The Impact of Regulation on Innovation
Online Appendix

Philippe Aghion   Antonin Bergeaud   John Van Reenen

Table of contents

A More Details of some Size-Related Regulations in France

The size-related regulations are defined in four groups of laws. The *Code du Travail* (labor laws), *Code du Commerce* (commercial law), *Code de la Sécurité Sociale* (social security) and in the *Code General des Impots* (fiscal law). The main bite of the labor (and some accounting) regulations comes when the firm reaches 50 employees. But there are also some other size-related thresholds at other levels. The main other ones comes at 10-11 employees.

For this reason we generally trim the analysis below 10 employees to mitigate any bias induced in estimation from these other thresholds. For more details on French regulation see *inter alia* Abowd and Kramarz (2003) and Kramarz and Michaud (2010), or, more administratively and exhaustively, Moins (2010).

A.1 Main Labor Regulations

The unified and official way of counting employees has been defined since 2004[1] in the *Code du Travail*,[2] articles L.1111-2 and 3. Exceptions to the 2004 definition are noted in parentheses in our detailed descriptions of all the regulations below. Employment is taken over a reference period which from 2004 was the calendar year (January 1^{st} to December 31^{st}). There are precise rules over how to fractionally count part-year workers, part-time workers, trainees, workers on sick leave, etc. (Moins, 2010). For example, say a firm employs 10 full-time workers every day but in the middle of the year all 10 workers quit and are immediately replaced by a different 10 workers. Although in the year as a whole 20 workers have been employed by the firm the standard regulations would mean the firm was counted as 10 employee firm.

In this case, this would be identical to the concept used in our main source of firm level data: FICUS. Garicano, Lelarge and Van Reenen (2016) extensively document that the discontinuity in the firm size distribution at 50 can be seen across a variety of alternative datasets with different definitions of employment. There is of course more measurement errors in some datasets than others due to differences in how the employment concept matches the regulatory definitions. Still, one concern is that there is under-reporting of employment in FICUS in order to avoid the regulation, we discuss this point in Appendix A.2. However, it is important to stress that what is crucial for our identification is not the exact position of the excess mass of

[1]Before that date, the concept of firm size was different across labor regulations.
[2]The text is available at the legifrance website

firms at precisely 49 employees (which is subject to measurement error and possible misreporting), but rather that the size distribution shifts around the region of the regulatory threshold, which it does robustly across a wide variety of datasets and employment size concepts.

Finally, recall that the employment measure in the FICUS data is average headcount number of employees taken on the last day of each quarter in the fiscal year (usually but not always ending on December 31st). All of these regulations strictly apply to the firm level, which is where we have the FICUS data. Some case law has built up, however, which means that a few of them are also applied to the group level.

From 200 employees

- Obligation to appoint nurses (Code du Travail, article R.4623-51)
- Provision of a place to meet for union representatives (*Code du Travail*, article R.2142-8)

From 50 employees:

- Monthly reporting of the detail of all labor contracts to the administration (*Code du Travail*, article D.1221-28)
- Obligation to establish a staff committee ("comité d'entreprise") with business meeting at least every two months and with minimum budget = 0.3% of total payroll (*Code du Travail*, article L.2322-1-28, threshold exceeded for 12 months during the last three years)
- Obligation to establish a committee on health, safety and working conditions (known as CHSCT) (*Code du Travail*, article L.4611-1, threshold exceeded for 12 months during the last three years)
- Appointing a shop steward if demanded by workers (*Code du Travail*, article L.2143-3, threshold exceeded for 12 consecutive months during the last three years)
- Obligation to establish a profit sharing scheme (*Code du Travail*, article L.3322-2, threshold exceeded for six months during the accounting year within one year after the year end to reach an agreement)
- Obligation to do a formal "Professional assessment" for each worker older than 45 (*Code du Travail*, article L.6321-1)
- Higher duties in case of an accident occurring in the workplace (*Code de la Sécurité sociale* and *Code du Travail*, article L.1226-10)

- Obligation to use a complex redundancy plan with oversight, approval and monitoring from Ministry of Labor in case of a collective redundancy for 9 or more employees (*Code du Travail*, articles L.1235-10 to L.1235-12; threshold based on total employment at the date of the redundancy)

From 25 employees:

- Duty to supply a refectory if requested by at least 25 employees (*Code du Travail*, article L.4228-22)
- Electoral colleges for electing representatives. Increased number of delegates from 25 employees (*Code du Travail*, article L.2314-9, L.2324-11)

From 20 employees:

- Formal house rules (*Code du Travail*, articles L.1311-2)
- Contribution to the National Fund for Housing Assistance;
- Increase in the contribution rate for continuing vocational training of 1.05% to 1.60% (*Code du Travail*, articles L.6331-2 and L.6331-9)
- Compensatory rest of 50% for mandatory overtime beyond 41 hours per week

From 11 employees:

- Obligation to conduct the election of staff representatives (threshold exceeded for 12 consecutive months over the last three years) (*Code du Travail*, articles L.2312-1)

From 10 employees:

- Monthly payment of social security contributions, instead of a quarterly payment (according to the actual last day of previous quarter);
- Obligation for payment of transport subsidies (Article R.2531-7 and 8 of the General Code local authorities, *Code general des collectivités territoriales*);
- Increase the contribution rate for continuing vocational training of 0.55% to 1.05% (threshold exceeded on average 12 months).

Note that, in additions to these regulations, some of the payroll taxes are related to the number of employees in the firm.

The additional requirements depending on the number of employees of the firm, but also limits on turnover and total assets are as follows (commercial laws, *Code du Commerce*, articles L.223-35 and fiscal regulations, *Code général des Impôts*, article 208-III-3):

In addition to the labor laws there are also some accounting regulations which are relatively minor in scope. From 50 employees there is first, no longer the possibility of a simplified presentation of Schedule 2 to the accounts. Second, there is a Requirement for LLCs, the CNS, limited partnerships and legal persons of private law to designate an auditor. From 10 employees, the possibility of a simplified balance sheet and income statement is lost.

A.2 Misreporting and compliance with regulations

One concern is that firms would under-report their employment level in FICUS in order to avoid the regulation. As Garicano, Lelarge and Van Reenen (2016) discuss, there is a lot of scrutiny of the employment numbers by unions, government and other agents as well as significant fines for non-compliance.

In a recent work, Askenazy, Breda and Pecheu (2022) argue that employment is systematically under-reported in FICUS in order to avoid the regulation. They argue that discontinuities are not observed in the matched employer employee data which comes from the Social Security (DADS) when they have reconstructed it to match the labor regulations. By contrast, Garicano, Lelarge and Van Reenen (2016) extensively document that the discontinuity in the firm size distribution at 50 can be seen across a variety of firm datasets with different definitions of employment. For example, their Figure 6 shows that even in DADS there is evidence of a change in the power law of the firm size distribution at 50 employees.

The bunching of firm size just below 50 employees in DADS is less apparent than in FICUS, especially when the hours (rather than headcount) data are used to construct Full Time Equivalents. This is the measure re-constructed by Askenazy, Breda and Pecheu (2022). Unfortunately, the hours data in DADS is problematic with up to 18% of the values imputed, no inclusion of agency workers (which are relevant for the regulation) before 2018, no way to distinguish 35 vs. 39 hour contracts, etc... Hence, the reconstructed measure using hours could also be generating significant measurement errors disguising the discontinuity at 50.

Rather than viewing the employment measure as the single “correct” one for regulatory purposes, we must recognize that all available datasets and empirical employment measures whether from FICUS or DADS has an imperfect mapping to the regulation. As documented above there are differences in how the employment concept matches the (multiple) regulatory definitions.

Fortunately, the methods in our paper (and those in Garicano, Lelarge and Van Reenen 2016) do not require to obtain the precise value of the cut-off in the empirical data. First, we need to only approximate that there is a margin of firms of employment below 50 who look different, not that this is exactly at 49. In the cross section, the innovation valley is around 45-49 employees, as is the differential response to market size shocks in the dynamic analysis.

Second, our approach utilizes differences away from the discontinuity at 50. In our baseline calibration we use the change in the gradient of the innovation-size relationship for firms in the 10 to 45 range vs. the 50-100 range to help identify the implicit tax of regulation. Similarly, in the extension where we calibrate the implicit tax using the dynamic analysis of the responses to export market size shocks, we use the data away from the threshold, again comparing responsiveness of smaller to larger firms. Hence, the identification of the aggregate costs of the regulation does not rely closely on the firms to the left of the threshold, and is therefore robust to possible under-reporting.

Askenazy, Breda and Pecheu (2022) collect data on the presence of works councils and profit sharing schemes. They show, using the FICUS data that there is a discontinuous jump in the likelihood of such institutions exactly around the 50 threshold. This is clearly supportive of the fact that FICUS employment is relevant for the establishment of such institutions, and supports the idea that FICUS employment is very relevant.

Finally, it is possible to extend our approach to allow for misreporting. In particular, Garicano, Lelarge and Van Reenen (2022) augment Garicano, Lelarge and Van Reenen (2016) with a new "regime" in which firms can cheat by under-reporting their number of workers, but with a risk of being detected and of being fined for that behavior. They assume that the probability p of detection is homogeneous across firms and that the fine in case of being caught lying is linear, but that the largest firms will for sure get caught if they lie. Specifically, with probability $(1-p)$ the firm gets away with misreporting, while with probability p it pays a variable fine τ_{nc} (so the more you lie about size, the more you pay) plus a fixed fine f_{nc}. With these assumptions they prove that the estimation regime in Garicano, Lelarge and Van Reenen (2016) goes through. The only difference is in the interpretation of the magnitude of the fixed cost of regulation which now also includes a misreporting term. Since the estimates of the fixed cost are trivial in magnitude, the misreporting bias is also immaterial for the welfare calculations.

In summary, although there is no perfect measure of employment, the use of FICUS appears adequate for our purposes.

A.3 Alternative ways of modelling the regulation

We have modelled the regulation as a variable tax on profits. We consider three possible extensions: (i) modelling the regulation having a fixed cost component; (ii) modelling the regulation as a labor tax rather than a profit tax; (iii) including capital inputs

Modelling the regulation as having a fixed cost component. Introducing a fixed cost component in our modelling of the regulation, generates new empirical implications. For example, the fixed component would still generate a hump in the size distribution below 50 and a subsequent sharp drop in the density of firm size to the right of 50. It would not, however, generate a permanent downward shift in the slope of the firm density by size distribution. Intuitively, the fraction of very large firms would be essentially unchanged, as such firms could spread the fixed cost over a very large number of units. The data, by contrast, shows this downward shift very clearly (see Figure D1). This is consistent with a strong role for the variable cost. Garicano, Lelarge and Van Reenen (2016) structurally estimate the magnitude of the fixed cost of the regulation using employment data and find it to be very small (less than the wage of a single worker), with just about all the regulatory cost loaded on the variable component.

Overall, adding a fixed cost would not generate markedly different slopes of the size-innovation relationship for large versus small firms.

Modelling the regulation as a labor tax rather than a profit tax. Modelling the regulation as a marginal tax on the labor input instead of a tax on profits, does not affect our main theoretical predictions, even though it somewhat complicates the model. Indeed, when the regulation is modelled as a labor tax, firms' marginal costs of production will depend upon the labor regulation they face, which in turn depends upon the firm's size. Limit pricing then implies that the equilibrium profit on each line will depend upon both, the labor tax of the current leader and the labor tax of the fringe firm on that line. Overall, moving from our baseline model to the model with labor tax amounts to introducing an extra state variable, namely the share of firms above the regulatory threshold which itself is endogenous. While assessing the magnitude of the effects, requires solving this more complicated variant of our baseline model,[3] nonetheless the logic remains the same as in our baseline model: namely, just below the threshold, firms will anticipate that if they successfully innovate then they will move beyond the threshold and therefore be subject to the labor tax. An innovating firm does not know the labor tax of the fringe firm they will face on the corresponding line, yet they reason in expected terms and clearly the expected profit goes down when they themselves become subject to the labor tax. Hence we still predict an innovation valley just below the threshold. Similarly, the expected profit from innovating for firms above the threshold, is reduced when the regulatory labor tax is introduced, to an extent which increases with the size of the firm, i.e. with its level of employment. Hence, introducing the labor tax above the threshold should again reduce the slope of the size-innovation curve above the threshold.

Including capital inputs. The current model has only labor as a productive input. Adding capital (or other inputs) would not make any fundamental difference to our current set-up because the regulation is a profit tax, so increase size is isomorphic to getting a larger (absolute) profit, since there is only one producer per line and this monopolist earns a fixed markup per line. However, if we instead followed approach (ii) and modelled the regulation as an implicit tax on labor, then the regulation does have different effects on other inputs that are not taxed. In particular, there would be an incentive to substitute into non-labor factors of production in order to mitigate the regulatory cost. In our data, firms who are approaching the 50 threshold so tend to increase capital investments, so that they can grow without necessarily adding more employees. Similarly, they increase employee hours, add more skilled workers, etc.

If these margins of adjustment were perfect substitutes for raw labor this would completely unravel the effects of the regulation. However, since these factors are generally not perfect substitutes, there will be some cost to such strategies. To what extent this reduces the overall cost of the regulation and its impact on welfare is an empirical question, which will hinge on the elasticity of substitutability. The degree of substitutability is partly due to other regulations – for example, there are strict rules in France on the number of hours a worker can work per week which limits the increasing hours margin. But it is partly also constrained by the technology of production.

Appendix D and subsection 6.4 of Garicano, Lelarge and Van Reenen (2016) introduced capital in the context of a CES production function with regulations modelled as a labor tax with fixed and variable components (so combining points (i) and (ii) above). For an elasticity of

[3] In Appendix C.5 we compute the equilibrium innovation rate $z(n)$ where both the share of lines S in which the current producer is a "large" firm beyond the threshold and the threshold number of lines $\overline{n}$ beyond which an n - line firm is considered to be "large", are constant over time. We show that the expression for $z(n)$ is similar to that in the baseline model, except that it now also depends on S which is endogenous.

substitution of one (Cobb-Douglas) the output loss barely changed from the baseline case of no substitution (the implied variable tax fell from 3.1% to 3%). Since most econometric estimates of the capital-labor substitution elasticity are less than unity, even this 0.1 percentage point change is likely to be an overestimate. This is in the context of a static model, but since the magnitudes are so small, it is unlikely that it would be so much larger in our dynamic model.

B Data Appendix

B.1 Patent data

Our first database is PATSTAT (EPO, 2016) Spring 2016's version which contains detailed information about patent applications from every patent office in the world. Among the very rich set of information available, one can retrieve the date of application, the technological class, the name of the patent holder (the *assignee*, the entity which owns the intellectual property rights) and the complete list of forward and backward citations.

We use a crosswalk built by Lequien et al. (2017) that associates each patent whose assignee is located in France with the official identifying number (or *SIREN*), which enables us to use most administrative firm level datasets. This matching use supervised learning based on a training sample of manually matched patents from the French patent office (INPI). It has the advantage over other matching protocols as it is specific to French firms to exploits additional information such as the location of innovative establishments (see Lequien et al., 2017 or Aghion et al., 2018 for more details).[4]

Because we stop our patent analysis in 2007, we are not affected by the truncation bias toward the end of the sample (see Hall, Jaffe and Trajtenberg, 2005) and we consider that our patent information are complete. In order to be as close to the time of the innovation as possible, we follow the literature and consider the filing year and not the granting year in our study. We use citations through to the last year (2016). When calculating a firm's quantile in the patent citation distribution, we do this based on a technology class (32 codes) by cohort-year.

Finally, we consider every patent owned by a French firm, regardless of the patent office that granted the patent rights, but we restrict to *priority patents* which correspond to the earliest patents which relate to the same invention. Therefore, if a firm successively fills the same patent in different patent offices, only the first application of this family will be counted.

B.2 Firm-level administrative data

Our second data source provides us with accounting data for French firms from the DGFiP-INSEE, this data source is called `FICUS` (1994-2007) (Insee & DGFiP, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007) and completed with `FARE` (2008-2012) (Insee & DGFiP, 2008, 2009, 2010, 2011, 2012). The data are drawn from compulsory

[4] If the firm shares a patent with another firm, then we only allocate a corresponding share of this patent to the firm.

reporting of firms and income statements to fiscal authorities in France. Since every firm needs to report every year to the tax authorities, the coverage of the data is all French firms from 1994 to 2012 with no limiting threshold in terms of firm size or sales. This dataset provides us with information on the turnover, employment, value-added, the four-digit NACE sector the firm belongs to. This corresponds to around 35 million observations.

The manufacturing sector is defined as category C of the first level of the NAF (*Nomenclature d'Activités Franqise*), the first two digits of which are common to both NACE (Statistical Classification of Economic Activities in the European Community) and ISIC (International Standard Industrial Classification of All Economic Activities). INSEE provides each firm with a detailed principal activity code (APE) with a top-down approach: it identifies the 1-digit section with the largest value added. Among this section, it identifies the 2-digit division with the largest value-added share, and so on until the most detailed 5-digit APE code (INSEE, 2016). It is therefore possible that another 5-digit code shows a larger value-added share than the APE identified, but one can be sure that the manufacturing firms identified produce a larger value-added in the manufacturing section than in any other 1-digit section, which is precisely what we rely on to select the sample of most of our regressions. The 2-digit NAF sector, which we rely intensively on for our fixed effects, then represents the most important activity among the main section of the firm. Employment each year is measured on average within the year and may therefore be a non-integer number.

B.3 Trade data

Customs data for French firms. Detailed data on French exports by product and country of destination for each French firm are provided by French Customs (French customs and indirect taxation authorities (DGDDI), 2023). These are the same data as in Mayer, Melitz and Ottaviano (2014) but extended to the whole 1994-2012 period. Every firm must report its exports by destination country and by very detailed product (at a level finer than HS6). However administrative simplifications for intra-EU trade have been implemented since the Single Market, so that when a firm annually exports inside the EU less than a given threshold, these intra-EU flows are not reported and therefore not in our dataset. The threshold stood at 250 000 francs in 1993, and has been periodically reevaluated (650 000 francs in 2001, 100 000 euros in 2002, 150 000 euros in 2006). Furthermore flows outside the EU both lower than 1 000 euros in value and 1 000 kg in weight are also excluded until 2009, but this exclusion was deleted in 2010.

Country-product bilateral trade flows. CEPII's database BACI (Gaulier and Zignago, 2010), based on the UN database COMTRADE, provides bilateral trade flows in value and quantity for each pair of countries from 1995 to 2015 at the HS6 product level, which covers more than 5,000 products.

C Theoretical Appendix

In this Theory Appendix we first present numerical solutions for our baseline (one innovation type) model and then more details of the two types of innovation (radical and incremental) model. Next, we detail the multi-period lived owner model and finally the extension to R&D as scientists (rather than lab-equipment).

C.1 Numerical Solutions for the baseline model

We solve the model numerically. To do so, we need to discretize the problem. That is, we need to move from a model with a continuum of products of size 1 to a model with a finite number of products K and a finite number of firms N.

The final good aggregator is adjusted as follows:

$$\ln y = \int_0^1 \ln y_j dj \text{ becomes } \ln y = \frac{1}{K}\sum_{j=1}^{K} \ln y_j$$

Unite price of a given intermediate good j is unchanged, but the demand:

$$y_j = \frac{y}{p_j} \text{ becomes } y_j = \frac{y}{p_j K}$$

And as a result:

$$\pi_j = \left(1 - \frac{1}{\gamma}\right) y \text{ becomes } \pi_j = \left(1 - \frac{1}{\gamma}\right) \frac{y}{K}$$

Finally, firm i's employment L_i is still equal to $n/(\omega\gamma)$ where:

$$\omega = \frac{wK}{y}.$$

The firm's maximization problem is still:

$$n\pi(n) + \beta nz\left[(n+1)\pi(n+1) - n\pi(n)\right] + \beta nx\left[(n-1)\pi(n-1) - n\pi(n)\right] - \zeta z^{\eta} n \frac{y}{K}$$

where the R&D cost function in our finite line model

$$C(z,n) = \zeta n z^{\eta} y \text{ has become } C(z,n) = \frac{\zeta}{K} n z^{\eta} y$$

With these changes in mind, equation (4) still applies and we can numerically solve the model in steady state. We proceed as follows:

1. There is a finite number N of firms and K of product lines, with $K > N$
2. $\mu(n)$ denotes the number of firms producing in exactly n product lines and $z(i)$ denotes its innovation intensity per line (which is taken from equation (4) in the model).
3. All firms produce at least one product, as a result, we must have $\mu(n) = 0$ for all $n \geq K - N$. For all i larger than 1

We therefore have $K - N + 1$ unknowns: $\mu(n)$ for $1 \leq n < K - N$ ($K - N - 1$ unknowns), x and z_e. The corresponding $K - N + 1$ independent equations are given by:

- The law of motion for μ:

$$\mu(n) = \frac{(n-1)\mu(n-1)z(n-1) + \mu(n+1)(n+1)x}{n(x + z(n))},$$

  for all $n \geq 2$ and $n < K - N$, recalling that $\mu(K - N) = 0$

- The definition of μ:

$$\sum_{n=1}^{K-N-1} \mu(n) = N$$

- The definition of x

$$x = z_e + \sum_{n=1}^{K-N-1} z(n)n\mu(n)/K$$

- The steady-state equation for the number of firms in the economy

$$\mu(1)x = z_e K$$

C.2 Radical vs. Incremental innovation

This details the summarized discussion in the main text in subsection V.A where we allow firms to choose to invest in radical vs. incremental innovation.

Innovation equation. we solve for $u(n)$ and $z(n)$ by taking the first order condition from equation (13), where z is the output-adjusted effort invested in incremental R&D and u is the output-adjusted effort invested in radical R&D. This yields the following two equations:

$$u(n) = \left(\frac{\beta}{\alpha\eta}\left[(n+k)\pi(n+k) - (n+1)\pi(n+1)\right]\right)^{\frac{1}{\eta-1}}$$

and

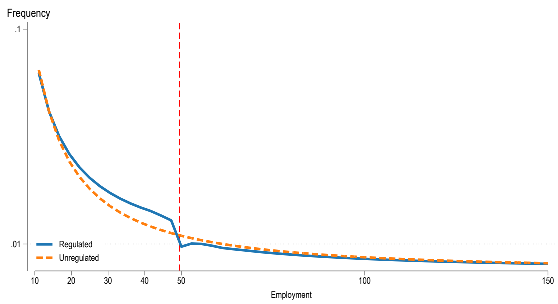
$$z(n) = \left(\frac{\beta}{\zeta\eta}\left[(n+1)\pi(n+1) - n\pi(n)\right]\right)^{\frac{1}{\eta-1}} - \left(\frac{\beta}{\alpha\eta}\left[(n+k)\pi(n+k) - (n+1)\pi(n+1)\right]\right)^{\frac{1}{\eta-1}}$$

With these two expressions, we can solve for the equilibrium size distribution and for the share of radical innovations over incremental innovations for each firm size. See Figures 8(a) and 8(b).

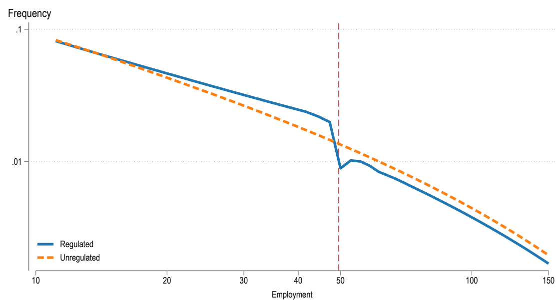
The equilibrium size distribution is depicted in Figure C1, first on a linear scale and then on a logarithmic scale. The size distribution is qualitatively similar to the baseline case.

Figure C1: Firm size distribution with two types of innovation

(a) Linear scale



(b) Log scale



Notes: These figures plot the value of $\mu(n)$ as a function of employment $L = n/(\gamma\omega)$. Left-hand side panel uses a linear scale and right-hand side panel a log-log scale. Extension with two types of innovation with $k = 4$ (see Section V.A)

A special case when $\eta = 2$. To give the intuition of what is happening in the model with two types of innovation, we first solve formally for u and z in equation (13) in the simple case where we take the overall cost of R&D to be quadratic and equal to $\zeta(u+z)^2 n/2 + \alpha u^2 n/2$. Thanks to the quadratic cost assumption, the first-order conditions can be conveniently summarized by the linear system:

$$\begin{pmatrix} \zeta & \zeta \\ \zeta & \alpha+\zeta \end{pmatrix} \begin{pmatrix} z \\ u \end{pmatrix} = \beta \begin{pmatrix} (n+1)\pi(n+1) - n\pi(n) \\ (n+k)\pi(n+k) - n\pi(n) \end{pmatrix}$$

As long as α and ζ are not equal to 0, this linear system solves into:

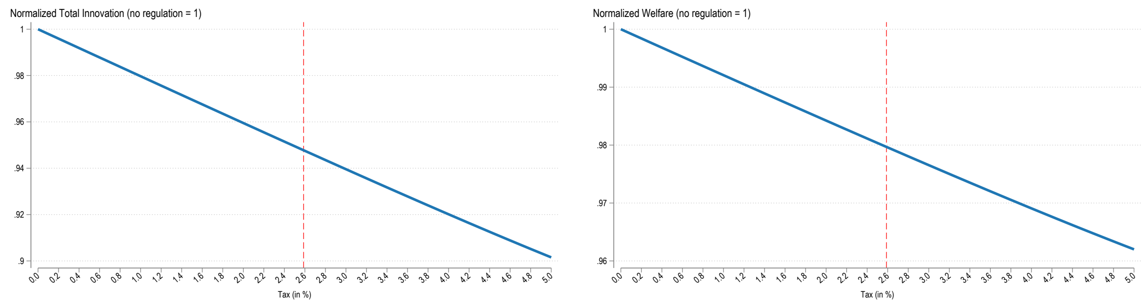
$$\begin{pmatrix} z \\ u \end{pmatrix} = \frac{\beta}{\zeta\alpha} \begin{pmatrix} \zeta+\alpha & -\zeta \\ -\zeta & \zeta \end{pmatrix} \begin{pmatrix} (n+1)\pi(n+1) - n\pi(n) \\ (n+k)\pi(n+k) - n\pi(n) \end{pmatrix}$$

Hence, we can see that the values of z and u will be impacted by the threshold as long as $n+k \geq \bar{n}$ and $n < \bar{n}$, which means that the regulation has a larger range of effects, but at the same time has a positive effect on u just before the threshold.

The solutions in the general case are presented in Table C1, where we have defined $\pi \equiv \frac{\gamma-1}{\gamma}\beta$

Figure C2: Total innovation and welfare with two types of innovation

(a) Total innovation (b) Welfare



Notes: These figures plot total innovation loss and total welfare loss in consumption equivalent against the value of τ. Parameter values can be found in Table C2.

Aggregate Effects on Innovation and Welfare. As discussed above and in the main text (see Section V.A) we can calibrate the general model using moments in our data and the literature. The parameter values are in Table C2 and the relationship between aggregate innovation, welfare and the regulation are in Figure C2. The aggregate losses are smaller than in the baseline model: 5.3% lower innovation and 2.1% lower welfare in the new two types of innovation model, compared to 5.7% and 2.2% in the baseline model. The lower losses are because radical innovation which creates greater social welfare is not discouraged. However, since the bulk of innovation is incremental, the aggregate losses are only modestly impacted.

Table C1: Solution in the Extended Model with two types of innovation (radical and incremental)

	$u(n)$	$z(n)$
$n < \bar{n} - k$	$\left(\frac{\pi}{\alpha\eta}(k-1)\right)^{\frac{1}{\eta-1}}$	$\left(\frac{\pi}{\zeta\eta}\right)^{\frac{1}{\eta-1}} - u(n)$
$\bar{n} - k \leq n < \bar{n} - 1$	$\left(\frac{\pi}{\alpha\eta}\left((k-1) - \tau(k+n)\right)\right)^{\frac{1}{\eta-1}}$	$\left(\frac{\pi}{\zeta\eta}\right)^{\frac{1}{\eta-1}} - u(n)$
$n = \bar{n} - 1$	$\left(\frac{\pi}{\alpha\eta}\left((k-1) + \tau(\bar{n}-k)\right)\right)^{\frac{1}{\eta-1}}$	$\left(\frac{\pi}{\zeta\eta}(1 - n\tau\right)^{\frac{1}{\eta-1}} - u(n)$
$n \geq \bar{n}$	$\left(\frac{\pi}{\alpha\eta}(1-\tau)(k-1)\right)^{\frac{1}{\eta-1}}$	$\left(\frac{\pi(1-\tau)}{\zeta\eta}\right)^{\frac{1}{\eta-1}} - u(n)$

	$u(n) + z(n)$	$\frac{u(n)}{z(n)+u(n)}$
$n < \bar{n} - k$	$\left(\frac{\pi}{\zeta\eta}\right)^{\frac{1}{\eta-1}}$	$\left(\frac{\zeta}{\alpha}(k-1)\right)^{\frac{1}{\eta-1}}$
$\bar{n} - k \leq n < \bar{n} - 1$	$\left(\frac{\pi}{\zeta\eta}\right)^{\frac{1}{\eta-1}}$	$\left(\frac{\zeta}{\alpha}(k-1)\left(1 - \frac{\tau(k+n)}{k-1}\right)\right)^{\frac{1}{\eta-1}}$
$n = \bar{n} - 1$	$\left(\frac{\pi}{\zeta\eta}(1 - n\tau\right)^{\frac{1}{\eta-1}}$	$\left(\frac{\zeta}{\alpha}(k-1)\left(1 - \frac{\tau(\bar{n}-k)}{k-1}\right)\frac{1}{1-\tau\bar{n}}\right)^{\frac{1}{\eta-1}}$
$n \geq \bar{n}$	$\left(\frac{\pi(1-\tau)}{\zeta\eta}\right)^{\frac{1}{\eta-1}}$	$\left(\frac{\zeta}{\alpha}(k-1)\right)^{\frac{1}{\eta-1}}$

Table C2: Calibrated parameter values in a model with two types of innovation

Parameter	**Value**
γ	1.3
η	1.5
ω	0.28
β/ζ	1.36
k	4
α	6.96
τ	0.0259

Notes: Calibration strategy is described in Section V.A.

C.3 Longer lived owners

In our baseline model, although firms can live forever we simplified the analytical problem by assuming the owners of firms only live for two periods. In this subsection, we show that the qualitative and quantitative predictions of the model carry over to a more complex environment where owners live longer.

C.3.1 Adding one extra period to the life of the owner

We first show how to extend our model by allowing firm owner to live for three periods instead of the two period baseline. In the first period, the owner inherits a firm of size n_1. She then chooses her level of innovation $Z_1(n_1) = n_1 z_1(n_1)$ and enters period 2 with a size n_2 (which can be either equal to n_1, n_1+1 or n_1-1). She chooses the level of innovation for period 2, $Z_2(n_2) = nz_2(n_2)$. In period 3, the owner collects profits, exits and ownership passes on to a new agent. Because the firm's owner only produces for two periods, we refer to this model as "the two period model" while the baseline model is denoted the "one period model".

We solve backwards: in period 2, the situation is the same as in the one period model and we know that for any size n, the innovation per line is:

$$z_2(n) = \left(\frac{\beta\pi}{\eta\zeta}\right)^{\frac{1}{\eta-1}} \times \begin{cases} 1 & \text{if } n < \bar{n}-1 \\ (1-\bar{n}\tau)^{\frac{1}{\eta-1}} & \text{if } n = \bar{n}-1 \\ (1-\tau)^{\frac{1}{\eta-1}} & \text{if } n \geq \bar{n} \end{cases} \qquad \text{where } \pi = \frac{\gamma-1}{\gamma}$$

In period 1, the firm maximizes the value function:

$$V_1(n) = \max_{z_1>0} \left\{ n\pi(n)y - nz_1^\eta \zeta y + \frac{1}{1+r}\mathbb{E}_{z_1}\left[V_2(n')\right] \right\},$$

where $V_2(n)$ is the value of being of size n in period 2.

$$V_2(n) = \max_{z_2>0} \left\{ n\pi(n)y - nz_2^\eta \zeta y + \frac{1}{1+r}\mathbb{E}_{z_2}\left[\pi(n')y'\right] \right\}.$$

We denote $v_i(n) \equiv V_i(n)/(ny)$ for $i = 1, 2$. Then, using the Euler equation we have:

$$v_2(n) = \pi(n)(1+\beta) + \beta z_2(\pi(n+1) - \pi(n)) + \beta x(\pi(n-1) - \pi(n)) - z_2^\eta \zeta$$

and finally:

$$z_1 = \left(\beta \frac{(n+1)v_2(n+1) - nv_2(n)}{\zeta\eta}\right)^{\frac{1}{\eta-1}}$$

It is thus possible to solve for equilibrium innovation given the number of lines in each period. Compared to the baseline case, the regulation will not only impact firms with a size $\bar{n}-1$ but also firms with a size $\bar{n}-2$ in period 1.[5]

Figure 11(a) plots the value of $\frac{z_1+z_2}{2}$, the average value of innovation per period, along with the value of z in the baseline model against employment. The main differences between the two is that in the two period model, the innovation valley is wider and extends to firms with an employment corresponding to $n=\bar{n}-2$. The fall in innovation at 49 employees is also less deep because the cost of the regulation is smoothed over two periods instead of one.

Size distribution. To solve for the size distribution, we look for solution where the distribution of firms in their period 2 is the same as the distribution of firms in their first period. We denote the share of firms of size n as $\mu(n)$ and we have at the steady-state:

$$\frac{\mu(n)}{2}\left(z_1(n)+z_2(n)\right)n=\frac{\mu(n-1)}{2}\left(z_1(n-1)+z_2(n-1)\right)(n-1)+(n+1)x\frac{\mu(n+1)}{2} \tag{C1}$$

Hence if we define $z=\frac{z_1+z_2}{2}$, the flow equation that determines the equilibrium size distribution is the same as in the baseline case. Figure 11(b) plots this distribution against the value of employment in the baseline case and in the two period model.

Calibration. In the baseline model, the calibration of τ, which governed the aggregate innovation loss followed directly from the comparison of the slopes of the innovation - firm size cross-sectional relationship in large vs. small firms. In our extended multi-period model, the calibration is slightly more involved and all parameters need to be estimated simultaneously. To illustrate why, we write the average innovation observed for a given size n, taken as the mean of the level of innovation for firms in their first and second period, respectively for large and small firms and consider the ratio.

$$\mathcal{R}=\frac{z_1(n\geq\bar{n})+z_2(n\geq\bar{n})}{z_1(n<\bar{n}-2)+z_2(n<\bar{n}-2)}$$

[5]In principle, the regulation can also impact firms with a size $\bar{n}$ in period 1 as they can reach a size $\bar{n}-1$ in the next period. However, we make the assumption that once the firm has crossed the threshold, the regulation continues to be enforced during the lifespan of the firm's owner, even if the firm becomes smaller than 50 (i.e. the regulations are "grandfathered". This simplifying assumption seems reasonable given that the nature of the regulation imposes many important adjustment costs that are hard to reverse.

For small firms, $z_1 + z_2$ is equal to:

$$\left(\frac{\beta\pi}{\zeta\eta}(1+\beta(1-x)) + \left(\frac{\beta\pi}{\zeta\eta}\right)^{\frac{\eta}{\eta-1}} \frac{\eta-1}{\eta}\beta\right)^{\frac{1}{\eta-1}} + \left(\frac{\beta\pi}{\zeta\eta}\right)^{\frac{1}{\eta-1}}$$

For large firms, $z_1 + z_2$ is equal to:

$$\left(\frac{\beta\pi(1-\tau)}{\zeta\eta}(1+\beta(1-x)) + \left(\frac{\beta\pi(1-\tau)}{\zeta\eta}\right)^{\frac{\eta}{\eta-1}} \frac{\eta-1}{\eta}\beta\right)^{\frac{1}{\eta-1}} + \left(\frac{\beta\pi(1-\tau)}{\zeta\eta}\right)^{\frac{1}{\eta-1}}$$

Hence:

$$\mathcal{R} = (1-\tau)^{\frac{1}{\eta-1}}\left(1-\mathcal{A}\right),$$

where $\mathcal{A} > 0$ is a function of model parameters and endogenous variable x which is small for small values of τ.[6] $\mathcal{A} > 0$ also implies that the value of τ required to ensure that the ratio of slopes match the data will be smaller than in the baseline model.

We use the same calibration strategy as our baseline approach. We set η and γ to the same values in Table 3 and set β to 0.96. We then use three moments from the data: $\mathcal{R}$, the long-term growth and the gap in the size distribution around the threshold to estimate the three remaining parameters τ, ω and ζ. Top panel of Table C3 displays the corresponding moments in the data and in the calibration.

The resulting parameter values are presented in the bottom panel of Table C3. They are very similar to those in the baseline model in Table 3, with the exception of the innovation cost ζ, which we discuss below. Most importantly, the key parameter τ is essentially unchanged, because we find $\mathcal{A}$ very close to 0 (0.0023), but slightly lower (2.58% vs 2.59% in the baseline).

Aggregate Innovation and Welfare. Once the value of μ and z are obtained using equation (C1), we can compute aggregate innovation and welfare. The formulae are the same as in the baseline model, but the value of innovation per line z is replaced by the average innovation over the two periods of production $(z_1 + z_2)/2$. The loss in total innovation and total welfare

[6]Formally:

$$\mathcal{A} = \frac{\left((1+\beta(1-x)) + \left(\frac{\beta\pi}{\zeta\eta}\right)^{\frac{1}{\eta-1}} \frac{\eta-1}{\eta}\beta\right)^{\frac{1}{\eta-1}} - \left((1+\beta(1-x)) + \left(\frac{\beta\pi}{\zeta\eta}\right)^{\frac{1}{\eta-1}} \frac{\eta-1}{\eta}\beta(1-\tau)^{\frac{1}{\eta-1}}\right)^{\frac{1}{\eta-1}}}{\left((1+\beta(1-x)) + \left(\frac{\beta\pi}{\zeta\eta}\right)^{\frac{1}{\eta-1}} \frac{\eta-1}{\eta}\beta\right)^{\frac{1}{\eta-1}} + 1}$$

Table C3: Calibrated parameter values in a model with two periods

Moments	**Data**	**Model**
Growth rate of GDP	1.63	1.62
Gap in the size distribution	3.05	2.96
Ratio of slope $\mathcal{R}$	0.949	0.947
Parameter	**Value**	**Baseline**
γ	1.3	1.3
η	1.5	1.5
ω	0.22	0.22
β/ζ	0.79	1.70
τ	0.0258	0.0259

Notes: Top panel presents the 3 moments used in the calibration of β/ζ, τ and ω. $\mathcal{R}$ corresponds to the ratio of the slope in the employment-innovation cross section for large firms vs small firms. Bottom panel presents the corresponding parameters. Calibration strategy is described in Section V.C.

are shown in Figures 12(a) and 12(b) along with the corresponding loss in the baseline model. The figures show that the loss in total innovation and welfare remains extremely similar in the new multi-period model compared to the baseline, essentially since the value of the implicit regulatory tax remains close to 2.6%.

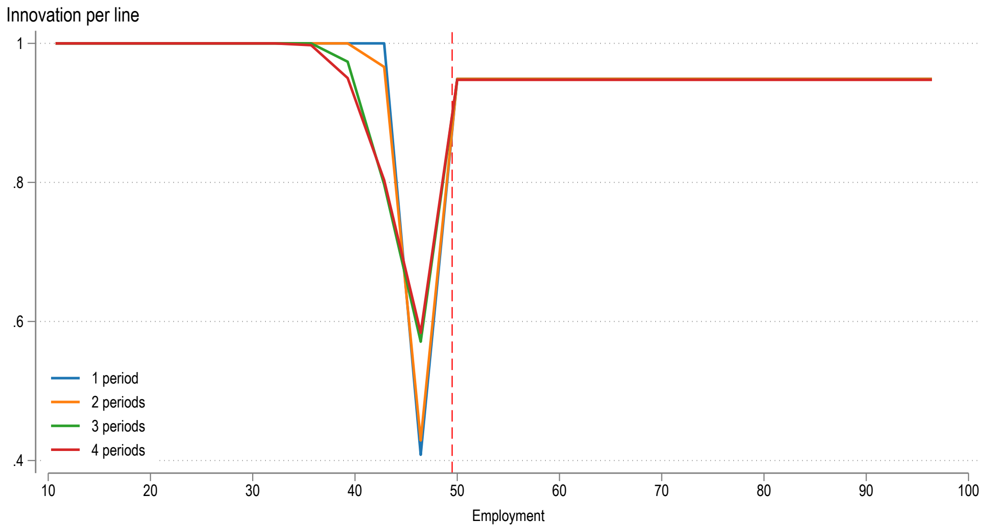
Figure 12(a) does show that the losses are slightly lower in the new model. This is because the calibration strategy must match the same empirical value of French growth of 1.63% (and hence total innovation). Because firms in the two period model will do more innovation on average each period, to be consistent with the empirical growth moment, the innovation cost parameter, ζ is estimated to be larger. Since ζ is held constant in the counterfactual unregulated economy, this means that overall innovation is reduced (slightly) less by any given regulatory tax.[7]

C.3.2 Adding more periods to the multi-period model

The model can be naturally extended to adding more periods to the firm owner's life through induction. Consider a case where a firm owner live for $k+1$ periods (k periods of innovation and one period where they simply collect profit). Then the value of innovation intensity in the last period z_k is the same as the value of z in the baseline model. The value of z_{k-1} comes from

[7]Note however, that the *level* of welfare for each value of τ is lower in the 2 period model than in the baseline model.

Figure C3: Innovation per line when adding more periods



Notes: Innovation per line in the baseline model and in models with 2, 3 and 4 production periods.

the difference between the value function in period k as in the two period model. Then using the value of z_{k-1} yields a value of v_{k-1} which gives a value of z_{k-2} and so on.

Formally, once we have solved for a J period model, extending to a model with $J+1$ periods can be done easily in the following way. If $V(n,p,J)$ and $z(n,p,J)$ respectively denote the value and innovation intensity of a firm with n lines in its period p in a model with J periods, then we have:

$$\forall n > 0, \forall p < J : V(n, p+1, J+1) = V(n, p, J) \text{ and } z(n, p+1, J+1) = z(n, p, J).$$

Hence, to move to a model with one more period, all we need is to solve for the first period which we do as in the two period model. Intuitively, as we extend the number of periods, the innovation valley widens and its depth reduces. This is illustrated in Figure C3 where we have reported the average value of innovation against size in the case of a 2, 3 and 4 period models, along with the baseline one period model.

Regarding the size distribution, we can use the same strategy as in the two period model by considering:

$$z = \frac{z_1 + \ldots + z_k}{k}$$

Generally speaking we have a sequence:

$$z(n, k-p-1, k)^{\eta-1} = \frac{\beta\pi}{\zeta\eta} + \beta\left(z(n, k-p, k)^{\eta-1}(1-x) + z(n, k-p, k)^{\eta}\frac{\eta-1}{\eta}\right) \qquad \forall p \in [[1, k-1]]$$

and

$$z(n,k,k)^{\eta-1} = \frac{\beta\pi}{\zeta\eta}$$

Importantly, if β is small enough, then extending the number of periods does not change the results materially from our baseline model.

C.3.3 Summary on extension to multi-period lived owners

We have shown that adding an extra period to a firm owner's life extends the "shadow" of the regulation further down the firm size distribution. As we might expect, the innovation valley becomes wider and flatter. A model calibration shows very similar aggregate innovation and welfare losses to our baseline case, however, suggesting that our simpler, more analytically tractable approach does not mislead us. Moreover, the theoretical findings on the shape of the innovation-size relationship generalize to having many more periods. Hence, our simple approach delivers losses in the right order of magnitude and is materially unchanged from moving to more complex dynamic models.[8]

C.4 R&D as scientific labor

This section solves the model outlaid in Section V.E. In this extension, R&D is performed by scientists, hence the workforce is now split between production and innovation workers. For each firm i, employment l_i is therefore given by:

$$l_i = \frac{n_i}{\omega\gamma} + \zeta n_i z_i^{\eta}. \tag{C2}$$

Aggregating over all firms, we get:

$$\mathcal{L} = \int_i l_i di = \frac{1}{\omega\gamma} + \zeta \int_i n_i z_i^{\eta} di = \frac{1}{\omega\gamma} + \zeta \sum_{n>0} \mu(n) n z^{\eta}(n)$$

Since $\mathcal{L}$ is fixed and exogenous, and since the right hand side terms of the above equation varies with the tax τ, then ω also varies with τ. More precisely, the equilibrium wage ω decreases with τ, since regulation costs decreases aggregate innovation (the second term of the right-hand side of the equation).

[8]In the working paper version, we show that qualitatively similar results are also found when considering another approach to modelling infinitely lived owners (Aghion, Bergeaud and Van Reenen, 2021).

Given that employment is now a function of both the number of products n and the intensity of innovation z, we denote it by $L(n, z)$. The cutoff threshold $\bar{l} = 50$ no longer corresponds to a single value of n but to a set of points in the space (n, z) such that:

$$z = \frac{1}{\zeta n}\left(\bar{l} - \frac{n}{\gamma\omega}\right) \tag{C3}$$

Figures C4 shows the equilibrium relationship between the number of products, employment and innovation intensity (which indirectly relates to the number of R&D workers). It is no longer possible to use the number of products as a measure of the size of the firms and we need to define profit per unit of final output is now equal to:

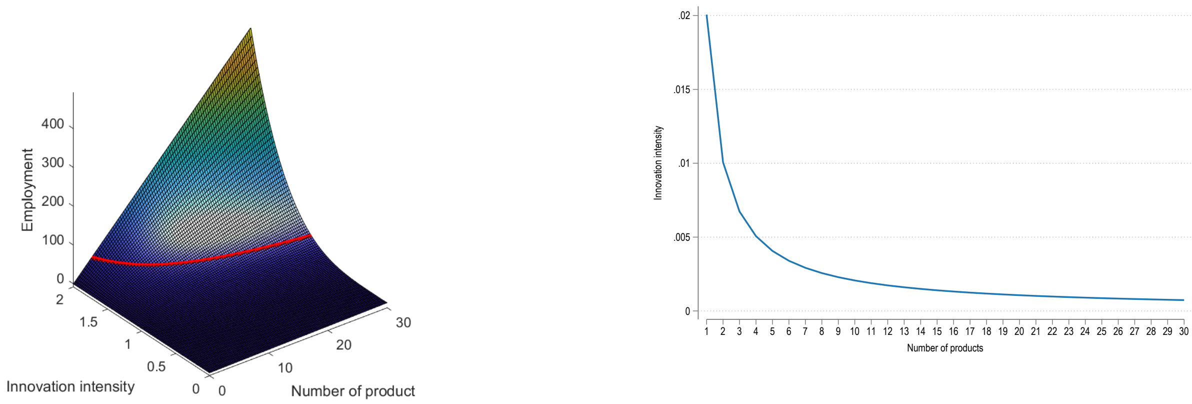
$$\pi(n, z) = \frac{\gamma - 1}{\gamma}\left(1 - \mathbb{1}\left[L(n, z) \geq \bar{l}\right]\tau\right).$$

Hence, the firm's maximization problem remains the same as before but with the two state variables n and z, that is:

$$\max_{n \geq 0, z \geq 0}\left\{n\pi(n, z)y - \zeta n z^{\eta} y + \frac{1}{1+r}\mathbb{E}\left[n'\pi(n', z')y'\right]\right\}.$$

Figure C4: Localization of employment threshold $\bar{l}$

(a) 3D plot　　(b) 2D projection

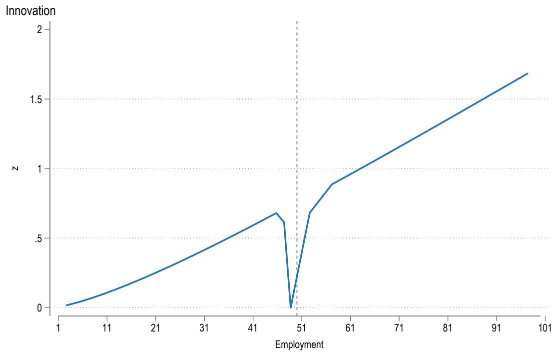


Notes: These Figures plot the relationship between employment L, innovation intensity z and number of products n. The left-hand side panel shows the 3D plot corresponding to the surface defined by equation (C2), where the z-axis corresponds to L. The curve in red corresponds to the intersection of the surface (n, z, L) with the surface $L = \bar{l}$. The right-hand side panel presents the set of pairs (z, n) which corresponds to an employment level of $\bar{l}$ according to equation (C3).

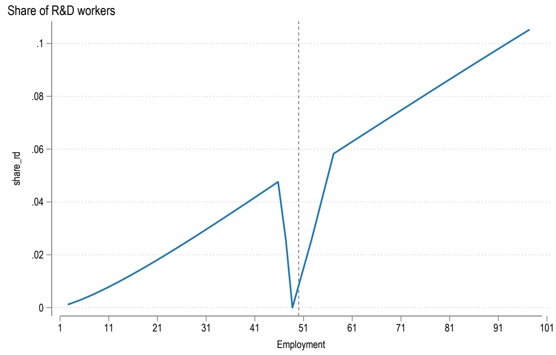
Solving this maximization problem for every value of n gives a function $Z(n) = nz(n)$ which we plot in Figure C5 against employment $L(n)$. We see that the innovation-employment cross-section relationship is qualitatively unchanged. In Figure C5, we also plot the corresponding relationship between firm's employment and its share of R&D workers.

Figure C5: Innovation-Employment cross-section with scientists $\bar{l}$

(a) Innovation



(b) Share of R&D workers



Notes: This is the total amount of innovation ($Z(n)$, left-hand side panel) and share of R&D workers in total employment ($\zeta n z^{\eta}(n)/L(n)$), right-hand side panel) by firms of different sizes (employment, $L = n/(\omega\gamma) + \zeta n z^{\eta}$) according to our theoretical model extension presented in Section V.E. We use arbitrary parameter values for illustrative purposes.

C.5 Modeling regulation as a tax on labor

In this extension, we let the regulation take the form of a marginal tax on the labor input so that the wage w becomes $w(1+\tau)$ if the firm crosses the threshold. Then, given that on each product line firms compete a la Bertrand, the incumbent producer $i(j)$ on a line j will set its price equal to the marginal cost of its competitor $i'(j)$, namely:

$$p_j = \frac{\gamma}{A_j} w(1+\tau_{i'(j)}) \text{ or } p_j = \frac{\gamma}{A_j} w,$$

depending on whether the competitor who was also the previous producer on the line, $i'(j)$, was larger or smaller than the threshold size. This yields:

$$y_j = \frac{A_j y}{\gamma w(1+\tau)} \text{ or } y_j = \frac{A_j y}{\gamma w}$$

The markup $m(j)$ on line j, defined as the unit price over unit cost depends on whether or not firm $i(j)$ is taxed and on whether the previous producer $i'(j)$ would be taxed, namely:

$$m(j) = \frac{\gamma(1+\tau_{i'(j)})}{1+\tau_{i((j))}}.$$

Such a line generates a profit per unit of final good which depends on the both, the labor tax $\tau_{i(j)}$ of the current producer and the labor tax of the previous producer $\tau_{i'(j)}$, i.e. on both, the size of the current producer $i(j)\tau_{i'(j)}$and the size of the previous producer $i'(j)$ on line j:

$$\pi(j) = 1 - \frac{1+\tau_{i(j)}}{\gamma(1+\tau_{i'(j)})}$$

Next, the equilibrium number of workers on a line also depends on the size of the fringe firm on that line. A line with a small fringe firm requires $y/(\gamma w)$ workers whereas a line with a large fringe firm only requires $y/(\gamma w(1+\tau))$ workers. Let us restrict attention to an equilibrium where S, the share of lines operated by a large firm, is constant. By the law of large numbers S also corresponds to the probability for an innovating firm, of facing a large firm as its competitive fringe on the corresponding line. It follows that total employment by a firm of size n is then equal to:

$$l(n) = n\frac{y}{\gamma w}\left(\frac{S}{1+\tau} + 1 - S\right).$$

This in turn implies that the threshold number of lines beyond which a firm is considered to be a large firm eligible to the regulatory labor tax, $\bar{n}$, is no longer constant but depends upon

S. Namely:

$$\bar{n} = \frac{\bar{l}\gamma w}{\left(\frac{S}{1+\tau} + 1 - S\right)}$$

Moving back to the R&D investment stage, for a given S at the steady state, a firm of size n will choose its innovation intensity z to maximize:

$$\Pi(n,S) + \beta n z(n)\left(\Pi(n+1,S) - \Pi(n,S)\right) - \beta n x\left(\Pi(n-1,S) - \Pi(n,S)\right) - n z^{\eta}\zeta,$$

where $\Pi(n,S) = n\left(1 - \frac{1+\tau\mathbb{1}(n\geq\bar{n})}{\gamma}\left(\frac{S}{1+\tau} + (1-S)\right)\right)$

In equilibrium, S is constant and equal to the share of products lines operated by large firms, namely:

$$S = \sum_{i>\bar{n}} \mu(i) i$$

where $\bar{n}$ itself depends upon S (see the above expression) and where μ follows a law of motion as in the baseline model. At the moment, we consider S, and therefore $\bar{n}$, to be constant and taken as given by the firm. This yields the following innovation-size cross relationship:

$$z(n,S) = \begin{cases} \left(\frac{\beta}{\zeta\eta}\left(1 - \frac{B}{\gamma}\right)\right)^{\frac{1}{\eta-1}} & \text{if } n < \bar{n} - 1 \\ \left(\frac{\beta}{\zeta\eta}\left(1 - \frac{B(1+\tau\bar{n})}{\gamma}\right)\right)^{\frac{1}{\eta-1}} & \text{if } n = \bar{n} - 1 \\ \left(\frac{\beta}{\zeta\eta}\left(1 - \frac{B(1+\tau)}{\gamma}\right)\right)^{\frac{1}{\eta-1}} & \text{if } n \geq \bar{n} \end{cases} \tag{C4}$$

where $B = \frac{S}{1+\tau} + (1-S) = 1 - \frac{\tau}{1+\tau}S < 1$

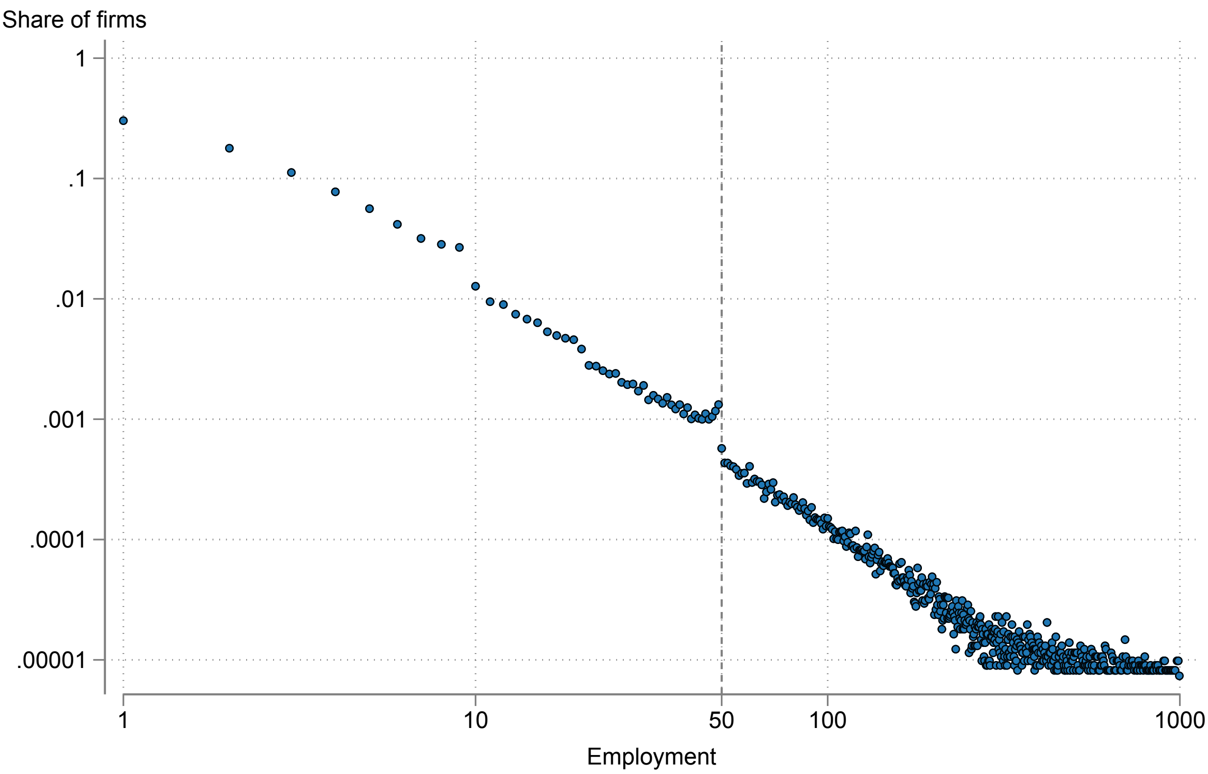
Hence the equilibrium $n \longrightarrow z(n,S)$ function looks similar to the equilibrium innovation-size relationship $z(n)$ in our baseline model with regulatory profit tax (see equation (4)).

D Additional Empirical Results

D.1 Size distribution of French firms

Figure D1 reports the size distribution of firms from FICUS in a log-log scale. In order to replicate results from Garicano, Lelarge and Van Reenen (2016), we use the year 2000, although choosing another year results in very similar relationship. The relationship is consistent with the well-know power law documented namely by Axtell (2001), but with two discontinuities: one at 50 employees and the other one at 10 employees, corresponding to size dependent regulation thresholds (see Appendix A).

Figure D1: Distribution of size



Notes: The data relate to the year 2000 for all firms.

D.2 Robustness of the cross-sectional innovation-size relationship

As noted in the main text the relationship between firm innovation and size are robust to a wide variety of alternative definitions. The baseline method in Figure 5 defines as innovative firm as one who has produced at least open patent over the sample period. In Figure D2 we consider using a narrower window around the year employment is measured. Panel A uses between patents in t, exactly the same year as employment as measured. Panel B uses patents filed two years before and two years after the employment measure (a five year window) and Panel C between four years before and after (a nine year window). Panel D measures innovation as log(1+patents count) in the same year as employment. Although the measures are somewhat noisier than using the whole period (which smooths things out), the same basic pattern of an innovation valley and a fall in the gradient after the regulatory threshold are apparent.

Figure D3 repeats these four definitions for the Figures comparing incremental and radical patents as measured by future citations (analogous to Figure 9).

D.3 Robustness of the dynamic effects of the market size shock on innovation

In the main text we noted the robustness of the decline in the impact of demand shocks to the left of the threshold and reported some of our tests. Here, we detail some more of these.

First, it is possible that the changing relationship between innovation and the market size shock around the threshold is driven by some kind of complex non-linearities in the innovation-employment relationship, and our quadratic controls are insufficient. To investigate this issue, we allow interactions between the demand shock and different size bins of firms in Table D1. Of all the 14 different size bins, only the interaction of the shock with the size bin just below the threshold (45-49 employees) is significantly different from zero and large in absolute magnitude.

Second, our results are robust to the particular way in which we define the upper and lower size cutoffs for our sample. Appendix Table D2 reproduces the baseline specification in column (1). Column (2) uses employment at t-2 instead of the initial year to define the sample, column (3) relaxes the upper threshold to include firms of up to 500 employees (instead of 100 employees in the baseline) and column (4) includes all firms below 100 employees (instead of dropping the firms with between zero and 9 workers). Column (5) restricts the sample to firms exporting in 1994 (instead of the restriction that a firm has to export in at least one year over the period 1994-2007). Column (6) includes all the non-exporting firms (see below for more details). The last three columns use three different definitions of the dependent variable instead of our basic measure $\tilde{\Delta} Y$: the log-difference in column (7), the difference in the Inverse Hyperbolic Sine in column (8) and the change in patents normalized on pre-sample patents in column (9). Our results are robust to all these tests.

Finally, one might be concerned that the quantiles of the citation distribution reported in Table 5 are arbitrary. Figure D4 reports the coefficient and confidence intervals on the key interaction term in our preferred specification for every quantile from the top 10th to the bottom 70th percentile in 5% intervals. As discussed in the text, it is clear that the negative effect of the regulation is only apparent for the less cited patents. There is no significant effect in a quantitative or statistical sense for patents in the top quartile of the citations distribution. The negative effect is driven by those in the bottom two-thirds of the citation distribution (with a monotonic decline of the effect for those between the 25th and 35th percentiles.

Extending to non-exporting firms. In column (6) of Table D2, we have extended the sample to all firms, while our baseline results restrict to exporting manufacturing firms. To do so, we need to calculate a demand shock for these firms that do not export. One natural way to do so would be to calculate the average demand shock at the sectoral level from firms that do export. However, our model includes sector-year fixed effects and even if the demand shock can be aggregated at a smaller sectoral level, most of the variance would be captured by these fixed effects.

To gain statistical power, we proceed as follows. We recalculate the same quantity as in equation (8) but at the sectoral level:

$$\Delta S_{k,t} = \sum_{s,c \in \Omega(k,1994)} \omega_{k,s,c,1994} \tilde{\Delta} I_{s,c,t}, \tag{C5}$$

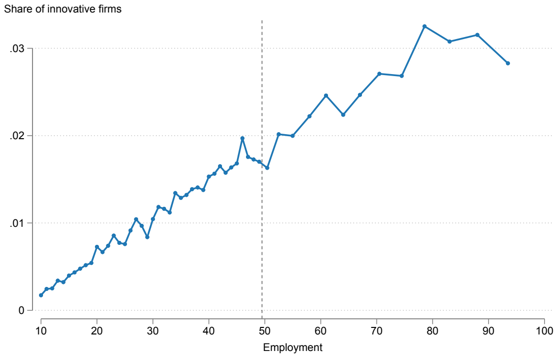
for each 5-digit sector k and year t. We use weights at the sector level taken during the year 1994 and covering all pairs of product-countries that firms in sector k exported to in 1994. Contrary to the baseline shock at the firm level $\Delta S_{i,t}$, we do not weight by the level of export intensity. Instead, we construct a weighted shock that used both $\Delta S_{k,t}$ and $\Delta S_{i,t}$ with weights depending on the level of export intensity of the firms:

$$\Delta S_{i,t}^{(k)} = \sigma_{i,t_0} \sum_{s,c \in \Omega(i,t_0)} \omega_{i,s,c,t_0} \tilde{\Delta} I_{s,c,t} + (1 - \sigma_{i,t_0}) \Delta S_{k,t}.$$

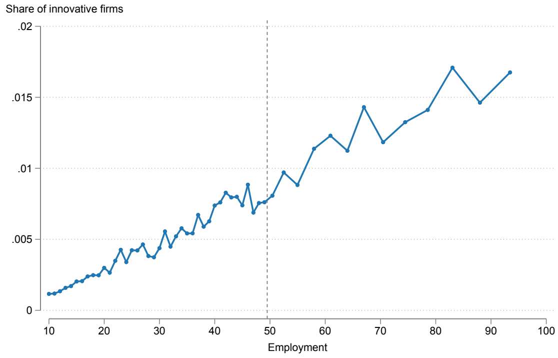
Hence, a non-exporting firms will have a shock equal to the sectoral component while an exporting firms will have a shock equal to a weighted average of the two components. Besides, the larger its export intensity, the closer this new shock is to the baseline one. As shown in column (6) of Table D2, our results are robust to using this shock which allows us to include much more observations (about 932,000 instead of 142,560).

Figure D2: Innovative firms at each employment level - robustness

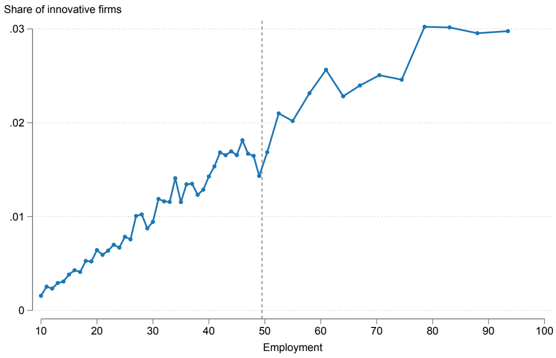
(a) Alternative A



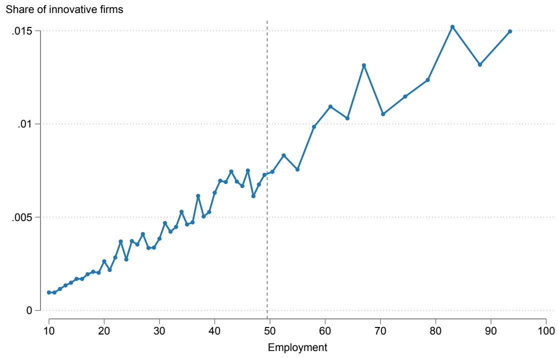
(b) Alternative B



(c) Alternative C



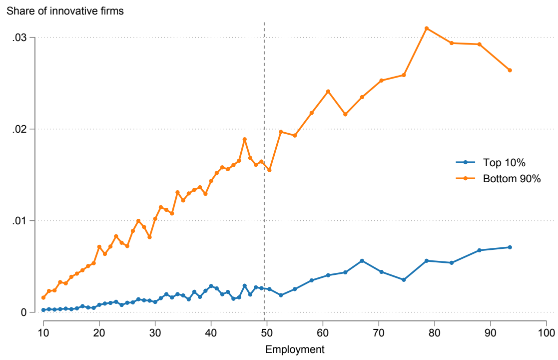
(d) Alternative D



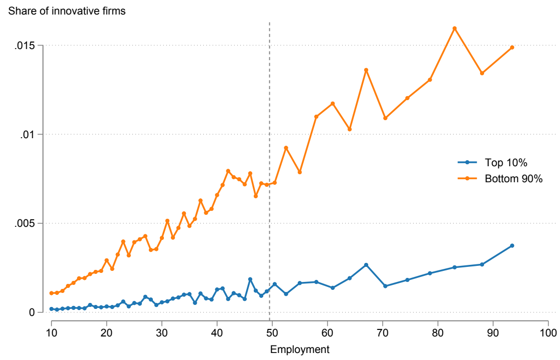
Notes: These Figures replicate Figure 5 using different definitions of the what counts as an innovative firm, based on the timing of patents. Alternatives A, B, C and D define an innovative firm as a firm having filed a priority patent application between $t-2$ and $t+2$ (A), at t (B), between $t-4$ and t (C). Alternative D uses the logarithm of 1 plus the number of patent application at t.

Figure D3: Innovative firms at each employment level and quality of innovation- robustness

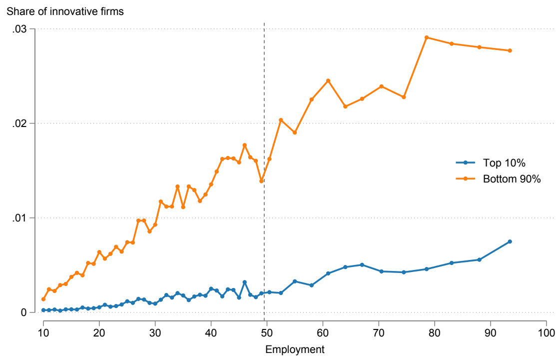
(a) Alternative A



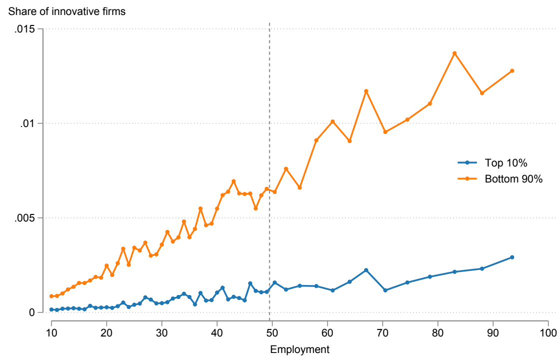
(b) Alternative B



(c) Alternative C

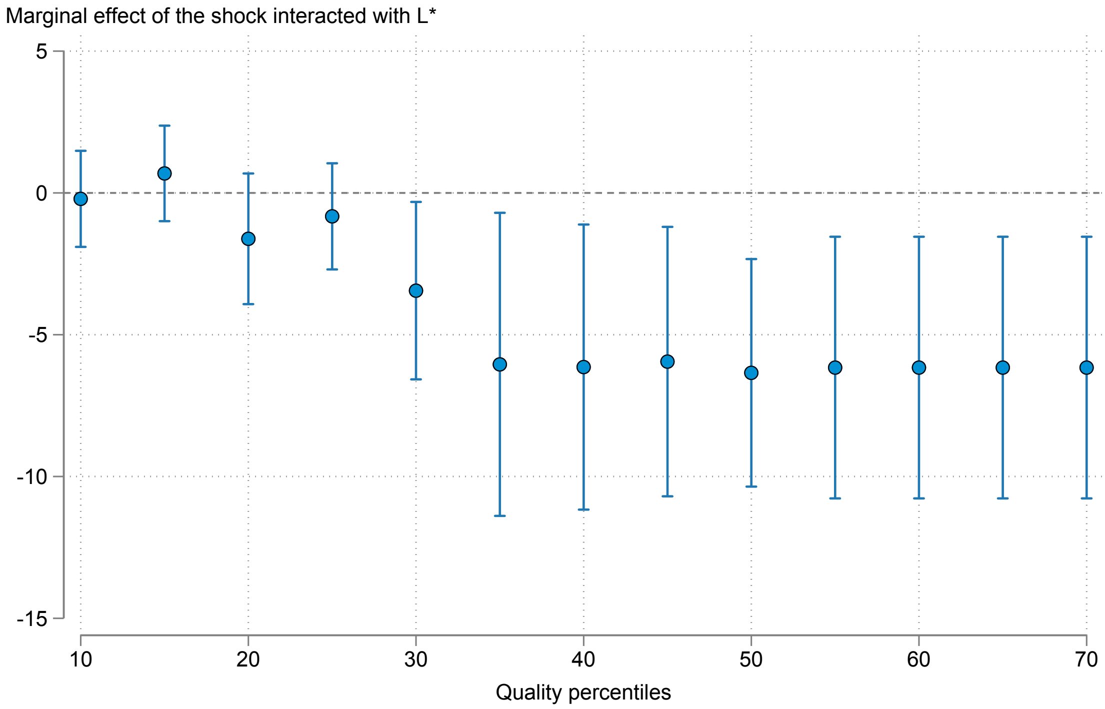


(d) Alternative D



Notes: These Figures replicate 8 using different definitions of the what counts as an innovative firm, based on the timing of patents. Alternatives A, B, C and D define an innovative firm as a firm having filed a priority patent application between $t-2$ and $t+2$ (A), at t (B), between $t-4$ and t (C). Alternative D uses the logarithm of 1 plus the number of patent application at t. The solid line considers the bottom 90% most cited patent and the dashed line the top 10% most cited.

Figure D4: Response to the Demand shock of patents of different quality



Notes: 95% confidence intervals around the estimated coefficient δ in equation (7). Each line corresponds to a separate estimation, where the dependent variable has been redefined by restricting to patents among the x% more cited in the year, with x equal to 10, 15 etc... up to 70. Note that the 65^{th} percentile threshold correspond to 0-citation patent and we include all patents for quality percentiles above 65. The estimated model is the same as in column 5 of Table 2.

Table D1: Placebo tests

	(1) 10-14	**(2)** 15-19	**(3)** 20-24	**(4)** 25-29	**(5)** 30-34	**(6)** 35-39	**(7)** 40-44	**(8)** **45-49**	**(9)** 50-54	**(10)** 55-59	**(11)** 60-64	**(12)** $\geq$ 65
$Shock_{t-2} \times L^{\star}_{t-2}$	-0.123	-0.905	1.300	-4.516**	1.323	3.368	2.356	**-6.159*** **	-5.834	2.638	1.829	0.292
	(1.045)	(1.112)	(2.324)	(2.039)	(2.540)	(3.157)	(2.799)	**(2.178)**	(4.597)	(4.892)	(4.327)	(2.877)
$L^{\star}_{t-2}$	0.015	-0.080	-0.045	0.126	-0.046	-0.063	0.248	**0.078**	-0.249	-0.079	-0.223	0.009
	(0.074)	(0.062)	(0.096)	(0.118)	(0.102)	(0.108)	(0.159)	**(0.132)**	(0.313)	(0.286)	(0.360)	(0.109)
$Shock_{t-2}$	-4.679	-4.424*	-5.027**	-4.173	-4.799*	-4.698*	-4.631*	**-5.268** **	-5.014**	-4.584*	-4.637*	-4.532*
	(2.934)	(2.639)	(2.370)	(2.517)	(2.497)	(2.535)	(2.560)	**(2.512)**	(2.517)	(2.491)	(2.475)	(2.555)
$log(L)_{t-2}$	-0.046	-0.058*	-0.051	-0.047	-0.048	-0.046	-0.057*	**-0.053**	-0.043	-0.046	-0.043	-0.051
	(0.033)	(0.032)	(0.035)	(0.031)	(0.032)	(0.032)	(0.032)	**(0.035)**	(0.031)	(0.037)	(0.034)	(0.037)
$Shock_{t-2} \times log(L)_{t-2}$ L	1.722*	1.671**	1.786**	1.690**	1.721**	1.636*	1.650*	**2.009** **	1.866**	1.670**	1.695**	1.664*
	(0.899)	(0.828)	(0.775)	(0.803)	(0.809)	(0.855)	(0.833)	**(0.819)**	(0.815)	(0.797)	(0.792)	(0.842)
Fixed Effects												
Sector×Year	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Number Obs.	142,560	142,560	142,560	142,560	142,560	142,560	142,560	**142,560**	142,560	142,560	142,560	142,560

Notes: These are based on the specification of column 5 of Table 2. The dependent variable is the Davis and Haltiwanger (1992) growth rate in the number of priority patent applications between $t-1$ and t. In each column $L^{\star}$ has been redefined as a dummy variable set to one if employment at $t-2$ is at different levels. These levels are defined as 10-14 (column 1), 15-19 (column 2), 20-24 (column 3) etc... up to over 65 (the baseline model is therefore in column 8). Innovation is measured by the number of new priority applications. All models include a 2-digit NACE sector interacted with a year fixed effect and a time fixed effect interacted with the initial level of export intensity. Estimation period: 1998-2007 . Standard errors are clustered at the 2-digit NACE sector level. ***, ** and * indicate p-value below 0.01, 0.05 and 0.1 respectively.

Table D2: Robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$Shock_{t-2} \times L^{\star}_{t-2}$	-6.159**	-5.471	-6.159**	-6.159**	-7.144**	-7.144**	-0.447**	-0.048*	-0.044*	-4.283**
	(2.307)	(3.741)	(2.307)	(2.307)	(2.814)	(1.278)	(0.161)	(0.024)	(0.022)	(1.563)
$L^{\star}_{t-2}$	0.078	0.085	0.078	0.078	-0.021	0.039	0.007	0.001	0.002	0.028
	(0.112)	(0.162)	(0.112)	(0.112)	(0.142)	(0.036)	(0.008)	(0.001)	(0.001)	(0.101)
$Shock_{t-2}$	-5.268**	-9.715*	-5.268**	-5.268**	-6.047**	-1.752*	-0.372**	-0.049*	-0.015	-4.169**
	(2.048)	(5.459)	(2.048)	(2.048)	(2.491)	(0.982)	(0.148)	(0.027)	(0.016)	(1.862)
$log(L)_{t-2}$	-0.053	-0.083	-0.053	-0.053	-0.046	-0.013	-0.002	-0.000	-0.000	-0.052
	(0.039)	(0.056)	(0.039)	(0.039)	(0.040)	(0.009)	(0.002)	(0.000)	(0.000)	(0.037)
$Shock_{t-2} \times log(L)_{t-2}$	2.009***	3.323*	2.009***	2.009***	2.281**	2.281**	0.142***	0.018*	0.007	1.548**
	(0.662)	(1.691)	(0.662)	(0.662)	(0.809)	(0.339)	(0.047)	(0.009)	(0.006)	(0.623)
Fixed Effects										
Sector×Year	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Number Obs.	142,560	94,252	142,560	142,560	103,318	932,694	142,560	142,560	142,560	189,864

Notes: These are based on the specification of column (5) of Table 2. The dependent variable is the Davis and Haltiwanger (1992) growth rate in the number of priority patent applications between $t-1$ and t. Each column considers a different sample. Column (1) replicates our baseline specification. Column (2) includes firms that have a workforce between 10 and 100 employees at $t-2$ (instead of the first year they appear in the sample). Column (3) (resp. (4)) includes firms that have a workforce between 10 and 500 (resp. 0 and 100) employees at t_0. Columns (5) and (6) are based on the same size limits as in column (1) but column (5) restricts to firm that first exported in 1994 (i.e.: $t_0 = 1994$, the earliest year in our dataset) and column (6) extends to non-exporting non-manufacturing firms. In Column (6), the shock is defined as a weighted average of a sectoral and the individual shocks, with weights equal to the initial export intensity (see Appendix D.2). Columns (7)-(9) also consider the same sample as column 1 but change the type of growth rate of the dependent variable. Column (7) considers the first difference in $log(1+Y)$, column (8) uses the Inverse Hyperbolic Sine $log(Y+\sqrt{1+Y^2})$, also in first difference and column (9) uses the first difference of Y/P_0, where P_0 is the yearly average number of priority patents filed by the firm before t_0 (the first year the firm appears in the database). Finally, column (10) reproduces the baseline but without trimming the shock. All models include a 2-digit NACE sector interacted with a year fixed effect and a time fixed effect interacted with the initial level of export intensity. Estimation period: 1998-2007 . Standard errors are clustered at the 2-digit NACE sector level. ***, ** and * indicate p-value below 0.01, 0.05 and 0.1 respectively.

D.4 Details (and robustness) of the estimates of the regulatory tax parameter and implications

In this section we give more details regarding the calculation of the aggregate innovation losses and test robustness to our main exercise in Section IV.

D.4.1 Static estimation of τ

Our theoretical model predicts a relationship between Z and employment $l = n/(\gamma\omega)$. Specifically, equation (4) shows that:

$$Z \propto l \text{ if } l < (\bar{n}-1)/(\gamma\omega) \text{ and } Z \propto l(1-\tau)^{\frac{1}{\eta-1}} \text{ if } l \geq \bar{n}/(\gamma\omega)$$

To map this into our data, we need to make an assumption on how Z relates to the number of patents filed by a firm. Our baseline estimates assume that $Z \propto \log P$, where P is the (smoothed) number of patent applications filed by the firm. We can therefore directly estimate τ from the innovation-size slopes for large firms vs. small firms.

In this subsection, we present some robustness tests around the baseline estimates. We report these in Table D3. Column (1) reports the baseline value and corresponding total innovation and welfare loss compared to an economy with $\tau = 0$.[9] Column (2) does the same as column (1) but includes firms with up to 250 employees (instead of 150 in the baseline). It is

[9] To compute these loss, we have kept all other parameters the same. The only other parameters directly affected by changed in the slope estimates is β/ζ. However, we know that this parameter plays little aggregate role.

clear that restricting the upper threshold to 100 employees does not exaggerate the impact of the regulation (if anything, it underestimates it). Column (3) does the same as column (1) but includes an intercept, assuming that $Z = a \log P + b$ for some parameters a and b. Columns (4) and (5) respectively assume that Z is proportional to the number of patents or to the share of firms with at least one patent at this level of employment. Finally, columns (6) and (7) assume that the relationship between Z and P also depends on the sector and year. We thus perform an estimation without binning the data and include an additive sector and year fixed effect (column (6)), and a multiplicative sector-year fixed effect (column (7)).

Although the exact magnitude of the implicit tax varies across the table, it is always non-trivial and our baseline estimate is just below the midpoint of the range of estimates of τ.

Table D3: Alternative estimation of τ

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Observations		Employment binned				Firm level	
τ	2.6%	3.7%	1.3%	1.2%	5.0%	3.8%	4.0%
Total Innovation loss (%)	5.65	8.24	2.87	2.65	11.49	8.48	8.97
Welfare loss (% of C equivalent)	2.22	3.22	1.13	1.04	4.47	3.31	3.50

Notes: This Table presents alternative OLS estimates of parameter τ based on the innovation-employment relationship of equation (4). Columns(1)-(5) bin observations at the employment level (one observation per level of employment) τ is computed as the ratio of two slope, respectively for firms between 10 and 45 employees and for firms between 50 and 100 (except column (2) which extends this to 250). The left-hand side variable is the log of the total number of patents computed as a five year average before t to which we add 1 for columns (1), (2) and (3). Column (4) uses the number of patents in level (as opposed to log) and column (5) the average of a dummy variable equal to 1 if the number of patents in the past five years is non-zero (which is equivalent to the share of firms with at least one patent at a specific level of employment). Columns (6) and (7) use the panel of firm-year (1,737,476 observations) to estimate the coefficient on the 2 year lag of employment on the log of the number of patents at $t+1$. Column (6) includes a 2-digit sector fixed effect and year fixed effects and column (7) includes sector-year fixed effects. Each estimation includes dummies for each employment level between 46 and 49.

D.4.2 Dynamic estimation of τ

We also estimate τ using a dynamic specification as presented in Section IV.B. The idea of the dynamic estimations is to use the response of total innovation Z for a firm of size n which receives a demand shock ε (see Section II.E):

$$Z(n, \varepsilon) = \left(\frac{\beta \tilde{\pi}(n)}{\zeta \eta} \right)^{\frac{1}{\eta - 1}} \omega \gamma l(n) (1 + \varepsilon)^{\frac{1}{\eta - 1}},$$

where $l(n) = n/(\omega\gamma)$ is the level of employment absent a shock and $\tilde{\pi}(n)$ is equal to:

Table D4: Sensitivity analysis for welfare

Robustness	**Loss in total welfare**
Baseline (full sample)	2.22%
1. $\gamma = 1.2$	0.81%
2. $\gamma = 1.50$	7.01%
3. $\eta = 1.3$	0.92%
4. $\eta = 2$	3.36%
5. $\omega = 0.19$	2.23%
6. $\omega = 0.25$	2.22%
7. $\beta/\zeta = 1.44$	1.61%
8. $\beta/\zeta = 1.95$	2.92%
11. τ	
Percentile 25^{th} ($\tau = 0.005$)	0.41%
Percentile 75^{th} ($\tau = 0.046$)	4.07%
Exporting manufacturing firms	
12. Static estimation ($\tau = 0.064$)	5.89%
13. Using dynamic model ($\tau = 0.061$)	5.57%

Notes: baseline uses parameter values: ($\eta = 1.5$, $\gamma = 1.3$, $\tau = 0.0259$, $\beta/\zeta = 1.70$ and $\omega = 0.22$), see Table 3. In the robustness where γ, η, ω or β/ζ are changed, we keep τ as in the baseline. Results in line 11 report the 25^{th} and 75^{th} percentile for the loss of innovation in a sample computed from 100,000 independent draws of τ from two normal distribution. The corresponding value of τ and β/ζ are computed as an average for each percentile. Results in rows 12-13 show the result when restricting to exporting manufacturing firms and the corresponding estimation of τ, either using the static baseline approach or the dynamic model described in Section IV.B. Loss in welfare is given in consumption equivalent and does not include initial quality (see section IV.C).

$$\begin{array}{ll} \pi & \text{if } n < \bar{n} - 1 \\ \pi(1 - \bar{n}\tau) & \text{if } n = \bar{n} - 1 \\ \pi(1 - \tau) & \text{if } n \geq \bar{n}. \end{array}$$

Hence, a shock of magnitude ε implies a change in Z such that:

$$\Delta Z(n, \varepsilon) \equiv Z(n, \varepsilon) - Z(n, 0) = \left(\frac{\beta \tilde{\pi}(n)}{\zeta \eta}\right)^{\frac{1}{\eta-1}} \omega \gamma l(n) \left[(1+\varepsilon)^{\frac{1}{\eta-1}} - 1\right], \tag{C6}$$

which yields:

$$\frac{\partial^2 \Delta Z}{\partial \varepsilon \partial l} = \left(\frac{\beta \tilde{\pi}(n)}{\zeta \eta}\right)^{\frac{1}{\eta-1}} \frac{\omega \gamma}{\eta - 1} (1+\varepsilon)^{\frac{2-\eta}{\eta-1}}.$$

We will now use our empirical exercise to estimate this relationship.

Comparing large and small firms Our main exercise is to consider equation (12) which can be linked to the value of $\frac{\partial^2 \Delta Z}{\partial \varepsilon \partial l}$ as explained in Section IV.B. Coefficients c_5 and c_4 are estimated by OLS using the exact same sample of firms as the one in column (5) of Table 2 (i.e. manufacturing exporting firms). The dependent variable, as in the baseline estimations, is the growth rate of the number of patents filed during the year, using the modified $\tilde{\Delta}$ operator (see Section III.C). Observations with employment between 45 and 49 are removed from the sample (as we did with the static model as behavioral responses are different for these firms). The sample has 131,633 observations. Finally, the model includes sector by year fixed effects.

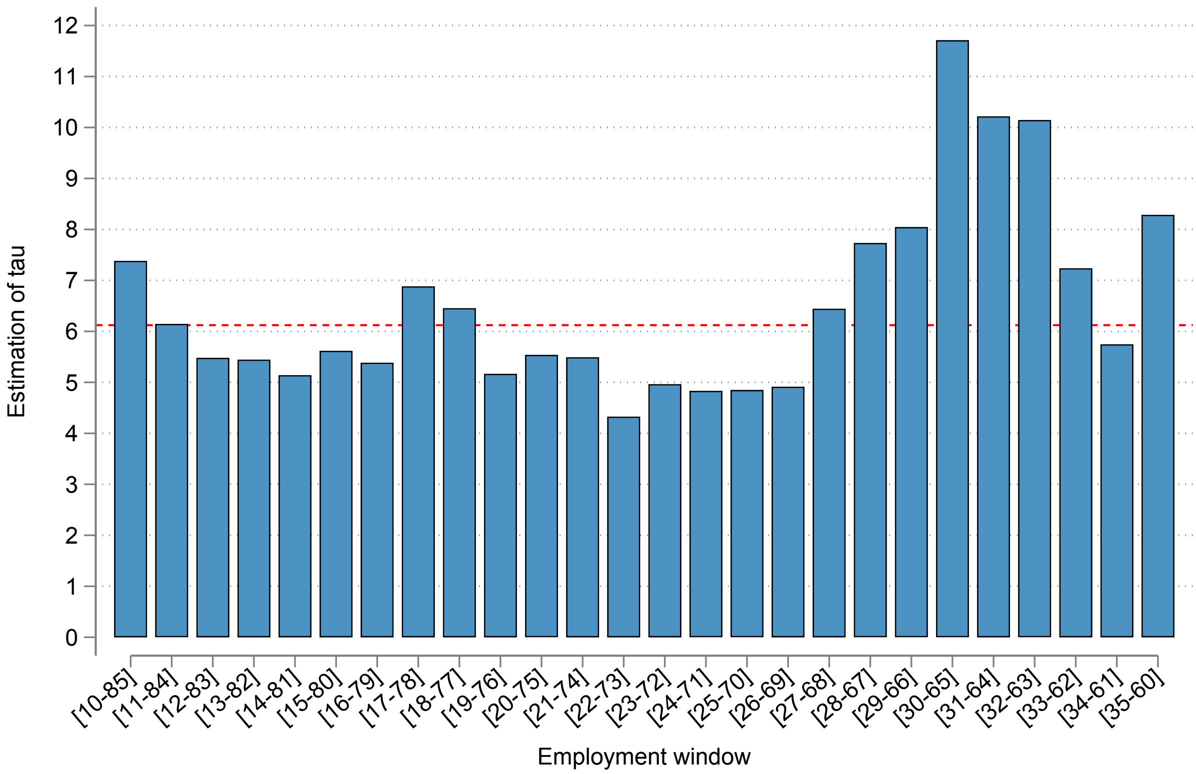
Once the estimated coefficients $\hat{c_5}$ and $\hat{c_4}$ are retrieved, we estimate equation (12) by taking the unweighted average of the shock ϵ and a value of η set to 1.5. As our baseline, we chose to match ϵ to ΔS. As an alternative, we have estimated the link between ϵ and ΔS by looking at the response of employment to a demand shock. In practice this does not impact the value of τ which is estimated to be about 6% if we exclude firms in the innovation valley (between 45 and 49 employees at t-2). Note that nothing substantial changes if we perform the exercises but include firms in the valley. In this case, the estimated value of τ is equal to 5.9%.

Restricting the sample to a narrower size range In the previous analysis, we compared firms with an initial employment ranging between 10 and 100 employees, which is consistent with the approach taken in the rest of the paper. However, we can adopt a coarser selection criterion by examining firms with an employment level between 10 and 100 at $t-2$, and then further reducing the sample by comparing firms that fall within a narrower range of x to $95-x$ employees at $t-2$ to estimate equation (12), where we vary x over the range from 10 to 40. In other words we consider size band windows from as wide as 10-85 employees to as narrow as 35-60 employees.

As we move x from 10 to 35, we include fewer and fewer firms that are closer to the threshold from both sides (e.g. the 35-60 window compares firms of size 35 to 45 with those of size 50 to 60 employees). This procedure restrict the sample and decreases the precision of the estimates, but at the same time allows to compare firms that are more closely related. The resulting values of τ are presented in Figure D5. The values of τ are essentially estimated between 5% and 6% which is in line with what we found in our baseline dynamic estimation. As we increase x above 30, the sample becomes very small (i.e. 35 to 60 employees) which accounts for the greater volatility of the estimates.

Using firms in the innovation valley As a third exercise, we propose an alternative dynamic estimation of τ which also uses the response of firms in the innovation valley to a demand shock, again using our estimates of $\frac{\partial^2 \Delta Z}{\partial \varepsilon \partial l}$. Consider the following model:

Figure D5: Estimating τ from dynamic responses: Robustness to narrowing the size window



Notes: Estimation of τ from the dynamic evaluation presented in Section IV.B and restricting to firms in the employment bracket displayed in the x-axis. Employment taken at at $t-2$ (see equation (12)). The horizontal line corresponds to the baseline 6%.

$$\Delta Z(n,\varepsilon)_{i,t} = c_1 l_{i,t-2} + c_2[\mathbb{1}\left(l_{i,t-2} < V(\bar{l})\right) * l_{i,t-2}] + c_3 \mathbb{1}\left(l_{i,t-2} \in V(\bar{l})\right) \tag{C7}$$
$$+ c_4[\mathbb{1}\left(l_{i,t-2} < V(\bar{l})\right) * l_{i,t-2} * \Delta S_{i,t-2}] + c_5[\mathbb{1}\left(l_{i,t-2} \in V(\bar{l})\right) * l_{i,t-2} * \Delta S_{i,t-2}]$$
$$+ c_6 \Delta S_{i,t-2} + \epsilon_{i,t}.$$

where $V(\bar{l})$ denotes the innovation valley (firms between 45 and 49 employees) and $l_{i,t-2} < V(\bar{l})$ indicates that employment is lower than 45. If we estimate this model for firms below 50 employees only, then the ratio of c_5 over c_4 is equal to:

$$\frac{c_5}{c_4} = \frac{\mathbb{E}\left[(1+\varepsilon)^{\frac{2-\eta}{\eta-1}} | l \in V(\bar{l})\right] (1-\tau\bar{n})^{\frac{1}{\eta-1}}}{\mathbb{E}\left[(1+\varepsilon)^{\frac{2-\eta}{\eta-1}} | l < V(\bar{l})\right]}$$

Hence, with a value of $\bar{n}$ we can estimate τ from these equations using the same strategy as in Section IV.B. We use a value of $\bar{n}$ of 14 which corresponds to the baseline calibrations.

Implementing this method leads us to an estimate of τ equal to 5.9%, very similar to our baseline dynamic approach.

D.5 Measuring different types of innovation

Our baseline approach simply uses patent counts. In the extensions of C.2, we take several approaches to examining the different types of innovation. In order to measure how "radical" a patent is, we use two alternative methods: citations and text-based measures of novelty. Then we also consider measures of how "labor saving" the patent is by looking at measures of automation and process innovation.

Citations. The first method uses the now classical approach of considering future citations. For every patent in a technology class by year of application cell, we calculate all the citations to that patent by all granted patents that were filed in 2016 or earlier. Since the last year we use in our analysis sample is 2007, this gives us a minimum of 10 future years of citation information. We then calculate which quantile of the citations distribution a given patent lies in. A patent which was in the top decile of citations, for example, would be counted as radical for the purposes of column (1) of Table 5.

We validated the use of this measure by presenting employment growth regressions. We regressed the change in the firm's log(employment) on a distributed lag of patent counts with sector by time dummies. Table D5 shows a representative example where we use patents from $t-1$ to $t-3$. To deal with zeros we add one to the patents before taking logs. Column (1) counts only "radical" patents in the top 10% of the technology-class-year cohort citation distribution and column (2) has incremental patents in the bottom 90%. The coefficients of all patents are positive and individually and jointly significant, indicating that patenting is associated with faster firm growth as we would expect. And consistent with our priors, the coefficients on the radical patents are much larger than incremental patents. Summing the coefficients to show the long-run effects in the base of the column we see that the radical patents have about 2.5 (=0.1141/0.0449) times the impact on employment growth compared to incremental patents. The base of the columns shows that in the long-run a doubling of incremental patents increases employment by 3.2% compared to 8.2% for radical patents.

More ambitiously, we can use these estimates to perform a back of the envelope calculation to see how much lower the loss of growth would be if we took into account that the regulation only affects incremental innovation. For example, using the approach of Table D5 radical innovations (the top 10% of the citations distribution) have about 2.5 times the effect of incremental patents, so an innovation index should give a weight of 5/7 to radical innovation and 2/7 to incremental innovation (instead of implicitly equal weights using the patent count). If the overall fall in patenting is 5.7% as estimated in Table 4 and this comes entirely from incremental innovation, we need to scale down the growth effect by 18/23 reflecting the lower impact of incremental

Table D5: Employment and patents for different level of quality

	Top 10%	Bottom 90%
$\log(P_{f,t-1})$	0.0572***	0.0203***
	(0.0096)	(0.0041)
$\log(P_{f,t-2})$	0.0366***	0.0152***
	(0.0126)	(0.0045)
$\log(P_{f,t-3})$	0.0201**	0.0093*
	(0.0095)	(0.0053)
Sum of coefficients	0.1140***	0.0449***
	(0.0079)	(0.0070)
Obs	171,179	171,179
R^2	0.0099	0.0099

Notes: The dependent variable is the change in the firm's log(employment). The right hand side is ln(1+patent count) between $t-1$ and $t-3$. Column (1) restricting to the top 10% most highly cited patents in a technology-class year and column (2) has the other 90% . Both models include a 2-digit NACE sector interacted with year fixed effects. Standard errors are clustered at the firm level. ***, ** and * indicate p-value below 0.01, 0.05 and 0.1 respectively.

innovation. This would imply a fall of 4.4% in growth (compared to the unregulated economy) compared to 5.7% in our baseline estimates. So the extension to different types of innovation does reduce the magnitude of the loss, but not by an enormous amount. Different assumptions will obviously change these exact magnitudes, but are unlikely (in our view) to overturn our main findings.

Google Patent Embedding. Our second, alternative measure of radical innovation involves a text-based analysis of novelty which is more involved and draws on some recent work by Google. In 2019, Google Patent released an embedding representation of each publication available in their public dataset (hereafter, "GP embedding"). As detailed in Srebrovic (2019), embeddings are a set of techniques in natural language processing that map a text to a vector of real numbers. By leveraging methods such as neural networks, this mapping allow to significantly reduce the dimensionality of a text input.

The GP embedding is a vector of 64 dimensions that have been constructed in order to predict a patent's CPC (Cooperative Patent Code) from its text (including all metadata, abstract and body of the patent description). Each element of the vector is a continuous variable ranging between -1 and +1. It therefore summarize the text content of a vector in a simple algebraic representation which has the advantage of allowing to calculate the distance between two patents by taking the dot product between the two corresponding embeddings.

Formally, for each patent p, we let $\mathcal{E}(\mathbf{p})$ denote its embedding representation. We then define the distance between a patent p and a patent q as:

$$d(p,q) = \mathcal{E}(\mathbf{p}).\mathcal{E}(\mathbf{q})$$

Measure of novelty using text. Using this distance measure, we can construct a novelty measure to capture radical innovation. The concept of novelty of a patent captures the extent to which a patent is significantly different from previous innovations in the same field. Typical measures of novelty look at the diversity of technological classes in the set of citing patents, or in the set of cited patents. These measures, sometimes called "originality" are limited by the fact that the average patent does not receive many citations.

Recently, the innovation literature has devoted much attention to using the text content of patent documents to refine some existing measures. For example, Kelly et al. (2018) shows how using the description of the innovation included in a patent publication can be used to build measures of similarity and novelty.

Here, we adapt their methodology. More precisely, we define novelty for each patent as the distance between its embedding and a reference point. This reference point is computed by calculating the unweighted average of all USPTO patents filed in the past 5 years and within the same technological class (we use 3-digit CPC classification, that is around 130 different categories). Formally, we define novelty $NOV(p)$ for each patent p as:

$$NOV(p) = \mathcal{E}(\mathbf{p})\frac{1}{N(k,t)}\sum_{q\in\mathcal{P}(k,t)}\mathcal{E}(\mathbf{q}) = \frac{1}{N(k,t)}\sum_{q\in\mathcal{P}(k,t)} d(p,q),$$

where k is the technological class of patent p,[10] $\mathcal{P}(k,t)$ is the set of USPTO patent filed between $t-5$ and $t-1$ and belonging to technological class k and $N(p,t)$ is its cardinal.

The static cross-section relationship between size and innovation when restricting attention to very novel patents (top 10%) and other patents respectively are shown in Figure 6(a). This graph is analogous to Figure 8. Likewise, results from regressions similar to that performed in Table 5 but using thresholds based on the value of novelty, are shown in Table D6.

Automation. Patents that protect automation technologies have been the subject of a large body of work recently (see e.g. Dechezlepretre et al., 2020; Webb, 2019; Mann and Püttmann, 2018 for reviews). These papers typically use at the description (or abstract) of the patents to

[10] In the case where a patent has more than one CPC code, we consider patents from all the CPC codes in which case k represents the set of technological classes. In other words, we use a weighted average.

identify the occurrence of words that are usually associated with labor-saving technologies.

To build our automation measure, we use the work of Mann and Püttmann (2018) who look at the wording of USPTO patents and build a classifier to distinguish between automation and non-automation technologies. To apply their work to our set of patents, we once again leverage the GP embedding. Specifically, we regress the binary variable from Mann and Püttmann (2018) (1 if patent is classified as automation and 0 otherwise) on each of the 64 coordinates of the patent's embedding. We then use the estimated coefficients to predict the probability of being an automation patent for every patent owned by a French firm. Formally, we define our score of automation $A(p)$ for each patent p as:

$$A(p) = \sum_{i=1}^{64} \hat{\beta}_i \mathcal{E}(p)_i,$$

where $\hat{\beta}_i$ is the estimated coefficients from a model restricted to USPTO patents:

$$Y_q = \sum_{i=1}^{64} \beta_i \mathcal{E}(q)_i + \nu_t + \varepsilon_q,$$

for a patent q published during year t. In this model, Y_q is equal to 1 if the patent has been classified as an automation patent and ε is an error term.

The underlying idea is that a linear combination of the embedding coordinates capture the feature included in the text that predict that a patent protects a labor-saving technology.

We show the cross-section relationship between size and innovation result in Figure 6(b) and the dynamic regression in Table D7.

As an alternative to the automation measure, we also used a measure of the extent to which a patent protects a process innovation (as opposed to a product innovation) using the classification of Arora et al. (2020). This uses the percentages of product or process related words in either the claims or the description of the patent publication document. It is likely the process innovations are more labor saving than product innovations, so the impact of regulation such fall more heavily on the product innovations. As with Mann and Püttmann (2018), this measure is only computed on USPTO patents, we leverage GP again, using the same methodology as the one described above to predict a corresponding value for our set of patents owned by French firms. We obtain broadly similar results. For example, splitting patents at the median level of "process", there are only significant negative effects of the threshold on below median levels of process innovation, i.e. for product innovation.

Table D6: Regression results for different levels of the novelty of innovation

Novelty	Top 10% (1)	Top 15% (2)	Top 25% (3)	Bottom 75% (4)	Bottom 85% (5)	Bottom 90% (6)
$Shock_{t-2} \times L^{\star}_{t-2}$	-1.236	-1.350	-1.796*	-5.111**	-5.599***	-5.644**
	(0.860)	(1.068)	(0.948)	(2.090)	(1.828)	(2.016)
$L^{\star}_{t-2}$	0.061*	0.069	0.095	0.064	0.046	0.062
	(0.032)	(0.056)	(0.104)	(0.116)	(0.100)	(0.105)
$Shock_{t-2}$	-2.852	-3.864*	-4.570**	-2.033	-2.378	-3.518
	(1.725)	(1.940)	(1.818)	(1.999)	(2.180)	(2.279)
$log(L)_{t-2}$	0.000	0.006	-0.027	-0.028	-0.052	-0.053
	(0.015)	(0.021)	(0.026)	(0.033)	(0.034)	(0.034)
$Shock_{t-2} \times log(L)_{t-2}$	0.991*	1.369**	1.585**	0.918	1.035	1.388*
	(0.553)	(0.643)	(0.596)	(0.662)	(0.720)	(0.745)
Fixed Effects						
Sector×Year	✓	✓	✓	✓	✓	✓
Number Obs.	142,560	142,560	142,560	142,560	142,560	142,560

Notes: estimation results of the same model as in column 5 of Table 2. The dependent variable is the Davis and Haltiwanger (1992) growth rate in the number of priority patent applications between $t-1$ and t, restricting to the top 10% most novel (column (1)), the top 15% most novel, the top 25% most novel (column (3)), the bottom 85% most novel (column (4)), the bottom 75% most novel (column (5)) and the bottom 90% most novel (column (6)). Definition of novelty is presented in Section V.A. All models include a 2-digit NACE sector interacted with a year fixed effect and a time fixed effect interacted with the initial level of export intensity. Estimation period is 1998-2007. Standard errors are clustered at the 2-digit NACE sector level. ***, ** and * indicate p-value below 0.01, 0.05 and 0.1 respectively.

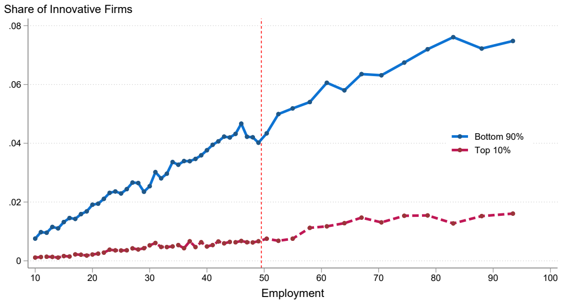
Table D7: Regression results for different levels of the automation

Automation	Top 10% (1)	Top 15% (2)	Top 25% (3)	Bottom 75% (4)	Bottom 85% (5)	Bottom 90% (6)
$Shock_{t-2} \times L^{\star}_{t-2}$	0.688	0.455	0.232	-5.978**	-5.908**	-6.137**
	(0.594)	(0.504)	(0.865)	(2.625)	(2.529)	(2.564)
$L^{\star}_{t-2}$	0.032	0.047	-0.030	0.104	0.043	0.054
	(0.025)	(0.036)	(0.037)	(0.126)	(0.115)	(0.114)
$Shock_{t-2}$	-0.040	-0.310	0.216	-5.096**	-4.830**	-5.425**
	(0.677)	(0.576)	(1.208)	(2.282)	(2.146)	(2.096)
$log(L)_{t-2}$	-0.007	-0.009	-0.015	-0.041	-0.046	-0.050
	(0.006)	(0.009)	(0.012)	(0.037)	(0.037)	(0.037)
$Shock_{t-2} \times log(L)_{t-2}$	-0.026	0.125	0.059	1.806**	1.857**	2.092***
	(0.213)	(0.194)	(0.414)	(0.758)	(0.714)	(0.692)
Fixed Effects						
Sector×Year	✓	✓	✓	✓	✓	✓
Number Obs.	142,560	142,560	142,560	142,560	142,560	142,560

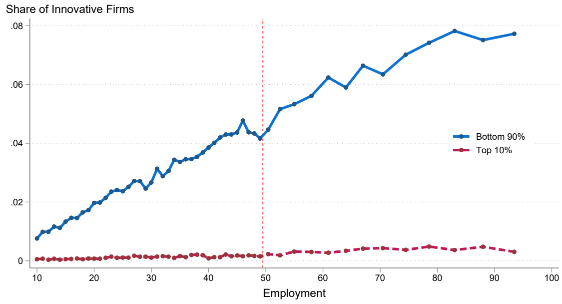
Notes: estimation results of the same model as in column (5) of Table 2. The dependent variable is the Davis and Haltiwanger (1992) growth rate in the number of priority patent applications between $t-1$ and t, restricting to the top 10% patents that score highest in terms of predictive automation measure (column 1) and respectively top 15%, top 25%, bottom 25%, bottom 85% and bottom 90% patents. Definition of automation is presented in Section V.B. All models include a 2-digit NACE sector interacted with year fixed effects. Estimation period: 1997-2007. Standard errors are clustered at the 2-digit NACE sector level. ***, ** and * indicate p-value below 0.01, 0.05 and 0.1 respectively.

Figure D6: Innovative firms at each employment level - novelty and automation

(a) Novelty



(b) Automation



Notes: These Figures replicate Figure 8 but split patents between top 10% and bottom 90% according to their level of novelty (left panel) or their predicted level of automation (right panel).

References

Abowd, John M., and Francis Kramarz. 2003. "The costs of hiring and separations." *Labour Economics*, 15(5): 499–530.

Aghion, Philippe, Antonin Bergeaud, and John Van Reenen. 2021. "The impact of regulation on innovation." National Bureau of Economic Research w28381.

Aghion, Philippe, Antonin Bergeaud, Matthieu Lequien, and Marc Melitz. 2018. "The Impact of Exports on Innovation: Theory and Evidence." National Bureau of Economic Research NBER Working Paper 24600.

Arora, Ashish, Sharon Belenzon, Wes Cohen, and Honggi Lee. 2020. "Big Firms and the Direction of Technical Change." Mimeo Duke University.

Askenazy, Philippe, Thomas Breda, and Vladimir Pecheu. 2022. "Under-Reporting of Firm Size Around Size-Dependent Regulation Thresholds: Evidence from France." AMSE Working Paper 2211.

Axtell, Robert L. 2001. "Zipf distribution of US firm sizes." *science*, 293(5536): 1818–1820.

Davis, Steven J, and John Haltiwanger. 1992. "Gross job creation, gross job destruction, and employment reallocation." *The Quarterly Journal of Economics*, 107(3): 819–863.

Dechezlepretre, Antoine, David Hemous, Morten Olsen, and Carlo Zanella. 2020. "Automating labor: evidence from firm-level patent data." Centre for Economic Performance, LSE CEP Discussion Papers dp1679.pdf.

EPO. 2016. "PATSTAT Spring 2016 Edition." https://www.epo.org/searching-for-patents/business/patstat.html, Accessed: 2023-26-04.

French customs and indirect taxation authorities (DGDDI). 2023. "DOU NC8 - 1993-2014." https://doi.org/10.34724/CASD.668.4645.V1, Accessed: 2023-26-04.

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

Garicano, Luis, Claire Lelarge, and John Van Reenen. 2022. "Firm Size Distortions and the Productivity Distribution: Evidence from France A Discussion and some Extensions." LSE mimeo.

Gaulier, Guillaume, and Soledad Zignago. 2010. "BACI: International Trade Database at the Product-Level. The 1994-2007 Version." CEPII research center Working Papers 2010-23.

Hall, Bronwyn, Adam Jaffe, and Manuel Trajtenberg. 2005. "Market Value and Patent Citations." *RAND Journal of Economics*, 36(1): 16–38.

INSEE. 2016. "NAF REV. 2 et CPF REV. 2 : Guide d'utilisation."

Insee & DGFiP. 1994. "FICUS : Annual structural statistics of companies from the SUSE scheme." https://doi.org/10.34724/CASD.68.1016.V1, Accessed: 2023-26-04.

Insee & DGFiP. 1995. "FICUS : Annual structural statistics of companies from the SUSE scheme." *https://doi.org/10.34724/CASD.68.1015.V1*, Accessed: 2023-26-04.

Insee & DGFiP. 1996. "FICUS : Annual structural statistics of companies from the SUSE scheme." *https://doi.org/10.34724/CASD.68.773.V1*, Accessed: 2023-26-04.

Insee & DGFiP. 1997. "FICUS : Annual structural statistics of companies from the SUSE scheme." *https://doi.org/10.34724/CASD.68.774.V1*, Accessed: 2023-26-04.

Insee & DGFiP. 1998. "FICUS : Annual structural statistics of companies from the SUSE scheme." *https://doi.org/10.34724/CASD.68.775.V1*, Accessed: 2023-26-04.

Insee & DGFiP. 1999. "FICUS : Annual structural statistics of companies from the SUSE scheme." *https://doi.org/10.34724/CASD.68.776.V1*, Accessed: 2023-26-04.

Insee & DGFiP. 2000. "FICUS : Annual structural statistics of companies from the SUSE scheme." *https://doi.org/10.34724/CASD.68.777.V1*, Accessed: 2023-26-04.

Insee & DGFiP. 2001. "FICUS : Annual structural statistics of companies from the SUSE scheme." *https://doi.org/10.34724/CASD.68.778.V1*, Accessed: 2023-26-04.

Insee & DGFiP. 2002. "FICUS : Annual structural statistics of companies from the SUSE scheme." *https://doi.org/10.34724/CASD.68.779.V1*, Accessed: 2023-26-04.

Insee & DGFiP. 2003. "FICUS : Annual structural statistics of companies from the SUSE scheme." *https://doi.org/10.34724/CASD.68.553.V1*, Accessed: 2023-26-04.

Insee & DGFiP. 2004. "FICUS : Annual structural statistics of companies from the SUSE scheme." *https://doi.org/10.34724/CASD.68.554.V1*, Accessed: 2023-26-04.

Insee & DGFiP. 2005. "FICUS : Annual structural statistics of companies from the SUSE scheme." *https://doi.org/10.34724/CASD.68.555.V1*, Accessed: 2023-26-04.

Insee & DGFiP. 2006. "FICUS : Annual structural statistics of companies from the SUSE scheme." *https://doi.org/10.34724/CASD.68.556.V1*, Accessed: 2023-04-26.

Insee & DGFiP. 2007. "FICUS : Annual structural statistics of companies from the SUSE scheme." *https://doi.org/10.34724/CASD.68.557.V1*, Accessed: 2023-26-04.

Insee & DGFiP. 2008. "FARE : Annual structural statistics of companies from the ESANE scheme." *https://doi.org/10.34724/CASD.42.551.V1*, Accessed: 2023-26-04.

Insee & DGFiP. 2009. "FARE : Annual structural statistics of companies from the ESANE scheme." *https://doi.org/10.34724/CASD.42.552.V1*, Accessed: 2023-26-04.

Insee & DGFiP. 2010. "FARE : Annual structural statistics of companies from the ESANE scheme." *https://doi.org/10.34724/CASD.42.124.V1*, Accessed: 2023-26-04.

Insee & DGFiP. 2011. "FARE : Annual structural statistics of companies from the ESANE scheme." *https://doi.org/10.34724/CASD.42.860.V1*, Accessed: 2023-26-04.

Insee & DGFiP. 2012. "FARE : Annual structural statistics of companies from the ESANE scheme." *https://doi.org/10.34724/CASD.42.1026.V1*, Accessed: 2023-26-04.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. 2018. "Measuring technological innovation over the long run." National Bureau of Economic Research w25266.

Kramarz, Francis, and Marie-Laure Michaud. 2010. "The shape of hiring and separation costs in France." *Labour Economics*, 17(1): 27–37.

Lequien, Matthieu, Martin Mugnier, Loriane Py, and Paul Trichelair. 2017. "Linking patents to firms: insights with French firms." mimeo.

Mann, Katja, and Lukas Püttmann. 2018. "Benign effects of automation: New evidence from patent texts." Available at SSRN 2959584.

Mayer, Thierry, Marc J. Melitz, and Gianmarco I. P. Ottaviano. 2014. "Market Size, Competition, and the Product Mix of Exporters." *American Economic Review*, 104(2): 495–536.

Moins, Laurent. 2010. *Code Du Travail 2010.* Lamy.

Srebrovic, Rob. 2019. "Expanding your patent set with ML and BigQuery." Google Cloud Data Analytics https://cloud.google.com/blog/products/data-analytics/expanding-your-patent-set-with-ml-and-bigquery.

Webb, Michael. 2019. "The impact of artificial intelligence on the labor market." Available at SSRN 3482150.