)
Share LW hours	-0.024	-0.114***	0.124*	-0.210**	0.122
	(0.024)	(0.032)	(0.064)	(0.094)	(0.188)
Interaction	-0.655*	-0.106	-0.630	2.799**	-1.617
	(0.366)	(0.473)	(0.843)	(1.180)	(2.050)
# observations	55,344	55,344	49,479	33,202	8,484
R^2	0.052	0.033	0.028	0.035	0.099

(C) Impact on the Number of Hierarchical Layers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Indicator:	Δ number hierarchical Layers	Probability of increase	Probability of decrease	Probability of stability	Δ Proba 3 layers	Δ Proba at least 2 layers	Δ Proba at least 1 layer
Mean :	-0.036	0.165	0.193	0.642	-0.015	-0.013	-0.008
Δ ln hourly cost	-0.045	-0.290***	-0.188**	0.478***	-0.231***	0.103	0.083**
at GMR	(0.122)	(0.068)	(0.076)	(0.091)	(0.082)	(0.069)	(0.040)
Share LW hours	0.188***	0.038	-0.118***	0.080**	0.011	0.097***	0.080***
	(0.048)	(0.028)	(0.028)	(0.034)	(0.023)	(0.030)	(0.025)
Interaction	-1.394**	-0.125	0.993***	-0.868*	0.218	-0.777*	-0.836***
	(0.647)	(0.368)	(0.374)	(0.455)	(0.329)	(0.402)	(0.316)
# observations	55,344	55,344	55,344	55,344	55,344	55,344	55,344
R^2	0.023	0.025	0.025	0.028	0.025	0.019	0.019

(D) Impact on Training across Layers

Indicator:	Δ Probability Training Hours in L_0			
	All firms	1-layer firms	2-layer firms	3-layer firms
Shares -1/+1:	0.07/0.25	0.08/0.20	0.07/0.25	0.06/0.28
Δ ln hourly cost	0.495***	-0.045	0.261	0.899**
at GMR	(0.171)	(0.608)	(0.237)	(0.369)
Share LW hours	-0.155**	-0.215	-0.135	0.008
	(0.079)	(0.161)	(0.119)	(0.245)
Interaction	0.726	3.877*	0.503	-0.498
	(0.946)	(2.148)	(1.379)	(2.630)
# observations	17,406	2,494	9,688	4,569
R^2	0.083	0.230	0.115	0.181
Indicator:	Δ Probability Training Hours in Managerial Layers			
	All firms	1-layer firms	2-layer firms	3-layer firms
Shares -1/+1:	0.07/0.24	0.08/0.16	0.07/0.25	0.06/0.27
Δ ln hourly cost	0.536***	-0.129	0.406*	0.562*
at GMR	(0.166)	(0.573)	(0.230)	(0.339)
Share LW hours	-0.130*	-0.255	-0.061	0.044
	(0.077)	(0.163)	(0.116)	(0.242)
Interaction	0.153	3.062	-0.291	-0.975
	(0.916)	(2.133)	(1.353)	(2.565)
# observations	17,406	2,494	9,688	4,569
R^2	0.081	0.247	0.105	0.194

Table D2: Impact of Minimum Wage Increases on Firms' Outcomes: Triple Differences, Continued

(E) Impact on Firms' Size

Indicator:	(1) Δ ln hours	(2) Δ ln # jobs	(3) Δ ln VA	(4) Δ ln Sales
Mean:	0.014	0.009	0.084	0.106
Δ ln hourly cost	-0.093	-0.161***	0.036	0.018
at GMR	(0.060)	(0.062)	(0.077)	(0.064)
Share LW hours	-0.024	-0.019	-0.025	-0.010
	(0.024)	(0.025)	(0.029)	(0.025)
Interaction	-0.655*	-0.617*	0.135	-0.137
	(0.366)	(0.371)	(0.395)	(0.337)
# observations	55,344	55,344	55,344	55,344
R^2	0.052	0.048	0.061	0.068

(F) Impact on Productivity

	(1)	(2)	(3)	(4)	(5)	(6)
Indicator:	Δ ln Labor		Δ ln TFP			
Measurement of TFP:	productivity		Accounting		Lev. Pet.	
Measurement of labor:	hours	jobs	hours	jobs	hours	jobs
Mean:	0.07	0.075	0.023	0.025	0.023	0.022
Δ ln hourly cost	0.129*	0.197***	0.043	0.052*	0.066**	0.069**
at GMR	(0.067)	(0.073)	(0.027)	(0.028)	(0.027)	(0.028)
Share LW hours	-0.002	-0.007	0.000	-0.003	0.007	0.008
	(0.025)	(0.028)	(0.011)	(0.011)	(0.011)	(0.012)
Interaction	0.790***	0.752**	0.261*	0.268*	0.086	0.056
	(0.332)	(0.363)	(0.144)	(0.149)	(0.149	(0.159)
# observations	55,344	55,344	53,436	53,436	53,436	53,436
R^2	0.044	0.040	0.081	0.074	0.159	0.184

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the 1-digit sector-commuting zone level in parentheses. Regression results from equation (4), estimated by OLS. All specifications contain 4-digit industry fixed effects and commuting zone fixed effects. In Panel (B), the composition of the estimation sample evolves across columns. In the first two columns, all sample firms are included. In column (3), only firms exhibiting at least 2 layers (Layer 0 and Layer 1) in 2003 and 2006 are included in the estimation sample. Similarly, in columns (4) to (5), only firms exhibiting at least 3 or 4 layers respectively, both in 2003 and 2006, are included in the estimation sample. In Panel (F) Columns (3) and (4), the productivity indices have been constructed using the accounting method, with the elasticities of value added to production factors (labor and capital) being evaluated by their share of compensation in value added. In Columns (5) and (6), the estimation method proposed by Levinsohn and Petrin (2003) is implemented using Wooldridge (2009) andd. Labor is measured in terms of hours in Columns (1), (3), and (5), and in terms of jobs in columns (2), (4), and (6).

Instrumented triple differences. Table D3 reports the estimations obtained when the increases in labor costs at the minimum wage are instrumented by increases in the corresponding wages paid-out to workers, net of the potentially endogenous payroll taxes (see Section 3.1 for rationale). This instrument is introduced both in level and in interaction.[77] The main take-aways of Table D3 are the following:

- Labor costs in Panel (A): results are similar to the un-instrumented version above. In particular, the main result that labor costs increase at the bottom hierarchical layer is preserved and significant, but the associated magnitude is reduced as compared with Table 1.
- Relative sizes of hierarchical layers in Panel (B): in the instrumented triple-difference specifications, we do obtain that low-wage firms contract more in terms of hours worked when exposed to increases in the minimum wage, as in the baseline version presented in the main text. The pattern is however only statistically significant in the pooled specification, estimated across all firms and hierarchical layers.
- Number of hierarchical layers in Panel (C): results are broadly similar to the un-instrumented version presented above (and to the baseline specification presented in the main text), although they tend to be slightly amplified quantitatively.
- Probability of training programs in Panel (D): unfortunately, none of the interaction terms remain significant, probably due to the reduced sample size (with training data only available for a third of the sample firms).
- Firms' sizes in Panel (E): in the instrumented, triple difference specifications, we obtain that firms with a high share of minimum wage workers contract more in terms of hours worked and jobs when exposed to increases in the minimum wage (than firms with a lower share of minimum wage workers). This effect is statistically significant and the implied magnitudes are similar to those of the simpler difference-in-difference specifications that are reported in the main text. There is no impact on sales or value added, as in the main text.
- Firms' productivities in Panel (F): results are broadly similar to the un-instrumented version above (and thus confirm results obtained with the baseline, double-difference specification which is presented in the main text). They tend to be slightly amplified in terms of magnitude.

[77] The value of the Stock-Yogo Cragg-Donald Wald F-statistic is 4.6e+04.

Table D3: Impact of Minimum Wage Increases on Firms' Outcomes: Instrumented Triple Differences

(A) Impact on Labor Costs

	(1)	(2)	(3)	(4)	(5)
Indicator:		Δ ln total hourly labor cost			
	Total	Layer 0	Layer 1	Layer 2	Layer 3
Mean:	0.095	0.110	0.097	0.073	0.106
Δ ln hourly cost	0.257***	0.076**	0.012	0.371***	0.128
at GMR	(0.034)	(0.038)	(0.070)	(0.089)	(0.218)
Share LW hours	0.115***	0.078***	0.051	0.160***	0.152
	(0.012)	(0.011)	(0.035)	(0.058)	(0.134)
Interaction	0.006	0.413***	-0.292	-1.244*	-1.125
	(0.162)	(0.155)	(0.453)	(0.691)	(1.502)
# observations	55,344	55,344	49,479	33,202	8,484

(B) Impact on the Relative Size of Hierarchical Layers

	(1)	(2)	(3)	(4)	(5)
Indicator:			Δ ln hours		
	Total	Layer 0	Layer 1	Layer 2	Layer 3
Mean:	0.014	0.016	-0.007	0.041	-0.043
Δ ln hourly cost	-0.222***	-0.747***	0.095	-0.281	-0.039
at GMR	(0.082)	(0.140)	(0.170)	(0.200)	(0.382)
Share LW hours	0.004	-0.099**	0.144*	-0.166	0.125
	(0.030)	(0.040)	(0.076)	(0.109)	(0.222)
Interaction	-1.073**	-0.353	-0.919	2.214	-1.635
	(0.474)	(0.597)	(1.041)	(1.382)	(2.545)
# observations	55,344	55,344	49,479	33,202	8,484

(C) Impact on the Number of Hierarchical Layers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Indicator:	Δ number hierarchical Layers	Probability of increase	Probability of decrease	Probability of stability	Δ Proba 3 layers	Δ Proba at least 2 layers	Δ Proba at least 1 layers
Mean :	-0.036	0.165	0.193	0.642	-0.015	-0.013	-0.008
Δ ln hourly cost	0.029	-0.510***	-0.469***	0.979***	-0.283***	0.195**	0.117**
at GMR	(0.156)	(0.087)	(0.095)	(0.112)	(0.103)	(0.089)	(0.050)
Share LW hours	0.223***	0.029	-0.154***	0.125***	0.024	0.122***	0.076**
	(0.059)	(0.034)	(0.033)	(0.040)	(0.027)	(0.037)	(0.031)
Interaction	-1.911**	-0.004	1.511***	-1.507***	0.020	-1.150*	-0.781**
	(0.818)	(0.469)	(0.458)	(0.560)	(0.416)	(0.509)	(0.387)
# observations	55,344	55,344	55,344	55,344	55,344	55,344	55,344

(D) Impact on Training across Layers

Indicator:	Δ Probability Training Hours in L_0			
	All firms	1-layer firms	2-layer firms	3-layer firms
Shares -1/+1:	0.07/0.25	0.08/0.20	0.07/0.25	0.06/0.28
Δ ln hourly cost	0.945***	-0.058	0.553*	1.489***
at GMR	(0.211)	(0.711)	(0.292)	(0.447)
Share LW hours	-0.210**	-0.169	-0.301**	0.045
	(0.091)	(0.183)	(0.142)	(0.300)
Interaction	1.470	3.223	2.683	-0.989
	(1.144)	(2.556)	(1.706)	(3.340)
# observations	17,406	2,494	9,688	4,569
Indicator:	Δ Probability Training Hours in Managerial Layers			
	All firms	1-layer firms	2-layer firms	3-layer firms
Shares -1/+1:	0.07/0.24	0.08/0.16	0.07/0.25	0.06/0.27
Δ ln hourly cost	1.107***	0.665	0.906***	1.139***
at GMR	(0.206)	(0.649)	(0.289)	(0.427)
Share LW hours	-0.109	-0.049	-0.138	0.100
	(0.089)	(0.172)	(0.137)	(0.300)
Interaction	-0.080	0.146	0.763	-1.684
	(1.112)	(2.334)	(1.664)	(3.324)
# observations	17,406	2,494	9,688	4,569

Table D3: Impact of Minimum Wage Increases... : Instrumented Triple Differences, Continued

(E) Impact on Firms' Size

Indicator:	(1) Δ ln hours	(2) Δ ln # jobs	(3) Δ ln VA	(4) Δ ln Sales	(5) Δ ln hourly cost at GMR	(6) Interaction (centered)
Mean:	0.014	0.009	0.084	0.106	0.072	0.009
Δ ln hourly cost at GMR	-0.222***	-0.274***	-0.100	-0.024		
	(0.082)	(0.083)	(0.099)	(0.082)		
Share LW hours	0.004	0.028	-0.028	-0.013	0.016***	-0.069***
	(0.030)	(0.031)	(0.036)	(0.030)	(0.003)	(0.001)
Interaction	-1.073**	-1.320***	0.166	-0.103		
	(0.474)	(0.480)	(0.520)	(0.437)		
Δ ln GMR					0.1984***	0.010
					(0.010)	(0.001)
Δ ln GMR × Share LW					-0.212***	1.797***
					(0.040)	(0.020)
# observations	55,344	55,344	55,344	55,344	55,344	55,344

(F) Impact on Productivity

Indicator:	(1) Δ ln Labor	(2)	(3)	(4) Δ ln TFP	(5)	(6)
Measurement of TFP:	productivity		Accounting		Lev. Pet.	
Measurement of labor:	hours	jobs	hours	jobs	hours	jobs
Mean:	0.070	0.075	0.023	0.025	0.023	0.022
Δ ln hourly cost at GMR	0.122	0.174*	-0.003	0.001	0.047	0.043
	(0.085)	(0.094)	(0.035)	(0.037)	(0.035)	(0.036)
Share LW hours	-0.032	-0.056*	-0.018	-0.025*	-0.002	-0.001
	(0.030)	(0.034)	(0.012)	(0.013)	(0.014)	(0.016)
Interaction	1.239***	1.486***	0.526***	0.603***	0.216	0.185
	(0.419)	(0.467)	(0.176)	(0.184)	(0.196)	(0.216)
# observations	55,344	55,344	53,436	53,436	53,436	53,436

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the 1-digit sector-commuting zone level in parentheses. Regression results from equation (4), estimated by 2SLS, using the increase in legal GMRs as an instrumental variable for the evolution of the corresponding total labor costs (and its interaction with share of low-wage workers as an instrumental variable for the interaction between total labor costs at GMR and the share of low-wage workers). All specifications contain 4-digit industry fixed effects and commuting zone fixed effects. In Panel (B), the composition of the estimation sample evolves across columns. In the first two columns, all sample firms are included. In column (3), only firms exhibiting at least 2 layers (Layer 0 and Layer 1) in 2003 and 2006 are included in the estimation sample. Similarly, in columns (4) to (5), only firms exhibiting at least 3 or 4 layers respectively, both in 2003 and 2006, are included in the estimation sample. Panel (E), Columns (5) and (6) report the first stage regressions. Value of the Stock-Yogo Cragg-Donald Wald F-statistic: 4.6e+04. In Panel (F) Columns (3) and (4), the productivity indices have been constructed using the accounting method, with the elasticities of value added to production factors (labor and capital) being evaluated by their share of compensation in value added. In Columns (5) and (6), the estimation method proposed by Levinsohn and Petrin (2003) is implemented using Wooldridge (2009) and Petrin and Levinsohn (2012). Labor is measured in terms of hours in Columns (1), (3), and (5), and in terms of jobs in columns (2), (4), and (6).

E Details and Proofs of the Theoretical Model

This Appendix contains the details of the equilibrium determination for Section 4.

E.1 Supply: Cost Minimization Programs

We consider in sequence the cases of firms with 0, 1, and 2 or more layers, respectively and solve the problem embodied in Equation 12.

Firms with 0 layers (self-employed entrepreneurs). The expected output of a self-employed entrepreneur with knowledge level z_0^0 is:

$$q \leq A(1 - e^{-\lambda z_0^0}) \Longleftrightarrow \frac{1}{\lambda} \ln \frac{A}{A - q} \leq z_0^0 \tag{23}$$

It is always optimal to saturate this output constraint, such that the cost function is:

$$\forall q \in [0; A[, \qquad C_0(z_0^0; q, k) = k(cz_0^0 + 1) = k\left(\frac{c}{\lambda} \ln \frac{A}{A - q} + 1\right) \tag{24}$$

Firms with 1 layer. If $n_1^0 \leq \frac{e^{\lambda z_1^0}}{h}$, the manager has enough time to deal with with all the problems that are transferred to her and the expected output of a firm with one layer is:

$$q \leq n_1^0 A(1 - e^{-\lambda z_1^1}) \Longleftrightarrow \frac{1}{\lambda} \ln \frac{n_1^0 A}{n_1^0 A - q} \leq z_1^1 \tag{25}$$

Notice that n_1^0 and z_1^1 have to be strictly positive to achieve strictly positive levels of output (while $n_1^1 = 1$ by definition of a 1-layer firm). The cost function to be minimized is:

$$\forall q \in \mathbb{R}^+, \qquad C_1(n_1^0, z_1^0, z_1^1; q, k) = k\left(\underbrace{cz_1^1 + 1}_{\text{Entrepreneur}} + \underbrace{n_1^0(cz_1^0 + 1)}_{\text{Production workers}}\right), \tag{26}$$

subject to the following three additional constraints:

$$q \leq n_1^0 A(1 - e^{-\lambda z_1^1}) \tag{27}$$

$$n_1^0 \leq \frac{e^{\lambda z_1^0}}{h} \tag{28}$$

$$0 \leq z_1^0 \tag{29}$$

The full Lagrangian associated with this program is thus:

$$\mathcal{L}_1(n_1^0, z_1^0; q, k) = k\left[cz_1^1 + 1 + n_1^0(cz_1^0 + 1)\right] - \phi\left[n_1^0 A(1 - e^{-\lambda z_1^1}) - q\right] - \theta\left[\frac{e^{\lambda z_1^0}}{h} - n_1^0\right] - \eta z_1^0 \tag{30}$$

The Kuhn-Tucker conditions are:

$$\frac{\partial \mathcal{L}_1}{\partial n_1^0} = k(cz_1^0+1) - \phi A(1-e^{-\lambda z_1^1}) + \theta = 0 \tag{31}$$

$$\frac{\partial \mathcal{L}_1}{\partial z_1^0} = kcn_1^0 - \frac{\theta \lambda e^{\lambda z_1^0}}{h} - \eta = 0 \tag{32}$$

$$\frac{\partial \mathcal{L}_1}{\partial z_1^1} = kc - \phi A n_1^0 \lambda e^{-\lambda z_1^1} = 0 \tag{33}$$

$$0 = \phi \left[n_1^0 A(1-e^{-\lambda z_1^1}) - q \right] \tag{34}$$

$$0 = \theta \left[\frac{e^{\lambda z_1^0}}{h} - n_1^0 \right] \tag{35}$$

$$0 = \eta z_1^0 \tag{36}$$

$$\theta \geq 0, \ \phi \geq 0, \ \eta \geq 0 \tag{37}$$

The marginal cost is thus simply $\frac{\partial \mathcal{L}_1}{\partial q} = \phi$. Condition 33 ensures that it is strictly positive, such that Equation 25 is necessarily saturated at the optimum, as in the 0-layer case (this is due to Condition 34).

In the case where $z_1^0 > 0$ (ie., production workers at the bottom are able to deal with at least some of the most common problems), then Condition 36 implies that $\eta = 0$. Plugging this into Condition 32 implies $\theta > 0$ and therefore $n_1^0 = \frac{e^{\lambda z_1^0}}{h}$ because of Condition 35. This case corresponds to the baseline Caliendo and Rossi-Hansberg (2012) specification.

However, in our setting, it is possible to get $z_1^0 = 0$ for some values of q.[78] Because of Condition 36, this relaxes the condition on η which is no longer necessarily 0 ($\eta \geq 0$). This in turn relaxes the constraint on θ, which is not necessarily strictly positive in Condition 32, such that cases where $\theta = 0$ are possible in Conditions 31, 32 and 35. This implies in particular that n_1^0 can become strictly smaller than $\frac{e^{\lambda z_1^0}}{h} = \frac{1}{h}$ (Condition 35 for $z_1^0 = 0$).

Firms with 2 or more layers. If Equations 8 and 9 hold, all managers at all layers have enough time to deal with with all the problems that are transferred to them and the expected output of a firm that is organized in L layers is ($L \geq 2$):

$$q \leq n_L^0 A(1-e^{-\lambda z_L^L}) \iff \frac{1}{\lambda} \ln \frac{n_L^0 A}{n_L^0 A - q} \leq z_L^L \tag{38}$$

The cost function to be minimized is:

$$\forall q \in \mathbb{R}^+, \quad C_L((n_L^l)_l, (z_L^l)_l; q, k) = k \left(\underbrace{cz_L^L + 1}_{\text{Entrepreneur}} + \underbrace{\sum_{l=1}^{L-1} n_L^l (cz_L^l + 1)}_{\text{Intermediate managers}} + \underbrace{n_L^0 (cz_L^0 + 1)}_{\text{Production workers}} \right), \tag{39}$$

$$= k \left(\sum_{l=0}^{L} n_L^l (cz_L^l + 1) \right) \tag{40}$$

[78] Caliendo and Rossi-Hansberg (2012) impose a parametric assumption in order to rule out this second case, strictly speaking for $L \geq 2$ (see their Appendix 2). With cumulative knowledge, Appendix E.2.2 (proof in the case where $L \geq 2$ and $z_L^0 = 0$) shows that this parametric assumption takes the following form: $c \leq \lambda h$. This condition ensures that $z_L^0 > 0$ because it becomes less costly to train production workers at the bottom than to transfer problems upwards, although managers have a productivity advantage as long as $h < 1$ since they can deal with $\frac{1}{h} > 1$ problems per unit of time instead of just one. However, it does not hold in our calibrations. In Section 5, we show that, in our main calibration, $h = 0.194$ and $c = 0.220$ (with λ normalized to 1). The Kuhn-Tucker conditions imply that the choices $z_1^0 = 0$, $n_1^0 < \frac{1}{h}$ are optimal if the target to attain in terms of output is relatively small: $q < A \frac{\lambda}{\lambda + ch} \frac{1}{h}$.

subject to the following constraints:

$$q \leq n_L^0 A(1 - e^{-\lambda z_L^L}) \quad (41)$$

$$\forall l \in \{0, ..., L-1\}, \quad n_L^0 \leq \frac{e^{\lambda z_L^l}}{h} n_L^{l+1} \quad (42)$$

$$(n_L^L = 1) \quad (43)$$

$$0 \leq z_L^0 \quad (44)$$

The full Lagrangian associated with this program is thus:

$$\mathcal{L}_L\left((n_L^l)_l, (z_L^l)_l; q, k\right) = k\left[\sum_{l=0}^{L} n_L^l (cz_L^l + 1)\right] - \phi\left[n_L^0 A(1 - e^{-\lambda z_L^L}) - q\right] - \sum_{l=0}^{L-1} \theta_l \left[\frac{e^{\lambda z_L^l}}{h} n_L^{l+1} - n_L^0\right] - \eta z_L^0 \quad (45)$$

The Kuhn-Tucker conditions are:

$$\frac{\partial \mathcal{L}_L}{\partial n_L^0} = k(cz_L^0 + 1) - \phi A(1 - e^{-\lambda z_L^L}) + \sum_{l=0}^{L-1} \theta_l = 0 \quad (46)$$

$$\frac{\partial \mathcal{L}_L}{\partial z_L^0} = kcn_L^0 - \theta_0 \frac{\lambda e^{\lambda z_L^0}}{h} n_L^1 - \eta = 0 \quad (47)$$

$$\frac{\partial \mathcal{L}_L}{\partial z_L^L} = kc - \phi A n_L^0 \lambda e^{-\lambda z_L^L} = 0 \quad (48)$$

$$\forall l \in \{1, ..., L-1\}, \quad \frac{\partial \mathcal{L}_L}{\partial n_L^l} = k(cz_L^l + 1) - \theta_{l-1} \frac{e^{\lambda z_L^{l-1}}}{h} = 0 \quad (49)$$

$$\forall l \in \{1, ..., L-1\}, \quad \frac{\partial \mathcal{L}_L}{\partial z_L^l} = kcn_L^l - \theta_l \frac{\lambda e^{\lambda z_L^l}}{h} n_L^{l+1} = 0 \quad (50)$$

$$0 = \phi\left[n_L^0 A(1 - e^{-\lambda z_L^L}) - q\right] \quad (51)$$

$$\forall l \in \{0, ..., L-1\}, \quad 0 = \theta_l \left[\frac{e^{\lambda z_L^l}}{h} n_L^{l+1} - n_L^0\right] \quad (52)$$

$$0 = \eta z_L^0 \quad (53)$$

$$\forall l \in \{0, ..., L-1\}, \quad \theta_l \geq 0, \ \phi \geq 0, \ \eta \geq 0 \quad (54)$$

Again, the marginal cost is simply $\frac{\partial \mathcal{L}_L}{\partial q} = \phi$. Equation 48 ensures that this marginal cost is strictly positive such that Equation 38 is saturated, as in the previous (0 and 1-layer) cases: this is due to Condition 51. Furthermore, Equations 49 and 50 insure that all θ_l ($\forall l \in \{0, ..., L-1\}$) are strictly positive, which implies that the set of constraints in Equation 42 are always saturated, even when $z_L^0 = 0$. Interestingly, this contrasts with the 1-layer case.

For the results and proofs which follow, it is convenient to combine the previous set of conditions into the following additional relations:

$$
\begin{aligned}
& & n_L^0 &= \frac{e^{\lambda z_L^{L-1}}}{h} & (55)\\
\forall l \in \{1, ..., L-1\}, & & n_L^l &= e^{\lambda(z_L^{L-1}-z_L^{l-1})} & (56)\\
& & 0 &= kcn_L^0 - \lambda n_L^1 k(cz_L^1+1) - \eta & (57)\\
\forall l \in \{1, ..., L-2\}, & & 0 &= kcn_L^l - \lambda n_L^{l+1} k(cz_L^{l+1}+1) & (58)\\
& & 0 &= kcn_L^{L-1} + \sum_{l=0}^{L-1} \lambda n_L^l k(cz_L^l+1) - \phi A(1-e^{-\lambda z_L^L})\lambda n_L^0 & (59)\\
& & 0 &= kc \underbrace{n_L^L}_{1} - \phi A \lambda e^{-\lambda z_L^L} n_L^0 & (60)
\end{aligned}
$$

Equation 55 comes from the bracket being saturated in Equation 52 for $l = L-1$ (and observing that $n_L^L = 1$). Equation 56 comes from the bracket being saturated in Equation 52 and Equation 55. Equation 57 combines Equation 47 with Equation 49 for $l = 1$. Equation 58 combines Equation 50 with Equation 49 (with a rank shift). Equation 59 combines Equation 50 for $l = L-1$ with Equation 46 (to retrieve θ_{L-1}) and Equation 49 (for all other θ_l). Equation 60 simply repeats Equation 48.

E.2 Additional Results from the Cost Minimization Programs

In their mainly theoretical contribution, Caliendo and Rossi-Hansberg (2012) rule out the case in which $z_L^0 = 0$ by imposing a convenient restriction on structural parameters (see footnote 78), which also has the benefit of simplifying several of their proofs. Unfortunately, this parametric restriction is rejected in the French data in our main calibration (which implies that some French firms are likely to feature $\eta > 0$ and $z_L^0 = 0$ for at least some organizational forms L). As a consequence, many derivations in Caliendo and Rossi-Hansberg (2012) no longer hold. Our assumption of skill overlap (aka cumulative knowledge at the worker level) also alters some of their results. It is however possible to derive the set of results which follows. They apply to regions of the firms' optimization problem, where they do not find it optimal to switch to a different organization (or to a different regime within a given organizational form). Formally, this implies that we consider marginal deviations in which the precise set of saturated constraints described in Section E.1 remains the same (or equivalently, in which the set of strictly positive Lagrange multipliers remains the same).

E.2.1 Characteristics of the Cost Function for a Given Number of Layers L

In terms of the cost function, our set-up features a result that is similar to part 2 of Proposition 1 from Caliendo and Rossi-Hansberg (2012).

Proposition 3. *Given a number of layers L, the marginal cost is always positive and so the cost curve is strictly increasing in q.*
Furthermore, the cost function remains homogeneous of degree one in k, which implies that the cost-minimizing firm organization given q is insensitive to k.

The first part of Proposition 3 results from the fact that, for $L \geq 1$, the first-order condition for z_L^L shows that ϕ must be positive; and for $L = 0$, the derivative of the cost function 24 is positive. The second part of Proposition 3 results from the fact that all Lagrange multipliers are proportional to k, such that k drops from all first order conditions. In our model with cumulative knowledge, and in contrast with the specification proposed in Caliendo and Rossi-Hansberg (2012), however, the marginal cost is not necessarily increasing in q: see Proposition 4. This result of non-monotonicity when knowledge is cumulative is likely to render the management of production costs in organizations more "complex" for firms.

E.2.2 Sensitivity Analysis with Respect to q of Firms' Optimal Employment and Skill Decisions

As in part 1 of Proposition 1 from Caliendo and Rossi-Hansberg (2012), we can evaluate how the values of z^l_L, n^l_L, and ϕ vary with q. However the proof is somewhat different with skill overlap such that the proposition below holds under a different set of sufficient conditions:

Proposition 4. *For a given number of layers L, we get:*

$$\frac{dz^l_L}{dq} \geq 0 \quad \forall\ l \in \{0, 1, ..., L\}$$

$$\frac{dn^l_L}{dq} \geq 0 \quad \forall\ l \in \{0, 1, ..., L-1\}$$

$$n^L_L = 1 \text{ by design, such that } \frac{dn^L_L}{dq} = 0$$

These results always hold for $L = 0$ (self-employed entrepreneurs), and under the following (sufficient) conditions,[79] *depending on firms' number of layers: $h < 1$ in firms with $L = 1$, or $h \leq 0.25$ in firms with $L \geq 2$ and $\eta = 0$ (thus $z^0_L \geq 0$), or $h \leq \frac{c}{\lambda} \leq 0.25$ in firms with $L \geq 2$ and $\eta > 0$ (thus $z^0_L = 0$).*
Furthermore, the previous inequalities are strict, except for n^0_L and z^0_L when the latter is constrained at zero ($\eta > 0$).

Proposition 5. *Under the same parameter restrictions as Proposition 4, $\frac{d\phi}{dq}$ can be positive or negative, depending on the number of layers and the parameter values:*

- *$\frac{d\phi}{dq} > 0$ for self-employed entrepreneurs ($L = 0$) as well as in 1-layer firms ($L = 1$) as long as $\theta > 0$, such that the employment of the bottom layer (n^0_L) reaches its upper bound (Equation 28);*
- *$\frac{d\phi}{dq} < 0$ in 1-layer firms ($L = 1$) where $\theta > 0$, such that n^0_L does not necessarily reach the upper bound in Equation 28;*
- *$\frac{d\phi}{dq}$ can take any sign in firms with more than 2 layers ($L \geq 2$).*

Proof for the case $L = 0$

Proof. In this class of firms, output is a monotonic function of z^0_0: $q = A\left(1 - e^{-\lambda z^0_0}\right)$, such that $z^0_0 = \frac{1}{\lambda} \ln \frac{A}{A-q}$ and $C_0(q) = k(cz^0_0 + 1)$. It is therefore immediate that $\frac{dz^0_0}{dq} > 0$ and $\frac{d\phi}{dq} > 0$. Note that in this case, z^0_0 can adjust, but not n^0_0. □

Proof for the case $L = 1$ In this class of firms, there are three different cases to consider:[80]

(a) $\eta = 0$ and $\theta > 0$, which implies $z^0_1 \geq 0$ and $n^0_1 = \frac{e^{\lambda z^0_1}}{h} > \frac{1}{h}$

(b) $\eta > 0$ and $\theta > 0$, which implies $z^0_1 = 0$ and $n^0_1 = \frac{e^{\lambda z^0_1}}{h} = \frac{1}{h}$

(c) $\eta > 0$ and $\theta = 0$, which implies $z^0_1 = 0$ and $n^0_1 \leq \frac{1}{h}$

[79] Notice that these conditions on the exogenous parameters are increasingly binding. Low values of h and $\frac{c}{\lambda}$ correspond to efficient communication or information technologies, respectively.

[80] Indeed, the case where $\eta = 0$ and $\theta = 0$ would imply $n^0_1 = 0$ from Equation 32, which is impossible.

Proof for the case $L = 1$, $\eta > 0$ and $\theta > 0$, which implies $z_1^0 = 0$ and $n_1^0 = \frac{e^{\lambda z_1^0}}{h} = \frac{1}{h}$. The proof for case (b) is analogous to the proof for $L = 0$. Indeed, we get that output simply takes the following form: $q = \frac{A}{h}(1 - e^{-\lambda z_1^1})$, such that $z_1^1 = \frac{1}{\lambda} \ln \frac{A/h}{A/h-q}$ and $C_1(q) = k(cz_1^1 + 1 + \frac{1}{h})$. It is therefore immediate that $\frac{dz_1^1}{dq} > 0$ and $\frac{d\phi}{dq} > 0$. Note that in this case, z_1^1 can adjust, but neither z_1^0 nor n_1^0.[81] □

Proof for the case $L = 1$, $\eta = 0$ and $\theta > 0$, which implies $z_1^0 \geq 0$ and $n_1^0 = \frac{e^{\lambda z_1^0}}{h}$. The proof for case (a) is somewhat more complex. The Kuhn-Tucker conditions can be used to deliver the following set of relations:[82]

$$n_1^0 kc + \lambda n_1^0 k(cz_1^0 + 1) = \phi\lambda q \tag{61}$$
$$\phi A\lambda e^{-\lambda z_1^1} n_1^0 = kc \tag{62}$$
$$A(1 - e^{-\lambda z_1^1})n_1^0 = q \tag{63}$$

Equation 61 combines Equations 31 and 32 and recognizes the values of q and n_1^0 in the (saturated) brackets of Equations 34 and 35, respectively. Equation 62 corresponds to the saturated constraint in Equation 35, while Equation 63 corresponds to the saturated constraint in Equation 34.
Differentiating each with respect to q, we obtain:

$$\phi + q\frac{d\phi}{dq} = \left(2n_1^0 kc + \lambda n_1^0 k(cz_1^0 + 1)\right)\frac{dz_1^0}{dq} \tag{64}$$
$$\frac{d\phi}{dq} = \phi\lambda\left(\frac{dz_1^1}{dq} - \frac{dz_1^0}{dq}\right) \tag{65}$$
$$1 = A\lambda n_1^0 \frac{dz_1^0}{dq} + A\lambda e^{-\lambda z_1^1} n_1^0 \left(\frac{dz_1^1}{dq} - \frac{dz_1^0}{dq}\right) \tag{66}$$

Combining all three of these, we find:

$$(kc + q\phi\lambda)\left(\frac{dz_1^1}{dq} - \frac{dz_1^0}{dq}\right) = \left(2n_1^0 kc + \lambda n_1^0 k(cz_1^0 + 1) - \phi A\lambda n_1^0\right)\frac{dz_1^0}{dq} \tag{67}$$

Multiplying Equation 62 by Equation 63 and subtracting kc, we get that $\phi A\lambda n_1^0 = \phi\lambda q + kc$, which can be plugged into the previous equation to deliver:

$$(kc + q\phi\lambda)\left(\frac{dz_1^1}{dq} - \frac{dz_1^0}{dq}\right) = (n_1^0 - 1)kc\frac{dz_1^0}{dq}. \tag{68}$$

Whenever $h < 1$, $n_1^0 > 1$ such that $\left(\frac{dz_1^1}{dq} - \frac{dz_1^0}{dq}\right)$ takes the same sign as $\frac{dz_1^0}{dq}$. Given Equation 66, that sign must be positive, which implies $\frac{dz_1^1}{dq} > \frac{dz_1^0}{dq} > 0$, and therefore $\frac{d\phi}{dq} > 0$ and $\frac{dn_1^0}{dq} > 0$. □

Proof for the case $L = 1$, $\eta > 0$ and $\theta = 0$, which implies $z_1^0 = 0$ and $n_1^0 < \frac{1}{h}$. Lastly, consider case (c). Again,

[81] Working on the restrictions implied by the two contraints $\eta > 0$ and $\theta > 0$, it is possible to characterize the domain under which this regime is likely to be optimal:

$$\frac{A}{h}\cdot\left[\frac{\frac{kc}{\theta\lambda} + k}{\frac{kch}{lambda} + \frac{kc}{\theta\lambda} + k}\right] > q > \frac{\lambda A}{h(ch + \lambda)}$$

[82] Solving this set of equations in z_1^0, we obtain the following non-linear equation, which can't be solved in closed form but delivers a unique solution:

$$k\left(cz_1^0 + \frac{c}{\lambda} + 1\right) = \frac{k^2 hcq}{\lambda e^{\lambda z_1^0}(Ae^{\lambda z_1^0} - hq)}$$

the Kuhn-Tucker conditions can be used to deliver the following set of relations:[83]

$$\begin{aligned}\phi A(1-e^{-\lambda z_1^1}) &= k && (69)\\ \phi A\lambda e^{-\lambda z_1^1} n_1^0 &= kc && (70)\\ A(1-e^{-\lambda z_1^1})n_1^0 &= q && (71)\end{aligned}$$

Equation 69 simply corresponds to Equation 31 with $\theta = 0$ and $z_1^0 = 0$. Equation 70 corresponds to the saturated constraint in Equation 35, while Equation 71 corresponds to the saturated constraint in Equation 34. Differentiating each of the previous relations with respect to q, we obtain:

$$\begin{aligned}(1-e^{-\lambda z_1^1})\frac{d\phi}{dq} &= -\phi\lambda e^{-\lambda z_1^1}\frac{dz_1^1}{dq} && (72)\\ \phi\lambda n_1^0\frac{dz_1^1}{dq} - \phi\frac{dn_1^0}{dq} &= n_1^0\frac{d\phi}{dq} && (73)\\ A(1-e^{-\lambda z_1^1})\frac{dn_1^0}{dq} + A\lambda e^{-\lambda z_1^1} n_1^0\frac{dz_1^1}{dq} &= 1 && (74)\end{aligned}$$

Combining Equations 72 and 73, we obtain:

$$\lambda n_1^0\frac{dz_1^1}{dq} = (1-e^{-\lambda z_1^1})\frac{dn_1^0}{dq} \tag{75}$$

Therefore, $\frac{dz_1^1}{dq}$ and $\frac{dn_1^0}{dq}$ must take the same sign, and Equation 74 implies that that sign must be positive: $\frac{dz_1^1}{dq} > 0$ and $\frac{dn_1^0}{dq} > 0$. Finally, Equation 72 implies that $\frac{d\phi}{dq} < 0$. □

Proof for the case $L \geq 2$ In this class of firms, there are two different cases to consider:

(a) $\eta = 0$, which implies $z_L^0 \geq 0$

(b) $\eta > 0$, which implies $z_L^0 = 0$

Proof for the case $L \geq 2$ and $\eta = 0$, which implies $z_L^0 \geq 0$. In this regime, $z_L^0 > 0$ implies $\eta = 0$ in Equation 54. Plugging Equations 55 and 56 into Equations 57 to 60, the latter simplify to:[84]

$$\begin{aligned}\lambda h(cz_L^1+1) &= ce^{\lambda z_L^0} && (76)\\ \forall l \in \{1,...,L-2\},\qquad \lambda(cz_L^{l+1}+1) &= ce^{\lambda(z_L^l - z_L^{l-1})} && (77)\\ \sum_{l=0}^{L-1} n_L^l kc + \lambda n_L^0 k(cz_L^0+1) &= \phi\lambda q && (78)\\ \phi A\lambda e^{-\lambda z_L^L} n_L^0 &= kc && (79)\\ A(1-e^{-\lambda z_L^L})n_L^0 &= q && (80)\end{aligned}$$

[83]In this regime, the firm's optimization problem can be solved for in closed form:

$$\begin{aligned}e^{-\lambda z_1^1} &= 1 + \frac{\lambda q + \sqrt{\lambda^2 q^2 + 4ac\lambda q}}{2Ac}\\ \phi A &= \frac{k}{1-e^{-\lambda z_1^1}}\\ n_1^0 &= \frac{q}{A(1-e^{-\lambda z_1^1})}\end{aligned}$$

[84]It should be noticed that Equation 77 does not operate when $L = 2$. However, in this regime, Equations 76, 78, 79 and 80 are sufficient to derive the same results as in the case where $L > 2$ and Equation 77 holds.

Differentiating each of the previous equations with respect to q, we obtain:

$$n_L^0 \frac{dz_L^0}{dq} = n_L^1 \frac{dz_L^1}{dq} \tag{81}$$

$$\forall l \in \{1, ..., L-2\}, \quad n_L^l \left(\frac{dz_L^l}{dq} - \frac{dz_L^{l-1}}{dq} \right) = n_L^{l+1} \frac{dz_L^{l+1}}{dq} \tag{82}$$

$$\phi \lambda q \frac{dz_L^{L-1}}{dq} + kcn_L^{L-1} \left(\frac{dz_L^{L-1}}{dq} - \frac{dz_L^{L-2}}{dq} \right) = \phi + q \frac{d\phi}{dq} \tag{83}$$

$$\phi \lambda \left(\frac{dz_L^L}{dq} - \frac{dz_L^{L-1}}{dq} \right) = \frac{d\phi}{dq} \tag{84}$$

$$A\lambda n_L^0 \frac{dz_L^{L-1}}{dq} + A\lambda n_L^0 e^{-\lambda z_L^L} \left(\frac{dz_L^L}{dq} - \frac{dz_L^{L-1}}{dq} \right) = 1 \tag{85}$$

Notice that the derivation of Equation 83 above is somewhat indirect and involves the following steps:

$$kc \sum_{l=0}^{L-1} \frac{dn_L^l}{dq} + \lambda k (cz_L^0 + 1) \frac{dn_L^0}{dq} + \lambda n_L^0 kc \frac{dz_L^0}{dq} = \phi\lambda + \lambda q \frac{d\phi}{dq}$$

$$\therefore kc \sum_{l=1}^{L-1} \lambda n_L^l \left(\frac{dz_L^{L-1}}{dq} - \frac{dz_L^{l-1}}{dq} \right) + kc\lambda n_L^0 \frac{dz_L^{L-1}}{dq} + \lambda^2 k (cz_L^0 + 1) n_L^0 \frac{dz_L^{L-1}}{dq} + \lambda n_L^0 kc \frac{dz_L^0}{dq} = \phi\lambda + \lambda q \frac{d\phi}{dq}$$

$$\therefore \phi\lambda q \frac{dz_L^{L-1}}{dq} - kc \sum_{l=1}^{L-1} n_L^l \frac{dz_L^{l-1}}{dq} + n_L^0 kc \frac{dz_L^0}{dq} = \phi + q \frac{d\phi}{dq}$$

$$\therefore \phi\lambda q \frac{dz_L^{L-1}}{dq} - kc \sum_{l=1}^{L-2} \left(n_L^l \frac{dz_L^l}{dq} - n_L^{l+1} \frac{dz_L^{l+1}}{dq} \right) - kcn_L^{L-1} \frac{dz_L^{L-2}}{dq} + n_L^0 kc \frac{dz_L^0}{dq} = \phi + q \frac{d\phi}{dq}$$

$$\therefore \phi\lambda q \frac{dz_L^{L-1}}{dq} - kc \left(n_L^1 \frac{dz_L^1}{dq} - n_L^{L-1} \frac{dz_L^{L-1}}{dq} \right) - kcn_L^{L-1} \frac{dz_L^{L-2}}{dq} + n_L^0 kc \frac{dz_L^0}{dq} = \phi + q \frac{d\phi}{dq}$$

$$\therefore \phi\lambda q \frac{dz_L^{L-1}}{dq} + kcn_L^{L-1} \left(\frac{dz_L^{L-1}}{dq} - \frac{dz_L^{L-2}}{dq} \right) = \phi + q \frac{d\phi}{dq}$$

In the previous set of equations, the third line uses the fact that $\sum_{l=0}^{L-1} n_L^l kc + \lambda n_L^0 k(cz_L^0 + 1) = \phi\lambda q$ from Equation 78. The fourth line uses the fact that $n_L^l \frac{dz_L^{l-1}}{dq} = n_L^l \frac{dz_L^l}{dq} - n_L^{l+1} \frac{dz_L^{l+1}}{dq}$ for $l \in \{1, ..., L-2\}$ from Equation 82, and the fifth line uses the fact that $n_L^0 \frac{dz_L^0}{dq} = n_L^1 \frac{dz_L^1}{dq}$ from Equation 81.

Returning to Equations 83 to 85 and combining them, we obtain:

$$\phi\lambda q \frac{dz_L^{L-1}}{dq} + kcn_L^{L-1} \left(\frac{dz_L^{L-1}}{dq} - \frac{dz_L^{L-2}}{dq} \right) = \phi A \lambda n_L^0 \frac{dz_L^{L-1}}{dq} + kc \left(\frac{dz_L^L}{dq} - \frac{dz_L^{L-1}}{dq} \right) + \phi\lambda q \left(\frac{dz_L^L}{dq} - \frac{dz_L^{L-1}}{dq} \right) \tag{86}$$

and using the fact that $\phi\lambda q = \phi A \lambda n_L^0 - kc$ (as implied by Equation 60 where q is replaced with its value in Equation 51), this simplifies to:

$$kcn_L^{L-1} \left(\frac{dz_L^{L-1}}{dq} - \frac{dz_L^{L-2}}{dq} \right) = kc \frac{dz_L^L}{dq} + \phi\lambda q \left(\frac{dz_L^L}{dq} - \frac{dz_L^{L-1}}{dq} \right) \tag{87}$$

This last result will be useful at a later stage of the proof. We now have to introduce the (sufficient) condition that $h \leq 0.25$, which secures monotone rankings in terms of employment across layers. To show this, we first recognize that Equation 52 as evaluated for $l = 0$ implies that $\frac{n_L^0}{n_L^1} = \frac{e^{\lambda z_L^0}}{h} \geq \frac{1}{h} \geq 4$. Second, our assumption of cumulative knowledge implies that z_L^l must increase across layers (ie. with respect to l for a given L). Equations 57 (where $\eta = 0$) and 58 thus imply that the ratio $\frac{n_L^l}{n_L^{l+1}}$ increases with l for $l \in \{0, ..., L-2\}$, and each such ratio is therefore greater than the first one, $\frac{n_L^0}{n_L^1} \geq \frac{1}{h} \geq 4$. The fact that $\frac{n_L^0}{n_L^1} \geq \frac{1}{h} \geq 4$ together with Equation 81 also imply the following rankings between $\frac{dz_L^0}{dq}$ and $\frac{dz_L^1}{dq}$, depending on the sign of $\frac{dz_L^0}{dq}$:

– if $\frac{dz_L^0}{dq} > 0$, then $\frac{dz_L^1}{dq} \geq \frac{1}{h}\frac{dz_L^0}{dq} \geq 4\frac{dz_L^0}{dq}$ (> 0)

– if $\frac{dz_L^0}{dq} \leq 0$, then $\frac{dz_L^1}{dq} \leq \frac{1}{h}\frac{dz_L^0}{dq} \leq 4\frac{dz_L^0}{dq}$ (≤ 0).

The challenge of the proof at this stage is to determine which of these two regimes is valid. To that end, we rely on recurrence reasoning in each regime:[85]

– *Recurrence equation, positive case* $\left(\frac{dz_L^0}{dq} > 0\right)$*:* Suppose that $\frac{dz_L^l}{dq} > 2\frac{dz_L^{l-1}}{dq} > 0$ for some $l \in \{1, ..., L-2\}$. Notice that in particular, this holds true for $l = 1$ whenever $\frac{dz_L^0}{dq} > 0$. Equation 82 implies that:

$$\frac{dz_L^{l+1}}{dq} = \frac{n_L^l}{n_L^{l+1}}\left(\frac{dz_L^l}{dq} - \frac{dz_L^{l-1}}{dq}\right) > \frac{n_L^l}{n_L^{l+1}}\left(\frac{dz_L^l}{dq} - \frac{1}{2}\frac{dz_L^l}{dq}\right) = \frac{n_L^l}{2n_L^{l+1}}\frac{dz_L^l}{dq} > 2\frac{dz_L^l}{dq} > 0$$

where the last inequality relies on the result $\frac{n_L^l}{n_L^{l+1}} \geq \frac{1}{h} \geq 4$ which was previously proved for $l \in \{0, ..., L-2\}$. By recurrence, this implies that whenever $\frac{dz_L^0}{dq} > 0$, all ratios $\frac{dz_L^l}{dq}$ for $l \in \{0, ..., L-1\}$ take the same sign (are strictly positive) and increase in absolute value.

– *Recurrence equation, negative case* $\left(\frac{dz_L^0}{dq} \leq 0\right)$*:* a similar reasoning holds with large inequalities and implies that all ratios $\frac{dz_L^l}{dq}$ for $l \in \{0, ..., L-1\}$ take the same sign (are negative or zero) and increase in absolute value.

However, if $\frac{dz_L^{L-1}}{dq} \leq \frac{dz_L^{L-2}}{dq} \leq 0$, then Equation 87 implies that $\frac{dz_L^L}{dq}$ must be negative or zero, since in that case:

$$\underbrace{(kc + \phi\lambda q)}_{>0}\frac{dz_L^L}{dq} = kcn_L^{L-1}\underbrace{\left(\frac{dz_L^{L-1}}{dq} - \frac{dz_L^{L-2}}{dq}\right)}_{\leq 0} + \phi\lambda q\underbrace{\frac{dz_L^{L-1}}{dq}}_{\leq 0}$$

This contradicts Equation 85, since the latter would imply:

$$A\lambda n_L^0 e^{\lambda z_L^L}\underbrace{\frac{dz_L^L}{dq}}_{\leq 0} + \lambda q\underbrace{\frac{dz_L^{L-1}}{dq}}_{\leq 0} = 1$$

which is impossible.

We can conclude from the previous developments that for $l \in \{0, ..., L-1\}$, all ratios $\frac{dz_L^l}{dq}$ must be strictly positive and increase with l.

[85] Solving these recurrence equations actually motivates the value of the upper bound for h (0.25) in the necessary condition that is required in this proof. Focusing on the first regime $\left(\frac{dz_L^0}{dq} > 0\right)$, let us define $\alpha > 0$ as a candidate value such that $\frac{dz_L^l}{dq} > \alpha\frac{dz_L^{l-1}}{dq} \implies \frac{dz_L^{l+1}}{dq} > \alpha\frac{dz_L^l}{dq}$. Our proof is based on the fact that Equation 82 implies that:

$$\frac{dz_L^{l+1}}{dq} = \frac{n_L^l}{n_L^{l+1}}\left(\frac{dz_L^l}{dq} - \frac{dz_L^{l-1}}{dq}\right) \geq \frac{1}{h}\left(\frac{dz_L^l}{dq} - \frac{dz_L^{l-1}}{dq}\right) \geq \frac{1}{h}\left(1 - \frac{1}{\alpha}\right)\frac{dz_L^l}{dq}$$

Thus, the following relation between parameters has to hold: $\frac{1}{h}\left(1 - \frac{1}{\alpha}\right) \geq \alpha \iff \alpha^2 h - \alpha + 1 \leq 0$. There are admissible real values for α if and only if this polynomial has real solutions. This happens if and only if $h \leq 0.25$. For $h = 0.25$, the only admissible candidate value is $\alpha = 2$. For $h \in \left[0; \frac{1}{4}\right]$, there is a entire range of admissible values for α: $\left[\frac{1-\sqrt{1-4h}}{2h}; \frac{1+\sqrt{1-4h}}{2h}\right]$. For $h \in \left[0; \frac{1}{4}\right]$, this interval always contains the value $\alpha = 2$ which is retained in the main proof.

Given Equation 84, the sign of $\frac{d\phi}{dq}$ is the same as the sign of $\left(\frac{dz_L^L}{dq} - \frac{dz_L^{L-1}}{dq}\right)$, but it can be either positive or negative in Equations 83 to 85.[86]

Lastly, differentiating Equations 55 and 56 with respect to q, we obtain that $\frac{dn_L^0}{dq} = \lambda n_L^0 \frac{dz_L^{L-1}}{dq} \geq \frac{dn_L^l}{dq} = \lambda n_L^l \left(\frac{dz_L^{L-1}}{dq} - \frac{dz_L^{l-1}}{dq}\right)$ for $l \in \{1, ..., L-1\}$, such that:

$$\frac{dn_L^0}{dq} > \frac{dn_L^1}{dq} > ... > \frac{dn_L^{L-1}}{dq} > 0 = \frac{dn_L^L}{dq}$$

□

Proof for the case $L \geq 2$ and $\eta > 0$, which implies $z_L^0 = 0$. We first notice that $\eta \geq 0$ in Equation 57 implies:

$$\begin{aligned} \eta \geq 0 &\iff n_L^0 k \left[c - \lambda h(cz_L^1 + 1)\right] \geq 0 \\ &\implies c \geq \lambda h \end{aligned}$$

This contradicts the parametric assumption of the kind that is introduced in Caliendo and Rossi-Hansberg (2012), which would precisely rule out the regime where $z_L^0 = 0$.

The proof in this regime is slightly altered and introduces the additional sufficient condition that $h \leq \frac{c}{\lambda} \leq 0.25$. It proceed as follows. Equations 77 to 80 are replaced with the following set of constraints (building on the fact that at the optimum, $n_L^0 = \frac{e^{\lambda z_L^{L-1}}}{h}$ and $n_L^1 = e^{\lambda(z_L^{L-1} - 0)} = h n_L^0$):[87]

$$\begin{aligned} \forall l \in \{1, ..., L-2\}, \quad \lambda(cz_L^{l+1} + 1) &= ce^{\lambda(z_L^l - z_L^{l-1})} && (88) \\ \sum_{l=2}^{L-1} n_L^l kc + n_L^0 k\left[hc + \lambda h(cz_L^1 + 1) + \lambda\right] &= \phi\lambda q && (89) \\ \phi A \lambda e^{-\lambda z_L^L} n_L^0 &= kc && (90) \\ A(1 - e^{-\lambda z_L^L}) n_L^0 &= q && (91) \end{aligned}$$

Differentiating each of the previous equations with respect to q we obtain:

$$\begin{aligned} n_L^1 \frac{dz_L^1}{dq} &= n_L^2 \frac{dz_L^2}{dq} && (92) \\ \forall l \in \{1, ..., L-2\}, \quad n_L^l \left(\frac{dz_L^l}{dq} - \frac{dz_L^{l-1}}{dq}\right) &= n_L^{l+1} \frac{dz_L^{l+1}}{dq} && (93) \\ \phi\lambda q \frac{dz_L^{L-1}}{dq} + kcn_L^{L-1}\left(\frac{dz_L^{L-1}}{dq} - \frac{dz_L^{L-2}}{dq}\right) &= \phi + q\frac{d\phi}{dq} && (94) \\ \phi\lambda\left(\frac{dz_L^L}{dq} - \frac{dz_L^{L-1}}{dq}\right) &= \frac{d\phi}{dq} && (95) \\ A\lambda n_L^0 \frac{dz_L^{L-1}}{dq} + A\lambda n_L^0 e^{-\lambda z_L^L}\left(\frac{dz_L^L}{dq} - \frac{dz_L^{L-1}}{dq}\right) &= 1 && (96) \end{aligned}$$

The first relation corresponds to Equation 93 as evaluated at $l = 1$ and recognizing that $\frac{dz_L^0}{dq} = 0$. Interestingly, Equation 94 coincides with Equation 83 despite the fact that Equation 89 does not coincide with

[86] Yet, it is possible to derive regimes where $\frac{d\phi}{dq} \geq 0$. Combining Equation 87 together with the result from the recurrence reasoning that $\frac{dz_L^{L-1}}{dq} > 2\frac{dz_L^{L-2}}{dq}$ implies that $\frac{d\phi}{dq} \geq 0$ whenever $n_L^{L-1} \geq 2$.

[87] Again, it should be noticed that Equation 88 does not operate when $L = 2$. However, in this regime, Equations 89, 90 and 91 are sufficient to derive the same results as in the case where $L > 2$ and Equation 88 holds.

Equation 78. This result is derived in the same way as previously:

$$\begin{aligned} kc\sum_{l=2}^{L-1}\lambda n_L^l\left(\frac{dz_L^{L-1}}{dq}-\frac{dz_L^{l-1}}{dq}\right)+k\lambda n_L^0\frac{dz_L^{L-1}}{dq}\left[hc+\lambda h(cz_L^1+1)+\lambda\right]+\lambda n_L^0 kch\frac{dz_L^1}{dq} &= \phi\lambda+\lambda q\frac{d\phi}{dq} \\ \therefore \phi\lambda q\frac{dz_L^{L-1}}{dq}-kc\sum_{l=2}^{L-1}n_L^l\frac{dz_L^{l-1}}{dq}+n_L^0 kch\frac{dz_L^1}{dq} &= \phi+q\frac{d\phi}{dq} \\ \therefore \phi\lambda q\frac{dz_L^{L-1}}{dq}-kc\sum_{l=2}^{L-2}\left(n_L^l\frac{dz_L^l}{dq}-n_L^{l+1}\frac{dz_L^{l+1}}{dq}\right)-kcn_L^{L-1}\frac{dz_L^{L-2}}{dq}+n_L^0 kch\frac{dz_L^1}{dq} &= \phi+q\frac{d\phi}{dq} \\ \therefore \phi\lambda q\frac{dz_L^{L-1}}{dq}-kc\left(n_L^2\frac{dz_L^2}{dq}-n_L^{L-1}\frac{dz_L^{L-1}}{dq}\right)-kcn_L^{L-1}\frac{dz_L^{L-2}}{dq}+kcn_L^1\frac{dz_L^1}{dq} &= \phi+q\frac{d\phi}{dq} \\ \therefore \phi\lambda q\frac{dz_L^{L-1}}{dq}+kcn_L^{L-1}\left(\frac{dz_L^{L-1}}{dq}-\frac{dz_L^{L-2}}{dq}\right) &= \phi+q\frac{d\phi}{dq} \end{aligned}$$

In the previous set of equations, the second line uses the fact that $\sum_{l=2}^{L-1} n_L^l kc + n_L^0 k\left[hc+\lambda h(cz_L^1+1)+\lambda\right] = \phi\lambda q$ from Equation 89. The third line uses the fact that $n_L^l\frac{dz_L^{l-1}}{dq} = n_L^l\frac{dz_L^l}{dq} - n_L^{l+1}\frac{dz_L^{l+1}}{dq}$ for $l\in\{1,...,L-2\}$ from Equation 93, and the fourth line incorporates Equation 92.

Since Equations 94 to 96 coincide with Equations 83 to 85, Equation 87 still holds. Furthermore, Equation 58 implies:

$$\forall l\in\{1,...,L-2\},\qquad \frac{n_L^l}{n_L^{l+1}}=\frac{\lambda}{c}(z_L^{l+1}+1) \tag{97}$$

In our framework of cumulative knowledge, as long as $h\leq\frac{c}{\lambda}\leq 0.25 \iff \frac{1}{h}\geq\frac{\lambda}{c}\geq 4$, this implies that the ratio $\frac{n_L^l}{n_L^{l+1}}$ increases with l for $l\in\{1,...,L-2\}$ and that each such ratio is greater than $\frac{\lambda}{c}\geq 4$.

The fact that $\frac{n_L^1}{n_L^2}\geq\frac{\lambda}{c}\geq 4$ together with Equation 92 also imply the following rankings between $\frac{dz_L^1}{dq}$ and $\frac{dz_L^2}{dq}$, depending on the sign of $\frac{dz_L^1}{dq}$:

- if $\frac{dz_L^1}{dq}>0$, then $\frac{dz_L^2}{dq}\geq\frac{\lambda}{c}\frac{dz_L^1}{dq}\geq 4\frac{dz_L^1}{dq}$ (>0)
- if $\frac{dz_L^1}{dq}\leq 0$, then $\frac{dz_L^2}{dq}\leq\frac{\lambda}{c}\frac{dz_L^1}{dq}\leq 4\frac{dz_L^1}{dq}$ (≤ 0).

As previously, the challenge of the proof at this stage is to determine which of these two regimes is valid. To that end, we rely on recurrence reasoning in each regime:

- *Recurrence equation, positive case* $\left(\frac{dz_L^1}{dq}>0\right)$*:* Suppose that $\frac{dz_L^l}{dq}>2\frac{dz_L^{l-1}}{dq}>0$ for some $l\in\{1,...,L-2\}$. Notice that in particular, this holds true for $l=2$ whenever $\frac{dz_L^1}{dq}>0$. Equation 93 implies that:

$$\frac{dz_L^{l+1}}{dq}=\frac{n_L^l}{n_L^{l+1}}\left(\frac{dz_L^l}{dq}-\frac{dz_L^{l-1}}{dq}\right)>\frac{n_L^l}{n_L^{l+1}}\left(\frac{dz_L^l}{dq}-\frac{1}{2}\frac{dz_L^l}{dq}\right)=\frac{n_L^l}{2n_L^{l+1}}\frac{dz_L^l}{dq}>2\frac{dz_L^l}{dq}>0$$

  where the last inequality relies on the result $\frac{n_L^l}{n_L^{l+1}}\geq\frac{\lambda}{c}\geq 4$ which was previously proved for $l\in\{1,...,L-2\}$. By recurrence, this implies that whenever $\frac{dz_L^0}{dq}>0$, all ratios $\frac{dz_L^l}{dq}$ for $l\in\{1,...,L-1\}$ take the same sign (are strictly positive) and increase in absolute value.
- *Recurrence equation, negative case* $\left(\frac{dz_L^1}{dq}\leq 0\right)$*:* a similar reasoning holds with large inequalities and implies that all ratios $\frac{dz_L^l}{dq}$ for $l\in\{1,...,L-1\}$ take the same sign (are negative or zero) and increase in absolute value.

However, if $\frac{dz_L^{L-1}}{dq} \leq \frac{dz_L^{L-2}}{dq} \leq 0$, then Equation 87 implies that $\frac{dz_L^L}{dq}$ must be negative or zero, since in that case:

$$\underbrace{(kc + \phi\lambda q)}_{>0} \frac{dz_L^L}{dq} = kcn_L^{L-1} \underbrace{\left(\frac{dz_L^{L-1}}{dq} - \frac{dz_L^{L-2}}{dq} \right)}_{\leq 0} + \phi\lambda q \underbrace{\frac{dz_L^{L-1}}{dq}}_{\leq 0}$$

This contradicts Equation 96, since the latter would imply:

$$A\lambda n_L^0 e^{\lambda z_L^L} \underbrace{\frac{dz_L^L}{dq}}_{\leq 0} + \lambda q \underbrace{\frac{dz_L^{L-1}}{dq}}_{\leq 0} = 1$$

which is impossible.

We can conclude from the previous developments that for $l \in \{1, ..., L-1\}$, all ratios $\frac{dz_L^l}{dq}$ must be strictly positive and increase with l. Furthermore, $\frac{dz_L^0}{dq}$ is trivially 0. The treatment of $\frac{dz_L^L}{dq}$ is slightly more indirect, but Equation 87 implies that:

$$\underbrace{(kc + \phi\lambda q)}_{>0} \frac{dz_L^L}{dq} = kcn_L^{L-1} \underbrace{\left(\frac{dz_L^{L-1}}{dq} - \frac{dz_L^{L-2}}{dq} \right)}_{>0} + \phi\lambda q \underbrace{\frac{dz_L^{L-1}}{dq}}_{>0}$$

such that $\frac{dz_L^L}{dq}$ is necessarily strictly positive.

As in the previous case, given Equation 95, the sign of $\frac{d\phi}{dq}$ is the same as the sign of $\left(\frac{dz_L^L}{dq} - \frac{dz_L^{L-1}}{dq} \right)$, but it can be either positive or negative in Equations 94 to 96.

Lastly, differentiating Equations 55 and 56 with respect to q, we obtain that $\frac{dn_L^0}{dq} = \lambda n_L^0 \frac{dz_L^{L-1}}{dq} \geq \frac{dn_L^1}{dq} = \lambda n_L^1 \frac{dz_L^{L-1}}{dq} \geq \frac{dn_L^l}{dq} = \lambda n_L^l \left(\frac{dz_L^{L-1}}{dq} - \frac{dz_L^{l-1}}{dq} \right)$ for $l \in \{2, ..., L-1\}$, such that:

$$\frac{dn_L^0}{dq} > \frac{dn_L^1}{dq} > ... > \frac{dn_L^{L-1}}{dq} > 0 = \frac{dn_L^L}{dq}$$

□

E.3 Introducing the Minimum Wage

E.3.1 Supply: Cost Minimization with a Minimum Wage Constraint

Our framework mimics the French institutions and assumes that the minimum wage is only legally binding in actual firms, ie. organizations featuring $L \geq 1$. In our framework, this constraint hits the bottom (production) layer as follows:

$$k(1 + cz_L^0) \geq \overline{w},$$

which is more conveniently written in terms of z_L^0 (see Section 4.2) as:

$$z_L^0 \geq \frac{\overline{w} - k}{ck} \equiv \overline{z}_L^0.$$

As shown below, this modified constraint only marginally alters the previous proofs, such that all previous results are preserved.

Firms with 1 layer. The full Lagrangian associated with the new set of constraints is altered in the following way:

$$\mathcal{L}_1(n_1^0, z_1^0) = k\left[cz_1^1 + 1 + n_1^0(cz_1^0+1)\right] - \phi\left[n_1^0 A(1-e^{-\lambda z_1^1}) - q\right] - \theta\left[\frac{e^{\lambda z_1^0}}{h} - n_1^0\right] - \eta\left[z_1^0 - \overline{z}_1^0\right] \tag{98}$$

The only Kuhn-Tucker condition that is altered is equation 36. It is replaced with:

$$0 = \eta\left[z_1^0 - \overline{z}_1^0\right] \tag{99}$$

Our previous developments only have to be slightly adjusted. In cases where the minimum wage is not binding, ie whenever $\eta = 0$ in Condition 99 (such that $z_1^0 \geq \overline{z}_1^0$), then θ has to be strictly positive in Condition 32. Combined with Condition 35, this implies that $n_1^0 = \frac{e^{\lambda z_1^0}}{h}$.

Alternatively, if the minimum wage binds such that the Lagrange multiplier $\eta \geq 0$ (and $z_1^0 = \overline{z}_1^0$), then θ is no longer constrained to be strictly positive in Conditions 31, 32 and 35. More precisely, if $\theta = 0$, then n_1^0 can become strictly smaller than $\frac{e^{\lambda \overline{z}_1^0}}{h}$ (Condition 35). Notice in particular that since the bound is higher than previously ($\overline{z}_1^0 \geq 0$ whenever the institutional minimum wage is not degenerate), all else equal, the organizational distortion is likely to occur more often. However, here, and in contrast with the baseline case without minimum wage, production workers are always "productive" (in the sense that they are always able to solve problems) since $z_1^0 = \overline{z}_1^0 > 0$.

It is possible to investigate more generally how the optimal choices for n_1^0, z_1^1 and ϕ evolve with $\overline{z}_1^0$ as solutions to firms' cost minimization programs:[88]

- In the regime where $\eta \geq 0$ and $\theta > 0$, we get that:

$$\frac{dz_1^0}{d\overline{z}_1^0} = 1, \quad \frac{dn_1^0}{d\overline{z}_1^0} = \lambda n_1^0 > 0, \quad \frac{dz_1^1}{d\overline{z}_1^0} = -\frac{hq}{Ae^{\lambda \overline{z}_1^0} - hq} < 0, \quad \frac{d\phi}{d\overline{z}_1^0} = -\frac{Akche^{\lambda \overline{z}_1^0}}{\left(Ae^{\lambda \overline{z}_1^0} - hq\right)^2} < 0$$

- In the regime where $\eta \geq 0$ and $\theta = 0$, we get that:

$$\frac{dz_1^1}{d\overline{z}_1^0} = -\frac{q}{Ae^{-\lambda \overline{z}_1^0}}\left[1 + \frac{\lambda q[\lambda q(c\overline{z}_1^0+1) + 2Ac]}{\sqrt{\lambda^2 q^2 (c\overline{z}_1^0+1)^2 + 4\overline{z}_1^0 c\lambda q(c\overline{z}_1^0+1)}}\right] < 0,$$

$$\frac{dz_1^0}{d\overline{z}_1^0} = 1, \quad \frac{dn_1^0}{d\overline{z}_1^0} = -\frac{\lambda q e^{-\lambda \overline{z}_1^0}}{A\left(1-e^{-\lambda \overline{z}_1^0}\right)^2}\cdot\frac{dz_1^1}{d\overline{z}_1^0} > 0, \quad \frac{d\phi}{d\overline{z}_1^0} = -\frac{\lambda k(c\overline{z}_1^0+1)e^{-\lambda \overline{z}_1^0}}{A\left(1-e^{-\lambda \overline{z}_1^0}\right)^2}\cdot\frac{dz_1^1}{d\overline{z}_1^0} > 0$$

Firms with 2 or more layers. The Lagrangian is altered as:

$$\begin{aligned}\mathcal{L}_L\left((n_L^0)_l, (z_l^L)_l\right) &= k\left[\sum_{l=0}^{L} n_L^l (cz_L^l + 1)\right] - \phi\left[n_L^1 A(1-e^{-\lambda z_L^L}) - q\right] \\ &\quad - \sum_{l=0}^{L-1}\theta_l\left[\frac{e^{\lambda z_L^l}}{h} n_L^{l+1} - n_L^0\right] - \eta\left[z_L^0 - \overline{z}_L^0\right]\end{aligned} \tag{100}$$

This means that condition 53 is simply altered as:

$$0 = \eta\left[z_L^0 - \overline{z}_L^0\right]$$

[88] Recall that $\frac{dn_1^1}{d\overline{z}_1^0} = 0$ since $n_1^1 = 1$.

The transposition of our previous developments and proofs is almost straightforward, and all previous results remain valid. To be precise, the regime where $\eta = 0$ is unaltered, while the analysis for the regime where $\eta > 0$ is altered in the following way:

- We still have $n_L^0 = \frac{e^{\lambda z_L^{L-1}}}{h}$ but $n_L^1 = e^{\lambda(z_L^{L-1} - \bar{z}_L^0)} = h n_L^0 e^{-\lambda \bar{z}_L^0}$;
- Equation 89 is altered as:

$$\sum_{l=2}^{L-1} n_L^l kc + n_L^0 k \left[hce^{-\lambda \bar{z}_L^0} + \lambda h(cz_L^1 + 1) + \lambda(c\bar{z}_L^0 + 1) \right] = \phi \lambda q$$

- Equations 92 to 96 still hold such that the remainder of the proof also holds.

The comparative statics with respect to $\bar{z}_L^0$ are more complex than in the case where $L = 1$:

- For $L = 2$, differentiating Equations 89 to 91 leads to the following results:

$$\begin{aligned} \frac{dz_2^2}{d\bar{z}_2^0} - \frac{dz_2^1}{d\bar{z}_2^0} &= -e^{\lambda z_2^2} \frac{dz_2^1}{d\bar{z}_2^0} \\ \frac{d\phi}{d\bar{z}_2^0} &= -\phi \lambda e^{\lambda z_2^2} \frac{dz_2^1}{d\bar{z}_2^0} \\ kcn_2^0(1 - he^{-\lambda \bar{z}_2^0}) &= -\left[\phi \lambda q(1 + e^{\lambda Z_2^2}) + kchn_2^0 \right] \frac{dz_2^1}{d\bar{z}_2^0} \end{aligned}$$

  The last equation implies $\frac{dz_2^1}{d\bar{z}_2^0} < 0$ such that $\frac{dz_2^2}{d\bar{z}_2^0} > 0$ and $\frac{d\phi}{d\bar{z}_2^0} > 0$

- For $L > 2$, no general results hold: see Figure 5.

E.3.2 Partial Equilibrium Effects of Minimum Wage: Proofs

This Section of the appendix contains the proofs of the two Propositions from Section 4. In contrast to the previous analyses in Section E.3.1, the optimal level of production q is endogenized (while k remains exogenously determined in partial equilibrium).

Proof of Proposition 1 : Suppose that a firm is indifferent between producing with M and L layers, and that the minimum wage is binding on z_M^0 but not on z_L^0. If so, then an increase in the minimum wage reduces profits for a firm with M layers but not for a firm with L layers:

$$\frac{d\pi_x}{d\bar{w}} = \frac{d\pi_x}{dz_x^0} = \frac{\partial \pi_x}{\partial z_x^0} + \frac{\partial \pi_x}{\partial n_x^0}\frac{dn_x^0}{dz_x^0} + \sum_{l=1}^{x} \frac{\partial \pi_x}{\partial z_x^l}\frac{dz_x^l}{dz_x^0} + \frac{\partial \pi_x}{\partial q_x}\frac{dq_x}{dz_x^0}, \ \forall x \in \{L, M\}$$

However, an envelope condition tells us that $\frac{\partial \pi_x}{\partial n_x^0} = \frac{\partial \pi_x}{\partial z_x^l} = \frac{\partial \pi_x}{\partial q_x} = 0$. Therefore, we have:

$$\frac{d\pi_x}{d\bar{w}} = \frac{\partial \pi_x}{\partial z_x^0} = -\frac{\partial C_x}{\partial z_x^0} = -(n_x^0 kc - \lambda n_x^1 k(cz_x^1 + 1)) = -\eta_x$$

which is zero for L but negative for M. As a result, we know that $\frac{d\pi_M}{d\bar{w}} < 0$ and $\frac{d\pi_L}{d\bar{w}} = 0$, and so a firm that is on the margin between the two will shift to production with L layers.

If, on the other hand the minimum wage is binding on both L and M, the result is ambiguous, because $\eta_M > 0$ and $\eta_L > 0$. Our intuition is that the multiplier η should be larger for an organization with more layers, where the minimum wage should be more binding, but at this stage, we are unable to prove that intuition.

Proof of Proposition 2: There are 3 different regimes to consider: they correspond to those where the minimum wage constraint is binding.

Proof for the case where $L = 1$, $\eta > 0$ and $\theta = 0$ such that $z_1^0 = \bar{z}_1^0$ and $n_1^0 \leq \frac{e^{\lambda \bar{z}_L^0}}{h}$. In this case, we have 3 first-order conditions which can be combined with the profit-maximization Condition 17, $q = \alpha k \left(\frac{\sigma}{\sigma-1}\phi\right)^{-\sigma}$, to deliver:

$$k(c\bar{z}_1^0 + 1) = \phi A(1 - e^{-\lambda z_1^1}) \tag{101}$$

$$kc = \phi A \lambda e^{-\lambda z_1^1} n_1^0 \tag{102}$$

$$A(1 - e^{-\lambda z_1^1}) n_1^0 = \alpha k \left(\frac{\sigma}{\sigma - 1}\phi\right)^{-\sigma} \tag{103}$$

Differentiating each condition with respect to $\bar{z}_1^0$, ie. the parameter corresponding to the lower bound in the minimum wage constraint, we obtain, first, $\frac{dz_1^0}{d\bar{z}_1^0} = 1$, and the following set of equations:

$$kc = A(1 - e^{-\lambda z_1^1})\frac{d\phi}{d\bar{z}_1^0} + \phi A \lambda e^{-\lambda z_1^1}\frac{dz_1^1}{d\bar{z}_1^0} \tag{104}$$

$$\phi \lambda n_1^0 \frac{dz_1^1}{d\bar{z}_1^0} = n_1^0 \frac{d\phi}{d\bar{z}_1^0} + \phi \frac{dn_1^0}{d\bar{z}_1^0} \tag{105}$$

$$A\lambda e^{-\lambda z_1^1} n_1^0 \frac{dz_1^1}{d\bar{z}_1^0} + A(1 - e^{-\lambda z_1^1})\frac{dn_1^0}{d\bar{z}_1^0} = -\sigma \alpha k \left(\frac{\sigma}{\sigma - 1}\right)^{-\sigma} \phi^{-\sigma-1}\frac{d\phi}{d\bar{z}_1^0} \tag{106}$$

Equation 104 can be solved for:

$$\frac{d\phi}{d\bar{z}_1^0} = \frac{kc}{A(1 - e^{-\lambda z_1^1})} - \frac{\phi \lambda e^{-\lambda z_1^1}}{1 - e^{-\lambda z_1^1}}\frac{dz_1^1}{d\bar{z}_1^0}.$$

Equation 105 then rearranges as:

$$A(1 - e^{-\lambda z_1^1})\frac{dn_1^0}{d\bar{z}_1^0} = A\lambda n_1^0 \frac{dz_1^1}{d\bar{z}_1^0} - \frac{kcn_1^0}{\phi}$$

and Equation 106 rearranges to deliver:

$$\phi A \lambda \left(1 + (1 - \sigma)e^{-\lambda z_1^1}\right)\frac{dz_1^1}{d\bar{z}_1^0} = (1 - \sigma)kc \iff \frac{\phi A \lambda}{kc}\frac{dz_1^1}{d\bar{z}_1^0} = \frac{\sigma - 1}{(\sigma - 1)e^{-\lambda z_1^1} - 1}$$

Since σ is typically calibrated at values that exceed 1, we obtain that $\frac{dz_1^1}{d\bar{z}_1^0}$ is of the same sign as $(\sigma - 1)e^{-\lambda z_1^1} - 1$. This implies:

$$\frac{dz_1^1}{d\bar{z}_1^0} < 0 \iff z_1^1 > \frac{\ln(\sigma - 1)}{\lambda},$$

which corresponds to our calibration to the French data.

Next, substituting the rearranged form of Equation 106 into Equation 105, we have:

$$A(1 - e^{-\lambda z_1^1})\frac{dn_1^0}{d\bar{z}_1^0} = \frac{kcn_1^0}{\phi}\left[\frac{\sigma + (1 - \sigma)e^{-\lambda z_1^1}}{(\sigma - 1)e^{-\lambda z_1^1} - 1}\right].$$

The numerator in the final square brackets is always positive, such that $\frac{dn_1^0}{d\overline{z}_1^0}$ is of the same sign as $(\sigma - 1)e^{-\lambda z_1^1} - 1$. This implies:

$$\frac{dn_1^0}{d\overline{z}_1^0} < 0 \Longleftrightarrow z_1^1 > \frac{\ln(\sigma-1)}{\lambda},$$

which, again, corresponds to our calibration to the French data.

Finally, Equation 104 rearranges to deliver $A(1-e^{-\lambda z_1^1})\frac{d\phi}{d\overline{z}_1^0} = -\frac{kc}{(\sigma-1)e^{-\lambda z_1^1}-1}$ such that:

$$\frac{d\phi}{d\overline{z}_1^0} > 0 \Longleftrightarrow z_1^1 > \frac{\ln(\sigma-1)}{\lambda}.$$

□

Proof for the case where $L = 1$, $\eta > 0$ and $\theta > 0$ such that $z_1^0 = \overline{z}_1^0$ and $n_1^0 = \frac{e^{\lambda \overline{z}_L^0}}{h}$. There are two relevant first-order conditions (for z_1^1 and ϕ, since z_1^0 is fixed by the minimum wage), which can be combined with profit maximization to deliver:

$$kch = \phi A \lambda e^{\lambda(z_1^0 - z_1^1)} \tag{107}$$

$$A(1-e^{-\lambda z_1^1})e^{\lambda z_1^0} = \alpha hk \left(\frac{\sigma}{\sigma-1}\phi\right)^{-\sigma} \tag{108}$$

Differentiating each equation with respect to z_1^0, we find:

$$\frac{d\phi}{d\overline{z}_1^0} = \phi\lambda\left(\frac{dz_1^1}{d\overline{z}_1^0} - 1\right) \tag{109}$$

$$\phi\lambda\left[e^{-\lambda z_1^1}\frac{dz_1^1}{d\overline{z}_1^0} + (1-e^{-\lambda z_1^1})\right] = -\sigma(1-e^{-\lambda z_1^1})\frac{d\phi}{d\overline{z}_1^0} \tag{110}$$

Combining and rearranging, we get:

$$\frac{dz_1^1}{d\overline{z}_1^0} = \frac{(\sigma-1)(1-e^{-\lambda z_1^1})}{e^{-\lambda z_1^1} + \sigma(1-e^{-\lambda z_1^1})} > 0$$

and substituting back into Equation 109:

$$\frac{d\phi}{d\overline{z}_1^0} = \phi\lambda\left(\frac{(\sigma-1)(1-e^{-\lambda z_1^1}) - e^{-\lambda z_1^1} - \sigma(1-e^{-\lambda z_1^1})}{e^{-\lambda z_1^1} + \sigma(1-e^{-\lambda z_1^1})}\right) = \phi\lambda\left(\frac{-1}{e^{-\lambda z_1^1} + \sigma(1-e^{-\lambda z_1^1})}\right) < 0.$$

□

Proof for the case where $L \geq 2$ and $\eta \geq 0$, such that $z_L^0 = \overline{z}_L^0$ (and $n_L^0 = \frac{e^{\lambda z_L^{L-1}}}{h}$). General results can only be derived in the case where the minimum wage is only marginally binding, ie. whenever $\eta = 0$ and $kcn_L^0 = \lambda n_L^1 k(cz_L^1 + 1)$.

There are four relevant first-order conditions, to be combined with profit maximization:

$$\begin{aligned} ce^{\lambda(z_L^l - z_L^{l-1})} &= \lambda(cz_L^{l+1}+1) && (111)\\ \sum_{l=0}^{L-1} n_L^l kc + \lambda n_L^0 k(cz_L^0+1) &= \lambda\alpha k \left(\frac{\sigma}{\sigma-1}\right)^{-\sigma}\phi^{1-\sigma} && (112)\\ kc &= \phi A\lambda e^{-\lambda z_L^L} n_L^0 && (113)\\ A(1-e^{-\lambda z_L^L})n_L^0 &= \alpha k \left(\frac{\sigma}{\sigma-1}\right)^{-\sigma}\phi^{-\sigma} && (114) \end{aligned}$$

Differentiating each equation with respect to $\overline{z}_L^0$, we obtain:

$$\begin{aligned} \forall l \in \{1,...,L-2\}, \qquad \frac{dz_L^{l+1}}{d\overline{z}_L^0} &= \frac{n_L^l}{n_L^{l+1}}\left(\frac{dz_L^l}{d\overline{z}_L^0} - \frac{dz_L^{l-1}}{d\overline{z}_L^0}\right) && (115)\\ \phi\lambda q \frac{dz_L^{L-1}}{d\overline{z}_L^0} + kcn_L^{L-1}\left(\frac{dz_L^{L-1}}{d\overline{z}_L^0} - \frac{dz_L^{L-2}}{d\overline{z}_L^0}\right) &= (1-\sigma)q\frac{d\phi}{d\overline{z}_L^0} && (116)\\ \frac{d\phi}{d\overline{z}_L^0} &= \phi\lambda\left(\frac{dz_L^L}{d\overline{z}_L^0} - \frac{dz_L^{L-1}}{d\overline{z}_L^0}\right) && (117)\\ A\lambda(1-e^{-\lambda z_L^L})n_L^0\frac{dz_L^{L-1}}{d\overline{z}_L^0} + A\lambda n_L^0 e^{-\lambda z_L^L}\frac{dz_L^L}{d\overline{z}_L^0} &= \frac{-\sigma q}{\phi}\frac{d\phi}{d\overline{z}_L^0} && (118) \end{aligned}$$

Combining Equation 117 and 118, we find:

$$A\lambda n_L^0\left[\frac{dz_L^{L-1}}{d\overline{z}_L^0} + \frac{e^{-\lambda z_L^L}}{\phi\lambda}\frac{d\phi}{d\overline{z}_L^0}\right] = \frac{-\sigma q}{\phi}\frac{d\phi}{d\overline{z}_L^0}$$

and therefore $A\lambda n_L^0 \frac{dz_L^{L-1}}{d\overline{z}_L^0} = -\frac{\sigma q + A n_L^0 e^{-\lambda z_L^L}}{\phi}\frac{d\phi}{d\overline{z}_L^0}$, so $\frac{dz_L^{L-1}}{d\overline{z}_L^0}$ and $\frac{d\phi}{d\overline{z}_L^0}$ take opposite signs. Plugging this into Equation 118 and rearranging, we obtain $\frac{dz_L^L}{d\overline{z}_L^0} = \frac{(1-\sigma)(1-e^{-\lambda z_L^L})}{\phi\lambda}\frac{d\phi}{d\overline{z}_L^0}$, so the sign of $\frac{dz_L^L}{d\overline{z}_L^0}$ is the same as the sign of $\frac{dz_L^{L-1}}{d\overline{z}_L^0}$ and opposite to $\frac{d\phi}{d\overline{z}_L^0}$.

Next, we can plug the solution for $\frac{dz_L^{L-1}}{d\overline{z}_L^0}$ into Equation 116:

$$kcn_L^{L-1}\left(\frac{dz_L^{L-2}}{d\overline{z}_L^0} - \frac{dz_L^{L-1}}{d\overline{z}_L^0}\right) = \left[-1 + (\sigma-1)e^{-\lambda z_L^L}\right] q \frac{d\phi}{d\overline{z}_L^0}.$$

Therefore, $\left(\frac{dz_L^{L-2}}{d\overline{z}_L^0} - \frac{dz_L^{L-1}}{d\overline{z}_L^0}\right)$ takes the same sign as $\frac{dz_L^{L-1}}{d\overline{z}_L^0}$ if and only if $1+(1-\sigma)e^{-\lambda z_L^L} > 0$. If this condition is satisfied, we have $\frac{dz_L^{L-2}}{d\overline{z}_L^0} = X_{L-2}\frac{dz_L^{L-1}}{d\overline{z}_L^0}$ where $X_{L-2} > 1$. We also know that $\frac{n_L^l}{n_1^{l+1}}$ is increasing in l for $l \in \{1,...,L-2\}$, and as before we have $\frac{n_L^1}{n_L^2} \geq 4$, so that $\frac{n_L^l}{n_1^{l+1}} \geq 4$ for all $l \leq L-2$. Then we can rely on a descending chain of inductive reasoning:

- if $\frac{dz_L^l}{d\overline{z}_L^0} = X_l \frac{dz_L^{l+1}}{d\overline{z}_L^0}$, where $X_l \geq \frac{1}{2}$, for $l \in \{2,...,L-2\}$, then $\frac{dz_L^{l-1}}{d\overline{z}_L^0} = \frac{dz_L^l}{d\overline{z}_L^0} - \frac{n_L^{l+1}}{n_L^l}\frac{dz_L^{l+1}}{d\overline{z}_L^0} = \left(1 - \frac{n_L^{l+1}}{n_L^l X_l}\right)\frac{dz_L^l}{d\overline{z}_L^0}$
- then $X_{l-1} \equiv 1 - \frac{n_L^{l+1}}{n_L^l X_l} \geq \frac{1}{2}$, so $\frac{dz_L^l}{d\overline{z}_L^0}$ takes same sign for all $l \in 1,...,L$.

Finally, we have $\frac{n_L^1}{n_L^2}\left(\frac{dz_L^1}{d\overline{z}_L^0} - \frac{dz_L^0}{d\overline{z}_L^0}\right) = \frac{dz_L^2}{d\overline{z}_L^0} = \frac{1}{X_1}\frac{dz_L^1}{d\overline{z}_L^0}$ where $\frac{dz_L^0}{d\overline{z}_L^0} > 0$, and therefore $\frac{dz_L^1}{d\overline{z}_L^0} = \frac{\frac{n_L^1}{n_L^2}}{\frac{n_L^1}{n_L^2} - \frac{1}{X_1}}\frac{dz_L^0}{d\overline{z}_L^0} > 0$. Therefore, $\frac{dz_L^l}{d\overline{z}_L^0} > 0$ for all $l \in 1,...,L$, and $\frac{d\phi}{d\overline{z}_L^0} < 0$. However, $\frac{dn_L^l}{d\overline{z}_L^0}$ is of ambiguous sign except for

$\frac{dn_L^0}{dz_L^0} = \lambda n_L^0 \frac{dz_L^{L-1}}{dz_L^0} > 0.$

It is important to note that, in the case of $L \geq 2$, we need to assume that the minimum wage is only marginally binding for the proof to work; in that case, $\eta = 0$, and we can use $kcn_L^0 = \lambda n_L^1 k(cz_L^1 + 1)$ to simplify Equation 112 and Equation 116 in the proof. Once the minimum wage is strictly above the value of z_L^0 that a firm would choose in the absence of the minimum wage, that condition no longer holds, and there will be a tendency for firms to decrease q by more than the above analysis suggests, leading to lower values of z_L^l and n_L^l. In our simulation of the model, we actually find that n_L^0 decreases for a sufficiently large minimum wage when $L = 3$.

□

F Sensitivity Analyses of Simulation Results

This Appendix contains a series of sensitivity analyses of the calibration and simulation results from Section 5. Our main goal is to assess whether our results can be safely interpreted as representative of the French economy. Specifically, we consider four alternative calibration strategies, based on increasingly comprehensive datasets, in order to check that our main results are not driven by some of the selections imposed by our identification strategy in our main estimation sample (see Appendix B): (i) one in which we include all firms in GMRs 1 through 5 with at least one (instead of five) paid employee; (ii) one in which we further include firms with at least 1 employee with a well defined but late RWT agreement (between 2002 and 2006) which were subject to SMIC rather than a GMR, as well as the firms from (i); (iii) one in which we use sample (i) but also incorporate a moment capturing the percentage of firms that consist of just one self-employed individual. There is no dataset available which measures this share for 2000-06. Rather, we measure it from Insee's FARE dataset for 2010-16 and assume in this robustness check that this value is also approximately relevant for 2006. Lastly, specification (iv) relies on sample (ii) but also inserts the same estimate of self-employment as the specification (iii).[89] None of the other aspects of the calibration strategy are altered, with the exception of the firm death rate (δ), which is calculated to be consistent with the different scenarios: $\delta = 0.327$ in scenarios (i) and (iii), and $\delta = 0.370$ in scenarios (ii) and (iv).

We present the moments used for each scenario in Panel (A) of Table F1, and the values actually attained in each simulated model in Panel (B) of Table F1. The results of the calibrations give us the parameter values found in Panel (C) of Table F1. They show that the fixed cost parameters f_E and f absorb most of the variations that are introduced, including the one corresponding to the insertion of an additional moment for self-employed entrepreneurs. In contrast, parameters c, h and A are only weakly affected, and the same holds true for the distribution of α (which only feature thinner tails), such that the partial equilibrium and general equilibrium results are globally preserved.

[89]Due to the fact that the sum of the shares across all possible organizational forms has to sum to 1, the first two specifications actually contain an implicit moment which pools all firms with 0 and 1 hierarchical layers, while in specifications (iii) and (iv), the implicit moment only concerns 1-layer firms.

Table F1: Calibration Sensitivity Analyses

(A) Target Values of Moments

Moment	(i)	(ii)	(iii)	(iv)
Share of agents who are teachers	0.098	0.098	0.098	0.098
Share of workers in new firms	0.209	0.241	0.209	0.241
Share of 0-layer firms			0.170	0.170
Share of 2-layer firms	0.352	0.267	0.292	0.222
Share of 3-layer firms	0.161	0.099	0.134	0.082
Average firm size	66.17	28.10	66.17	28.10
Average size of 2-layer firms	77.60	42.53	77.60	42.53
Average size of 3-layer firms	210.30	122.90	210.30	122.90
Average (hourly) wage (€)	19.70	20.04	19.70	20.04
Share of workers (in firms) bound by minimum wage	0.048	0.055	0.048	0.055

(B) Simulated Values of Moments

Moment	(i)	(ii)	(iii)	(iv)
Share of agents who are teachers	0.098	0.122	0.098	0.124
Share of workers in new firms	0.216	0.208	0.223	0.217
Share of 0-layer firms (additional moment)			0.172	0.166
Share of 2-layer firms	0.349	0.261	0.293	0.215
Share of 3-layer firms	0.161	0.097	0.134	0.082
Average firm size	66.43	29.98	66.00	30.23
Average size of 2-layer firms	77.42	47.12	77.75	46.59
Average size of 3-layer firms	209.81	132.30	212.32	142.87
Average (hourly) wage (€)	19.69	20.60	19.60	19.69
Share of workers (in firms) bound by minimum wage	0.048	0.057	0.048	0.058

(C) Calibrated Values of Parameters

Parameter	(i)	(ii)	(iii)	(iv)
Cost of knowledge acquisition, c	0.225	0.332	0.235	0.330
Time cost of problem communication, h	0.192	0.222	0.195	0.227
Output per problem solved, A ($\sim$ 100s of €/hour)	0.221	0.234	0.282	0.260
Fixed cost of entry, f_E	4.580	4.566	63.240	21.920
Fixed cost of production, f	4.774	2.471	0.116	0.217
Minimum wage $\bar{w}$, relative to average wage	0.909	0.864	0.913	0.863
2nd-node of α distribution ($\alpha = 50$)	-11.79	-11.72	-11.51	-11.47
3rd-node of α distribution ($\alpha = 500$)	-16.33	-13.88	-12.41	-12.72
4th-node of α distribution ($\alpha = 1000$)	-16.81	-32.25	-79.25	-66.57

Notes: Each column presents the values of moments or parameters computed on alternative datasets: Column (i) one includes all firms in GMRs 1 through 5 with at least 1 paid employee; Column (ii) further includes firms with at least 1 employee and late RWT agreements (subject to SMIC), as well as the firms from (i); Column (iii) uses sample (i) but adds a moment from Insee's FARE dataset for 2010-16 which measures the percentage of firms that consist of one self-employed individual; and Column (iv) uses sample (ii) but with the same self-employment moment added. Due to the fact that the sum of the shares across all possible organizational forms has to sum to 1, Columns (i) and (ii) contain an implicit moment which pools all firms with 0 and 1 hierarchical layers, while in Columns (iii) and (iv), the implicit moment only concerns 1-layer firms.

G Framework with Unemployment

In this Appendix, we examine whether a rigid labor market, ie. a feature often attributed to the French economy, is likely to affect our results. To this end, we extend the model by augmenting it with an endogenous unemployment rate, determined by workers' job-search effort. Consequently, the calibration exercise targets an additional empirical moment, which corresponds to the empirically measured unemployment rate.

To be precise, we assume that, in each period, non-entrepreneurs (including workers, managers, fixed-cost workers, and teachers) must search for a job. If they find a job, they receive net wage k (ie. the total wage $k(1+cz)$, minus the training costs kcz), while they receive an unemployment benefit b if they don't. The probability that they find a job is e, which they choose optimally subject to a job search cost $c(e) = \frac{\ln(1-e)+e}{-\chi}$. As a result, non-entrepreneurs select the optimal amount of effort: $e = \frac{\chi(k-b)}{1+\chi(k-b)}$. The free-entry and production conditions for entrepreneurs require that they compare their (expected or known) profits as entrepreneurs to a worker's expected utility $V = ek + (1-e)b - c(e)$ rather than just k. In the simulations which follow, we assume that unemployment benefits b are set according to a replacement rate that is applied over expected wages. This replacement rate is set to 0.57, a value that is in line with the rate applied in France. Lastly, the labor-market clearing condition also changes: if u represents the unemployment rate (within the population "at risk", ie. $1-M$), Equation 20 is replaced with:

$$1 - u = M \left(\frac{\left[\delta f_E + \int_{\overline{\alpha}}^{+\infty} \left(C(q(\alpha);1) + f\right) g(\alpha) d\alpha\right]}{1 - G(\overline{\alpha})} - u \right) \tag{119}$$

Table G1 reports the calibrated moments and parameters. Introducing the possibility of being unemployed for workers alters mainly the values that are obtained for the parameters driving occupational choices, namely f_E and f, but only marginally. The distribution of α is also estimated to have slightly thinner tails than in the main calibration.

Table G2 replicates the same general equilibrium simulation scenarios as in the main text. The simulation shows that our main results are all preserved and that the unemployment margin is not a significant adjustment force in the model when we simulate firms exposed to increasingly stringent minimum wage constraints. First, the unemployment rate is unaffected up to the third digit when the minimum wage constraint is introduced (columns (1) and (2) in Panel (A) of Table G2). It only adjusts significantly for large increases relative to the baseline scenario that is calibrated to the actual French data. In contrast, the share of "teachers" nearly doubles in the "+16%" scenario. While our model is very crude in terms of modeling the public sector, it does suggest that the public sector may be the primary adjustment margin as the demand for skills increases in the economy, in order to render more expensive production workers more productive. Ultimately, the global "cost" of introducing a minimum wage constraint in terms of total output is negligible in the baseline calibration, as in the Section 5.2, and reaches 21% in the highest scenario, as compared with 19% in the specification of the main text. The global cost is thus somewhat amplified, but not by much. The calibrated values obtained for the production threshold $\overline{\alpha}$ and for the demand conditions across classes of firms in Panel (A) of Table G2 show that selection is somewhat reduced in the sense that firms with lower values of α remain active in equilibrium. If anything, average firm size is slightly higher across all scenarios, as is the $Q-$productivity index, which implies that firms tend to adjust to the minimum wage constraint by increasing their productivity by more than in the full-employment scenario that is presented in the main text.

Table G1: Baseline Calibration to French Data

(A) Empirical Targets

Moment	Data	Model
Share of agents who are teachers	0.098	0.098
Share of workers in new firms	0.203	0.222
Share of 2-layer firms	0.446	0.448
Share of 3-layer firms	0.228	0.228
Average firm size	93.48	92.31
Average size of 2-layer firms	87.23	87.04
Average size of 3-layer firms	214.00	215.46
Average (hourly) wage (€)	20.21	20.26
Share of workers (in firms) bound by minimum wage	0.047	0.047
Unemployment rate (France, 2006)	0.088	0.088

(B) Calibrated Values of Parameters

Parameter	Value
Cost of knowledge acquisition, c	0.220
Time cost of problem communication, h	0.194
Output per problem solved, A ($\sim$ 100s of €/hour)	0.240
Fixed cost of entry, f_E	5.920
Fixed cost of production, f	5.785
Minimum wage $\bar{w}$, relative to average wage	0.914
Job search parameter χ	135.84
2nd-node of α distribution ($\alpha = 50$)	-12.13
3rd-node of α distribution ($\alpha = 500$)	-15.43
4th-node of α distribution ($\alpha = 1000$)	-24.55

Notes: Panel (A) shows the empirical moments to which the model detailed in Appendix G is calibrated. They are all computed from the DADS, except the share of teachers, which is sourced from INSEE's Tableaux de l'Economie Française (2020 edition, data for 2018), as in the main text (see Section 5). Panel (B) displays the values of the structural parameters that are obtained in the calibration.

Table G2: Minimum Wages with Unemployment in General Equilibrium

(A) Aggregate Implications

	(1) Baseline	(2) No MW	(3) + 4%	(4) + 8%	(5) + 16%	(6) + 24%
Minimum wage, relative to k	1.04	1.00	1.09	1.13	1.25	1.61
Net (real) wage, k (€/hour)	17.74	17.74	17.73	17.68	17.18	14.23
Average wage per worker, $k(cz+1)$	20.26	20.25	20.49	21.02	21.93	22.98
Share of firms bound by MW	0.19	0.00	0.41	0.57	0.94	1.00
Share of workers bound by MW	0.05	0.00	0.46	0.85	0.93	0.99
(as shares of total number of agents)	*Occupational choices and total skills:*					
Fixed-cost workers	0.24	0.24	0.24	0.24	0.24	0.23
Workers/managers	0.58	0.58	0.57	0.55	0.52	0.41
Firms (entrepreneurs), M	0.01	0.01	0.01	0.01	0.01	0.00
Teachers ($\propto$ total skills in economy)	0.09	0.09	0.10	0.11	0.15	0.26
Unemployed	0.09	0.09	0.09	0.09	0.09	0.11
(In real terms when relevant)	*Production and productivity:*					
Total output ($\sim$ €/hour)	12.13	12.13	12.09	11.99	11.71	9.64
Profits per firm ($\sim$ €/hour)	5.66	5.66	5.65	5.56	5.49	5.93
Average firm size (in jobs)	92.31	92.37	90.32	84.92	82.98	85.90
Average revenue per worker (€/hour)	28.38	28.32	28.86	29.63	29.91	30.43
Q-productivity index, $q/C_L(q,1)$ (€/hour)	17.84	17.84	17.82	17.76	17.35	14.21
Production cutoff, $\overline{\alpha}$	51.67	51.67	51.59	51.16	51.71	62.11

(B) Firms' Organizations

	(1) Baseline	(2) No MW	(3) + 4%	(4) + 8%	(5) + 16%	(6) + 24%
(as shares of total number of firms)	*Structure of firm population:*					
1-layer firms	0.32	0.32	0.36	0.43	0.53	0.50
1-layer firms with $n_1^0 < e^{\lambda \overline{z}_0}/h$	0.00	0.00	0.00	0.00	0.00	0.90
2-layer firms	0.45	0.45	0.46	0.57	0.47	0.50
3-layer firms	0.23	0.23	0.18	0.00	0.00	0.00
	Hierarchical organizations:					
Fraction of agents in $l = 0$	0.51	0.51	0.50	0.50	0.49	0.40
Fraction of agents in $l = 1$	0.07	0.07	0.06	0.05	0.03	0.00
Fraction of agents in $l = 2$	0.00	0.00	0.00	0.00	0.00	0.00
(Share bound by MW)	*Minimum Wage constraint at the firm level:*					
Firms with $L = 1$	0.00	0.00	0.00	0.00	0.88	1.00
Firms with $L = 2$	0.43	0.00	0.50	1.00	1.00	1.00
Firms with $L = 3$	0.00	0.00	1.00	1.00	-	-
(Share bound by MW)	*Minimum Wage constraint at the worker level:*					
Firms with $L = 1$	0.00	0.00	0.00	0.00	0.09	0.16
Firms with $L = 2$	0.00	0.05	0.08	0.84	0.84	0.83
Firms with $L = 3$	0.00	0.00	0.38	0.01	0.00	0.00
(Relative to k)	*Wages of entrepreneurs*					
In 1-layer firms L1	1.95	1.95	1.95	1.96	2.02	2.12
In 2-layer firms L2	2.27	2.27	2.31	2.41	2.48	2.53
In 3-layer firms	2.54	2.54	2.55	2.58	-	-
(Relative to k)	*Wages of workers and managers:*					
In $l = 0$	1.08	1.08	1.09	1.13	1.25	1.61
In $l = 1$	1.58	1.57	1.62	1.73	1.75	1.73
In $l = 2$	1.79	1.79	1.79	1.83	-	-

Notes: This table presents the results from simulations computing the general equilibrium of the version of our model that incorporates job search and unemployment benefits (as described in Appendix G). Total output can be interpreted as per agent and per hour, in €: $P. \int M\mu(\alpha)q(\alpha)\mathrm{d}\alpha$, since P is normalized to 1. Average output per worker (and per hour) corresponds to $\frac{P. \int M\mu(\alpha)q(\alpha)\mathrm{d}\alpha}{\# \text{ Workers} + \# \text{ Managers}}$. Average revenue per worker (and per hour) corresponds to $\frac{p(\alpha). \int M\mu(\alpha)q(\alpha)\mathrm{d}\alpha}{\# \text{ Workers} + \# \text{ Managers}}$.

H Framework without Organizations: Unproductive Managers

This Appendix aims to compare the quantification of our main model specification with a comparable set-up where the endogenous organizational (productivity) channel is essentially shut down. In this framework however, we continue to assume that knowledge (skills) remains both relevant for production ($\lambda < \infty$), and costly to acquire ($c \geq 0$). We also continue to assume that managers are essential to "bundling" production workers in firms, which enables them to reach their optimal size given their idiosyncratic demand shifter, α. However, in contrast to the specification in the main text, we now assume that they are otherwise irrelevant for production such that parameter h becomes irrelevant.[90]

In such a setting, no firm finds it optimal to increase the number of managerial layers beyond $L = 1$, and managers do not need skills, but all firms still have to optimize n_1^0, the number of production workers, as well as their level of skills z_1^0. The firms' cost minimization program thus takes the following (simplified) form:

$$C(q, k) = \min_{n_1^0, z_1^0} C_1(q, k) = kn_1^0(cz_1^0 + 1) + k \tag{120}$$

subject to:

$$\overline{z}_1^0 \leq z_1^0 \tag{121}$$

$$q \leq n_1^0 A(1 - e^{-\lambda z_1^0}), \tag{122}$$

where $\overline{z}_1^0 = 0$ in the economy without minimum wage, and $\overline{z}_1^0 = \frac{\overline{w} - k}{ck}$ (as in Equation 21) when a minimum wage is introduced.

Solving this problem shows that the optimal level of knowledge of production workers, z_1^0, verifies the following relationship:

$$\frac{c}{\lambda} e^{\lambda z_1^0} \left(1 - e^{-\lambda z_1^0}\right) = cz_1^0 + 1 \qquad \text{if} \qquad z_1^0 > \overline{z}_1^0 \tag{123}$$

$$z_1^0 = \overline{z}_1^0 \quad \text{otherwise} \tag{124}$$

In particular, Equations 123 and 124 do not depend on α, the only parameter that is heterogeneous across firms, which implies that all firms select the same level of skills for any of their production workers. Therefore, in such an economy, there is no wage heterogeneity, neither within firms, nor across firms. In addition, Equations 123 and 124 do not depend on k either, which implies that the level of skills in unaffected by the

[90]This implies that $h = 0$ is an admissible value and that the condition $c > \lambda h$ still holds formally.
Notice that the set-up of this Appendix is a little less extreme than the limit case considered in Caliendo and Rossi-Hansberg (2012), where knowledge, and not only managers, becomes irrelevant for production: $\lambda \to \infty$, and $c \geq 0$ (which implies that $c \leq \lambda h$, ie. the parametric condition imposed by the French data no longer holds). In such a setting, the distribution of production problems becomes a mass point located at $d = 0$ and there is no heterogeneity in terms of the knowledge that is required to solve them all. Both workers' and managers' knowledge become irrelevant. In such a setting, organizations still emerge as soon as their production is higher than A, but only $L = 1$ managerial layer is required to "bundle" the optimal number of production workers (as in this Appendix). In addition, the optimal level of skills is trivially $z_1^0 = 0$ whatever the idiosyncratic demand conditions, α, which implies that all workers earn the same wage, k.
In such a setting, minimum wages start to be binding as soon as their value exceeds k. As in the model of the present Appendix, firms have to train all workers up to the level $\overline{z}_1^0 = \frac{\overline{w} - k}{ck}$, although this strategy is useless to them. Introducing such a binding minimum wage in this environment however appears to have drastic consequences: indeed, no complex firm survives, because the free entry condition (Equation 19) and the zero-profit cutoff equation (Equation 18) become generically incompatible. This result is due to the fact that, as soon as a minimum wage is introduced, the profit function becomes homogeneous of degree 1 in k, which implies that the general equilibrium price adjustments in k become entirely incapable of balancing either the free entry condition or the zero-profit cutoff equation. The cutoff $\overline{\alpha}$ is the only endogenous variable that is left to balance these two generically incompatible equations, which implies that no equilibrium exists where both minimum wages are binding and complex firms (with $L = 1$ managerial layer) are active. Rather, the cost of introducing any minimum wage is immediately huge: it forces the economy to turn into a state where only $L = 0$ firms (ie. self-employed agents, who are not bound by the minimum wage) operate, thus reducing drastically the production of all varieties (to A), even those which are highly valued by consumers.

general equilibrium price adjustments.

The profit maximization program in this set-up remains identical to Equation 15. It implies that in equilibrium, both $n_1^0(\alpha, k)$ and $q(\alpha, k)$ are linear in α:

$$n_1^0(\alpha, k) = \alpha k \left(\frac{\sigma - 1}{\sigma}\frac{\lambda}{c}\frac{A}{k}\right)^{\sigma} \frac{e^{-\sigma\lambda z_1^0}}{A(1 - e^{-\lambda z_1^0})} \tag{125}$$

$$q(\alpha, k) = \alpha k \left(\frac{\sigma - 1}{\sigma}\frac{\lambda}{c}\frac{A}{k}\right)^{\sigma} e^{-\sigma\lambda z_1^0}, \tag{126}$$

where, again, z_1^0 is an endogenous object but appears to be determined independently of α and k (see Equations 123 and 124).

Minimum wages. When minimum wages are introduced in such an economy and become binding, skills increase mechanically following Equation 124. As there is no wage heterogeneity between workers, the latter are all paid exactly the minimum wage, as soon as the latter begins to be binding. In addition, firms' sizes adjust according to Equations 125 and 126, where $z_1^0 = \overline{z}_1^0$: the increase in production workers' skills has a negative impact on both n_1^0 and q, but the latter is dampened by the general equilibrium adjustment of prices, k.[91]

To compare the qualitative and quantitative predictions arising from this set-up with our main results, we calibrate this model to the French data. As shown in Table H1, we rely on the same data moments as in the main text, except those relating to the number of layers in firms (since the number of layers is at most one in the present set-up) and to the share of workers who are bound by the minimum wage (since this share can only be 100% in the present set-up).[92] As a consequence, in order to achieve (just) identification, we have to reintroduce two additional targets, mainly to identify the compensation of teachers relative to workers and the distribution of α.[93] In the absence of public data on the former aspect and to ease the comparison with our baseline calibration, we simply introduce the value that is attained in the main text and use it as a calibration target. In terms of the distribution of α, we abandon the flexible specification of our main calibration and simply reintroduce the Pareto assumption that was used in Caliendo and Rossi-Hansberg (2012), so that $G(\alpha) = 1 - \alpha^{-\gamma}$. We also follow their calibration procedure in this respect, and thus introduce a calibration

[91] Lastly, both the increase in skills and the decrease in firm sizes alter the free entry condition (Equation 19) and the zero-profit cutoff equation (Equation 18), which in turn determine the production cut-off $\overline{\alpha}$.

A specific feature of this set-up with "unproductive managers" is that the production cut-off $\overline{\alpha}$ is unaffected by minimum wages when the distribution of α is a Pareto, $G(\alpha) = 1 - \alpha^{-\gamma}$. This is essentially due to the fact that marginal costs are constant (in contrast to the set-up in the main text). To see this, start from the fact that, evaluated at the optimal q, the profit function can be written as:

$$\pi(\alpha, k) = \frac{\alpha k}{\sigma - 1}\left(\frac{\sigma}{\sigma - 1}\right)^{-\sigma} \phi^{1-\sigma} - k(1 + f)$$

where $\phi = \frac{kc}{A\lambda e^{-\lambda z_1^0}}$ is the (constant) marginal cost. Plugging this into the zero-profit condition on entry and simplifying gives us the following expression for the production cut-off:

$$\overline{\alpha} = \left(\frac{1 + f}{(\gamma - 1)\delta f_E}\right)^{\frac{1}{\gamma}}$$

which varies only with fixed parameters γ, δ, f, and f_E - but not $\overline{z_1^0}$. Notice that this result does not hold when managers are productive (ie in the model of the main text where the distribution of α is constrained to a Pareto).

[92] In other terms, the set-up without productive managers is not able to rationalize the empirical share of workers who are bound by the minimum wage (which is only 5% in Table 5) and the existence of firms with several hierarchical layers (68% in Table 5).

[93] Additionally, we present the calibrated minimum wage is absolute terms rather than as a percentage of the average wage (since the latter is 1 by definition).

target corresponding to the coefficient obtained from a log-log regression of the cumulative distribution of firm sizes (-1.095 in their Figure VII), which ultimately identifies the Pareto coefficient, γ.

Calibration results. The calibration results are displayed in Panel (B) of Table H1. The model neglects both hierarchical organizations and wage heterogeneity, and thus pools high-skilled workers in the same category as low-skilled workers. This implies that the cost associated with knowledge acquisition is underestimated, at a value that is (roughly) ten times lower than in the main specification. In addition, in the model, high-skilled workers are assumed to be as likely as low-skilled workers to solve production problems, which implies that the overall productivity parameter, A is estimated to be approximately three times larger than in the main specification. The fixed costs of entry are also estimated to be approximately ten times higher than in the main specification, while the fixed costs of production are in contrast in the same order of magnitude.

Table H1: Calibration of Model with Unproductive Managers

(A) Empirical Targets

Moment	Data	Model
1. Share of agents who are teachers	0.098	0.98
2. Share of workers in new firms	0.203	0.213
3. Average firm size	93.48	93.51
4. Average (hourly) wage (€)	20.21	20.19

(B) Additional Targets

Moment	Target	Model
1. Average wage, relative to k (from main calibration)	1.142	1.153
2. Coefficient for log-log regression of cumulative distribution of firm size from Caliendo and Rossi-Hansberg (2012)	-1.095	-1.131

(C) Calibrated Values of Parameters

Parameter	Value
Cost of knowledge acquisition, c	0.037
Output per problem solved, A ($\sim 10^2$ €/hour)	0.739
Fixed cost of entry, f_E	61.541
Fixed cost of production, f	6.244
Minimum wage $\bar{w}$	0.202
Pareto parameter for α, γ	1.090

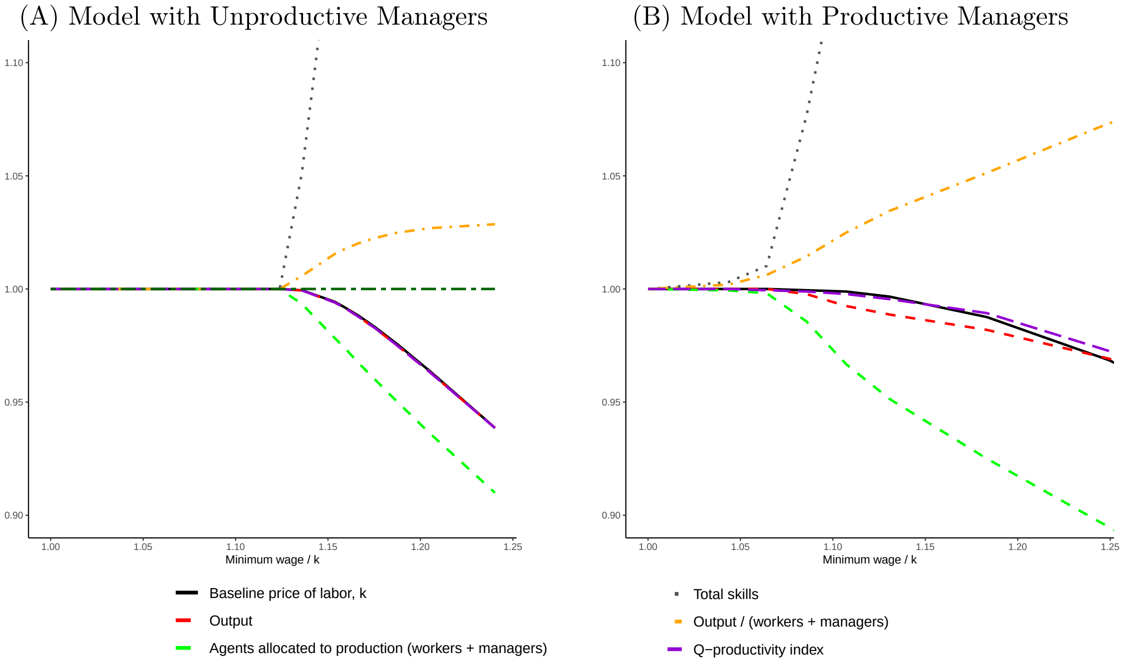
Notes: In Panel (A), all moments are computed from the DADS, except the share of teachers, which is sourced from INSEE's Tableaux de l'Economie Française (2020 edition, data for 2018), as in the main text (see Section 5).

Lastly, Figure H1 replicates Figure 6 and simulates the impact of increasingly stringent minimum wages in such an economy on the main aggregate outcomes. Given the lack of heterogeneity in wages across firms in this setting, firms find it optimal to train their workers (and pay them accordingly) even in absence of a minimum wage. As a consequence, we obtain that the minimum wage only starts to be binding for values above 1.12 k, while it started to be binding for at least some workers for all values exceeding k in the main calibration. At this point, skills adjust even more quickly than in the baseline set-up of the main text. Most importantly, the absence of the organizational channel[94] implies that output per worker increases only moderately, thanks to skills only. As a consequence, the $Q-$productivity index and aggregate output both

[94] Hierarchical layers range between $L = 0$ and $L = 1$, such that the scope for adjustment is extremely limited.

decrease four times faster than in the baseline set-up, when measuring this slope between the point where the minimum wage starts to be binding and the maximal value that is considered in Figure H1 ($1.24k$).

Figure H1: Comparison of Models with *vs.* w/o Organizational Productivity Channel



Notes: This figure presents simulations results for alternative scenarios where the minimum wage is either removed or set at increasingly high levels (relative to k, the baseline price of labor in general equilibrium). Panel (A) considers the economy which is described in this Appendix, where managers do not contribute directly to production (but remain essential to "bundle" workers into firms in cases where it is optimal to achieve large output targets given the value of the demand shifter). Panel (B) replicates Figure 6 for comparison. It describes the economy where managers contribute directly to production such that complex firms' organizations emerge in equilibrium, as in the main text.