APPENDICES NOT FOR PUBLICATION FOR

Daron Acemoglu, Philippe Aghion, Claire Lelarge,
John Van Reenen, and Fabrizio Zilibotti

"Technology, Information and the Decentralization of the Firm"

Appendix A: Theory

In this appendix, we provide a more detailed outline of the model presented in this theory section, and present proofs of the claims and propositions stated there.

The economy is assumed to be populated by a set $\mathcal{F}$ of firms. Time is discrete. New technologies $k = 1, 2, ...$ become available sequentially and randomly to firms. As a new opportunity of technology adoption materializes, firms must choose the form of its implementation. In each period, one (and only one) new technology becomes available to the firm $i \in \mathcal{F}$ with positive probability $p_i \in (0, 1]$. The speed at which new opportunities become available can differ across firms, and p_i measures the speed at which firm i climbs the technology ladder. The realizations of technological opportunities are independent across firms and over time.

Each technology can be implemented successfully or unsuccessfully. The successful implementation of a technology increases the firm's productivity by a factor $\gamma > 1$, while unsuccessful implementation leaves the productivity of the firm unchanged. So the law of motion of firm i's productivity is given by

$$y_{i,t} = \gamma^{S_i(t)} y_{i,t-1}, \tag{8}$$

where $y_{i,t}$ is the productivity (and revenue) of the firm at time t and $S_i(t)$ is an indicator function taking the value 1 if a technological opportunity arises for firm i at time t and is successfully implemented, and zero otherwise. Whether the technology is successful or not depends on an action taken by the firm, which we denote by $x_{i,k,t} \in \{L, R, \varnothing\}$, with $x_{i,k,t} = \varnothing$ standing for not attempting the technology, and L and R denoting two alternative ways of implementing the new technology, or two "actions". We will see below that the firm will never choose $x_{i,k,t} = \varnothing$, so the relevant choice is between L and R. One of these two actions, denoted by $x^*_{i,k} \in \{L, R\}$, leads to successful implementation, while the other leads to an unsuccessful outcome. As stated in the text, we refer to the action leading to successful implementation as the *correct action*, and assume that it is given by (1) as specified in the text. We also assume that conditional on x^*_k, $x^*_{i,k}$ and $x^*_{i',k}$ for any $i \neq i'$ are independent, or stated alternatively, the sequence $\left\{x^*_{i,k}\right\}_{i\in\mathcal{F}}$ is exchangeable in the sense that any permutation of this sequence is equally likely. This implies that $\left\{x^*_{i,k}\right\}_{i\in\mathcal{F}}$ is a Bernoulli sequence with a parameter of $1-\varepsilon$ or ε (depending on whether $x^*_k = L$ or R). We also assume that

$$\Pr(x^*_k = L) = q_0 = \frac{1}{2}.$$

As discussed in the text, each firm is owned by a principal, who maximizes the present discounted value (PDV) of profits, with discount factor $\beta \in (0, 1)$, conditional on information, h^t, and their initial productivity, $y_{i,t-1} < \infty$. Let us first ignore any costs on the side of firms, so profits are equal to revenue (productivity), and therefore the objective function of firm i is given by

$$V\left(y_{i,t-1}, h^t\right) = \mathbb{E}\left[\sum_{j=0}^{\infty} \beta^j y_{i,t+j} \mid h^t\right]. \tag{9}$$

We assume that $\beta < \gamma^{-1}$ so that the firm's value $V\left(y_{i,t-1}, h^t\right)$ always remains finite.

The above specification makes it clear that it is always advantageous for the firm to implement a technology whenever it becomes available, so we can restrict attention to $x_{i,k,t} \in \{L, R\}$. This is because, by delaying the implementation of technology k available at date t, the firm may obtain more information about this technology, but by assumption, it will never have an opportunity to adopt this technology again. Instead, at $t+1$, technology $k+1$ would become available to firm i with probability p_i. Consequently, a firm's productivity and value will depend on how many technologies have become available in the past to this firm and how successful the firm has been in implementing them. We can also note that none of the results will be affected if we allow firms to implement past unsuccessful new technologies again in the future. In this case, the cost of unsuccessful implementation will be simply the delay in realizing the increase in productivity.

Success in the implementation of technologies depends on the organization of the firm. The two alternative organizational forms available to each firm are: centralization and delegation. In the first, the principal (owner) manages the firm and chooses the action $x_{i,k}$, while in the second, the action choice is delegated to a manager.

The principal has no special skills in identifying the right action, so she bases her decision on the publicly available information h^t. Without loss of any generality, for the decision regarding technology k, we can restrict attention to the history that is payoff relevant to technology k. Moreover, throughout we will focus on a (representative) firm i that has access to technology k at time t. In this case, with a slight abuse of terminology, we will refer to the payoff-relevant history for the implementation of technology k as a *history* h_k^i. As discussed further below, under centralization the principal will choose $x_{i,t,k} = L$ when the posterior that $x_k^* = L$ given history h_k^i,

$$q\left(h_k^i\right) = \Pr\left(x_k^* = L \mid h_k^i\right)$$

is greater than 1/2.

Recall that $\pi\left(d_{i,k}; h_k^i\right)$ is the posterior probability that the firm chooses the correct action conditional on the history h_k^i and the organizational form $d_{i,k}$. We start our analysis by comparing $\pi\left(d_{i,k}; h_k^i\right)$ under the two organizational forms, delegation and centralization (Lemmas 1 and 2). Next we show that profit-maximizing firms always choose the organization form that maximizes $\pi\left(d_{i,k}; h_k^i\right)$ (Lemma 3). Finally, we turn to the main testable implications of the theory, linking distance to frontier (n_k^i) and heterogeneity (ε) to the probability that firms choose either delegation or centralization (Propositions 1 and 2).

As noted in the text, when the principal delegates the implementation of a new technology to an informed manager, the probability of success is constant and equal to δ, i.e.,

$$\pi\left(1; h_k^i\right) = \delta. \tag{10}$$

If, on the other hand, the principal retains authority, the probability of success is a stochastic variable that depends on h_k^i, thus both on the firm's distance to the frontier, n_k^i, and on the experiences of firms further ahead, captured by the number of firms $\tilde{n}_k^i$ out of n_k^i for whom L was the right action. The following lemma provides the expression for $\pi\left(0; h_k^i\right)$.

Lemma 1 Given a history h_k^i, the probability of success for a principal who retains authority is

$$\pi\left(0; h_k^i\right) = \begin{cases} \varepsilon + q\left(h_k^i\right)(1-2\varepsilon) & \text{if } q\left(h_k^i\right) \geq 1/2 \\ \varepsilon + \left(1 - q\left(h_k^i\right)\right)(1-2\varepsilon) & \text{if } q\left(h_k^i\right) < 1/2. \end{cases} \tag{11}$$

where

$$q\left(h_k^i\right) = \frac{(1-\varepsilon)^{\tilde{n}_k^i} \varepsilon^{n_k^i - \tilde{n}_k^i}}{(1-\varepsilon)^{\tilde{n}_k^i} \varepsilon^{n_k^i - \tilde{n}_k^i} + (1-\varepsilon)^{n_k^i - \tilde{n}_k^i} \varepsilon^{\tilde{n}_k^i}} = \frac{1}{1 + \left(\frac{1-\varepsilon}{\varepsilon}\right)^{n_k^i - 2\tilde{n}_k^i}} \tag{12}$$

is the posterior probability that the right choice is L.

Proof. The probability of success for a firm choosing centralization, conditional on action $x_{i,k}$ is:

$$\pi\left(0;h_k^i \mid x_{i,k}=L\right) = q\left(h_k^i\right)(1-\varepsilon)+(1-q\left(h_k^i\right))\varepsilon, \tag{13}$$
$$\pi\left(0;h_k^i \mid x_{i,k}=R\right) = (1-q\left(h_k^i\right))(1-\varepsilon)+q\left(h_k^i\right)\varepsilon, \tag{14}$$

where $q\left(h_k^i\right)$ is given by (12). In equation (13), $q\left(h_k^i\right)(1-\varepsilon)$ is the posterior probability that the reference action is L and that the correct action for the firm coincides with the reference action, whereas $(1-q\left(h_k^i\right))\varepsilon$ is the posterior probability that the reference action is R and that the correct action for the firm differs from the reference action. Equation (14) has a similar form, with $(1-q\left(h_k^i\right))(1-\varepsilon)$ as the posterior probability that the reference action is R and that the correct action for the firm coincides with the reference action and $q\left(h_k^i\right)\varepsilon$ as the posterior probability that the reference action is R and that the correct action for the firm differs from the reference action.

Thus

$$\begin{aligned}\pi\left(0;h_k^i\right) &= \max\left\langle \pi\left(0,h_k^i \mid x_{i,k}=L\right), \pi\left(0,h_k^i \mid x_{i,k}=R\right)\right\rangle \\ &= \begin{cases} \varepsilon+q\left(h_k^i\right)(1-2\varepsilon) & \text{if } q\left(h_k^i\right)\geq 1/2 \\ \varepsilon+\left(1-q\left(h_k^i\right)\right)(1-2\varepsilon) & \text{if } q\left(h_k^i\right)<1/2. \end{cases},\end{aligned}$$

establishing the result. ■

To understand the expression of $\pi\left(0;h_k^i\right)$, note that the principal chooses L if $q\left(h_k^i\right)\geq 1/2$ and R if $q\left(h_k^i\right)\leq 1/2$. The first row of (11) gives then the probability that, when $q\left(h_k^i\right)\geq 1/2$, the firm correctly infers that the reference action is L *and* the reference action coincides with the correct action. The second row gives the probability that, when $q\left(h_k^i\right)<1/2$, the firm incorrectly infers that the reference action is R *and* the reference action does not coincide with the correct action (hence, due to a double mistake, the firms makes the successful adoption).

The expression of $q\left(h_k^i\right)$ in (12) follows from the Bayes' rule, given the Bernoulli assumption and $q_0=1/2$. It may be useful to note that since $\varepsilon\in(0,1/2)$, $q\left(h_k^i\right)>1/2$ whenever $\tilde{n}_k^i>n_k^i/2$, and $q\left(h_k^i\right)<1/2$ whenever $\tilde{n}_k^i<n_k^i/2$. More importantly, $q\left(h_k^i\right)$ is a random variable that depends on the realization of the stochastic vector h_k^i. Since h_k^i consists of a deterministic component, n_k^i, and a stochastic one, $\tilde{n}_k^i$, we can determine the first moment of $q\left(h_k^i\right)$ conditional on the sample size n_k^i, $E\left(q_k^i|n_k^i\right)$, where expectation is taken over possible realizations of $\tilde{n}_k^i$.

The following lemma establishes how this conditional expectation changes with n_k^i and the limiting behavior of $q\left(h_k^i\right)$. In the rest of the analysis, without loss of any generality, we assume that $x_k^*=L$.

Lemma 2 Let $E\left(q_k^i|n_k^i\right)$ denote the expectation of $q\left(h_k^i\right)$ conditional on sample size n_k^i and suppose that $x_k^*=L$. Then $E\left(q_k^i|n_k^i\right)$ is an increasing function of n_k^i. Moreover, $p\lim_{n_k^i\longrightarrow\infty} q\left(h_k^i=\left\{n_k^i,\tilde{n}_k^i\right\}\right)=1$.

Proof. (1) To see that $E\left(q_k^i|n_k^i\right)$ is increasing in n_k^i, let $(q_n)_k^i$ denote the posterior probability of firm i that $x_k^*=L$ conditional on a history of length n, i.e., $\left(q_{n_k^i}\right)_k^i\equiv E\left(q_k^i|n_k^i\right)$. We need to prove that $(q_n)_k^i$ is increasing in n when $x_k^*=L$. First note that $(q_n)_k^i$ is a sufficient statistic for the history h_k^i. Thus, by the Law of Iterated Expectations, we have

$$(q_{n+1})_k^i = E\left((q_{n+1})_k^i \mid (q_n)_k^i\right) > (q_n)_k^i\,.$$

Therefore, it is sufficient to prove that

$$E\left((q_{n+1})_k^i \mid (q_n)_k^i\right) > (q_n)_k^i\,.$$

Consider a firm with a posterior $(q_n)_k^i$ conditional on a sample of n and suppose that it obtains an additional observation, i.e., observes the realization of a $n+1^{st}$ firm. Since, given $x_k^* = L$, L and R will be revealed to be the right choice for the $n+1^{st}$ firm with respective probabilities $1-\varepsilon$ and ε, Bayes' rule implies that

$$(q_{n+1})_k^i = \begin{cases} \frac{(q_n)_k^i \times (1-\varepsilon)}{(q_n)_k^i \times (1-\varepsilon) + \left(1-(q_n)_k^i\right)\times\varepsilon} & \text{with prob. } 1-\varepsilon \\ \\ \frac{(q_n)_k^i \times \varepsilon}{\left(1-(q_n)_k^i\right)\times(1-\varepsilon) + (q_n)_k^i \times\varepsilon} & \text{with prob. } \varepsilon \end{cases}$$

Therefore,

$$\begin{aligned} E\left((q_{n+1})_k^i \mid (q_n)_k^i\right) &= (1-\varepsilon)\times \frac{(q_n)_k^i \times (1-\varepsilon)}{(q_n)_k^i \times (1-\varepsilon) + \left(1-(q_n)_k^i\right)\times\varepsilon} \\ &\quad +\varepsilon \times \frac{(q_n)_k^i \times \varepsilon}{\left(1-(q_n)_k^i\right)\times(1-\varepsilon) + (q_n)_k^i \times\varepsilon}. \end{aligned}$$

This implies that

$$\begin{aligned} E\left((q_{n+1})_k^i \mid (q_n)_k^i\right) & - (q_n)_k^i \\ &= \frac{\left(1-(q_n)_k^i\right)^2 (1-2\varepsilon)^2 \times (q_n)_k^i}{\left(\varepsilon\times(q_n)_k^i + \left(1-(q_n)_k^i\right)\times(1-\varepsilon)\right)\left((q_n)_k^i\times(1-\varepsilon)+\varepsilon\times\left(1-(q_n)_k^i\right)\right)} \\ &> 0, \end{aligned}$$

establishing the desired result.

(2) Next, to prove $p\lim_{n_k^i \longrightarrow\infty} q_{i,k}\left(h_k^i\right) = 1$ note that from the viewpoint of $t=0$, $q\left(h_k^i\right)$ is a random variable, since the history h_k^i is a random vector. We need to show that for almost all histories $q\left(h_k^i\right)$ will become arbitrarily close to 1. We will do this by using the Continuous Mapping Theorem (e.g., van der Vaart, 1998, Theorem 2.3). First, when $x_k^* = L$, by the strong law of large numbers $p\lim_{n_k^i\to\infty} \tilde{n}_k^i / n_k^i = 1-\varepsilon$. The Continuous Mapping Theorem then implies that any continuous function, $G\left(\tilde{n}_k^i, n_k^i\right)$, converges in probability to $\lim_{n_k^i\to\infty} G\left((1-\varepsilon)\,n_k^i, n_k^i\right)$. (12) is such a function, so

$$p \lim_{n_k^i \longrightarrow\infty} q_{i,k}\left(h_k^i\right) = \lim_{n_k^i \longrightarrow\infty} \frac{1}{1+\left(\frac{\varepsilon}{1-\varepsilon}\right)^{n_k^i[2(1-\varepsilon)-1]}} = 1,$$

where the last equality follows from $\varepsilon < 1/2$. ■

Lemma 2 establishes the intuitive result that, as the history relevant to technology k expands, the principal learns the reference action x_k^* with increasing precision.

Next, we establish that profit-maximizing firms always choose the organization form which maximizes $\pi\left(d_{i,k}; h_k^i\right)$. Although firms have a dynamic objective function, given by (9), the maximization program is equivalent to a sequence of static problems. Intuitively, the current organization choice only affects the PDV of the firm via its effect on current productivity, $y_{i,t}$, so the optimal strategy simply maximizes the probability of successful implementation of new technologies in each period.

Lemma 3 A firm i maximizing (9) chooses, for all technologies k,

$$d_{i,k}^*\left(h_k^i\right) \in \{0,1\} = \arg\max_{d_{i,k}} \pi\left(d_{i,k}; h_k^i\right).$$

In particular, given history h_k^i, firm i will choose $d_{i,k}^*\left(h_k^i\right) = 1$ (delegation) if $\pi\left(0; h_k^i\right) < \delta$, and $d_{i,k}^*\left(h_k^i\right) = 0$ (centralization) if $\pi\left(0; h_k^i\right) > \delta$, where $\pi\left(0; h_k^i\right)$ is given by (11).

Proof. First, we note that the value of the firm admits a recursive representation, so that for $s^* \in \mathcal{S}$ that maximizes (9) with starting productivity y_i and corresponding history h_k^i, let this maximum value be $V_i\left(y_i, h_k^i\right)$. Then

$$V_i\left(y_i, h_k^i\right) = \max_{d_{i,k}\in\{0,1\}} \mathbb{E}\left[V_i\left(y_i, h_k^i \mid d_{i,k}\right) \mid h_k^i\right], \text{ and} \tag{15}$$

$$\begin{aligned} V_i\left(y_i, h_k^i \mid d_{i,k}\right) &= p_i\pi\left(d_{i,k}; h_k^i\right)\left(\gamma y_i + \beta\mathbb{E}\left[V_i\left(\gamma y_i, h_{k+1}^{i+1}\right) \mid h_k^i\right]\right) \\ &\quad + \left(1 - p_i\pi\left(d_{i,k}; h_k^i\right)\right)\left(y_i + \beta\mathbb{E}\left[V_i\left(y_i, h_{k+1}^i\right) \mid h_k^i\right]\right). \end{aligned}$$

In particular, with probability p_i, the opportunity to implement the next technology arrives and it is successfully implemented with probability $\pi\left(d_{i,k}; h_k^i\right)$. If it is successfully implemented, y_i increases to γy_i, and otherwise it stays at y_i. Future probabilities of opportunities and success are unaffected by these choices or realizations. The resulting optimal policy $d_{i,k}^*$ in this recursive representation is the organizational form induced by s^*.

To characterize the form of the value function in (15) and the optimal policy $d_{i,k}^*$, we guess and verify that $V_i\left(y_i, h_k^i \mid d_{i,k}\right)$ takes a linear form, namely

$$V_i\left(y_i, h_k^i \mid d_{i,k}\right) = A_i\left(h_k^i \mid d_{i,k}\right) y_i.$$

We then substitute the guess into the value function so as to solve for the unknown coefficient $A_i\left(h_k^i \mid d_{i,k}\right)$. This yields the following recursive equation for A_i:

$$\begin{aligned} A_i\left(h_k^i \mid d_{i,k}\right) &= p_i\pi\left(d_{i,k}; h_k^i\right)\gamma + \left(1 - p_i\pi\left(d_{i,k}; h_k^i\right)\right) + \\ &\quad \beta\left(p_i\pi\left(d_{i,k}; h_k^i\right)\gamma + \left(1 - p_i\pi\left(d_{i,k}; h_k^i\right)\right)\right) \max_{d_{i,k+1}\in\{0,1\}} \mathbb{E}\left[A_i\left(h_{k+1}^i \mid d_{i,k+1}\right) \mid h_k^i\right]. \end{aligned} \tag{16}$$

The optimal organizational form is simply that which maximizes $A_i\left(h_k^i \mid d_{i,k}\right)$. This is equivalent, in turn, to choose organization so as to maximize $\pi\left(d_{i,k}; h_k^i\right)$ (note that the term $\mathbb{E}\left[A\left(h_{k+1}^i \mid d_{i,k+1}\right) \mid h_k^i\right]$ is outside of the firm's control). Namely, the firm will choose organizational form $d_{i,k} = 1$ if $\pi\left(1; h_k^i\right) > \pi\left(0; h_k^i\right)$, and $d_{i,k} = 0$ otherwise. Since the program is recursive, we have established that choosing organization period-by-period so as to maximize $\pi\left(d_{i,k}; h_k^i\right)$ maximizes the value of the firm.

The second part of the Lemma follows immediately from the previous analysis. ■

Lemmas 1, 2 and 3 enable us to characterize how the organizational form changes with distance to frontier, n_k^i, and heterogeneity, ε. Consider, first, distance to frontier. Suppose, in particular, that a firm is at the technology frontier, so that it is the first firm implementing technology k. In this case, $n_k^i = 0$, and so $q\left(h_k^i\right) = 1/2$ and $\pi\left(0; h_k^i\right) = q_0 = 1/2 < \delta$. Thus, when no public information is available, the firm will choose delegation, $d_{i,k}^*\left(h_k^i\right) = 1$. Next, consider the other extreme where the firm is far behind the technology frontier, so that many other firms have implemented the same technology before. In this case, we have $n_k^i \to \infty$ and $p\lim_{n_k^i \longrightarrow\infty} q\left(h_k^i\right) = 1$ from Lemma 2. Equation (11) then implies $p\lim_{n_k^i \longrightarrow\infty} \pi\left(0, h_k^i\right) = 1 - \varepsilon > \delta$, where the inequality follows from the fact that $\delta \in (1/2, 1-\varepsilon)$. Hence, the firm will (almost surely) choose centralization, $d_{i,k}^*\left(h_k^i\right) = 0$. This discussion establishes the following proposition:

Proposition 1 (Distance to Frontier) Consider the adoption decision of technology k by firm i, and suppose that Assumption 1 holds and $x_k^* = L$. Then:
(i) For a firm at the frontier, i.e., $n_k^i = 0$, the principal chooses delegation, $d_{i,k}^*\left(h_k^i\right) = 1$.
(ii) For a firm sufficiently far from the frontier, i.e., $n_k^i \to \infty$, the principal chooses almost surely centralization, i.e., $p\lim_{n_k^i \longrightarrow\infty} d_{i,k}^*\left(h_k^i = \left\{n_k^i, \tilde{n}_k^i\right\}\right) = 0$.

From this proposition and the fact that, $E\left(q_k^i | n_k^i\right)$ is increasing in n_k^i (Lemma 2), one might expect a more general result, such that as distance to frontier n_k^i increases, decentralization becomes more likely. Unfortunately, though intuitive, this result is not correct because of integer issues, though it is true when integer issues are ignored, for example by smoothing the relationship between $E\left(q_k^i|n_k^i\right)$ and n_k^i. Therefore, in the empirical analysis, we disregard the integer problem and focus on the prediction that centralization increases with the distance to the frontier. We will proxy distance to the frontier with the gap between the productivity of a firm and the highest productivity (or more precisely the highest percentile productivity) in the same industry. It is clear that firms further from the frontier in terms of having high n_k^i's are less productive, since these are the firms that have been unlucky and have had fewer opportunities to adopt technologies, and they are also likely to be the ones with relatively low p_i's, that is, those that are slower in climbing their technology ladder.

We next turn to heterogeneity. Let us define $\Pr\left(d_{i,k}^*\left(h_t^i\right)=1\right)$ as the ex ante probability that firm i facing technology k at time t will choose decentralization. This probability is clearly a function of the parameters of the model, in particular ε, which measures the extent of heterogeneity, and the firm's distance to frontier. In particular, recall that greater ε translates into a greater heterogeneity in the firm's environment. The following proposition establishes that greater heterogeneity—higher ε—encourages decentralization (proof in Appendix A):

Proposition 2 (Heterogeneity) Consider the adoption decision of technology k by firm i. Given the distance to frontier, we have

$$\frac{\partial \Pr\left(d_{i,k}^*\left(h_t^i\right)=1\right)}{\partial\varepsilon}\geq 0,$$

so that an increase in ε makes delegation more likely.

This result is proved using the following three lemmas (which are themselves proved below).

Lemma 4 *For all $n_k^i\in\mathbb{N}$ and $\tilde{n}_k^i\in\mathbb{N}$ with $\tilde{n}_k^i\leq n_k^i$, we have*

$$\pi\left(0;\left(n_k^i,\tilde{n}_k^i\right)\right)=\pi\left(0;\left(n_k^i,n_k^i-\tilde{n}_k^i\right)\right).$$

This lemma states that firms updates their beliefs *symmetrically* after signals suggesting either L or R to be the more likely correct action.

Lemma 5 *Either $\pi\left(0;\left(n_k^i,\tilde{n}_k^i\right)\right)<\delta$ for all $\tilde{n}_k^i\in[0,n_k^i]$, or there exists a unique integer, $Q\left(\varepsilon,n_k^i\right)\in\mathbb{Z}_+$, such that*

$$\pi\left(0;\left(n_k^i,\tilde{n}_k^i\right)\right)\geq\delta\Leftrightarrow\begin{cases}\text{either} & \tilde{n}_k^i\leq Q\left(\varepsilon,n_k^i\right),\\ \text{or} & \tilde{n}_k^i\geq n_k^i-Q\left(\varepsilon,n_k^i\right).\end{cases}$$

In the latter case, $Q\left(\varepsilon,n_k^i\right)\leq\left(n_k^i-1\right)/2$ and $Q\left(\varepsilon,n_k^i\right)$ is non-increasing in ε.

This lemma states that the posterior that the firm will choose the correct action will be greater than the threshold for decentralization, δ, if the number of successful L actions in the past are either smaller or greater than a specific threshold depending on the integer $Q\left(\varepsilon,n_k^i\right)$. Finally, we have the following technical result:

Lemma 6 *Let*

$$\chi\left(n_k^i,\tilde{n}_k^i,\varepsilon,\bar{Q}\right)\equiv\Pr\left[\tilde{n}_k^i\leq\bar{Q}\right]+\Pr\left[\tilde{n}_k^i\geq n_k^i-\bar{Q}\right]\tag{17}$$

Then,

$$\frac{\partial\chi\left(n_k^i,\tilde{n}_k^i,\varepsilon,\bar{Q}\right)}{\partial\bar{Q}}\geq 0,$$

and for any $\bar{Q}\leq\left(n_k^i-1\right)/2$,

$$\frac{\partial\chi\left(n_k^i,\tilde{n}_k^i,\varepsilon,\bar{Q}\right)}{\partial\varepsilon}\leq 0.$$

Lemmas 4, 5 and 6 prove Proposition 2. In particular, unless $\pi\left(0;\left(n_k^i,\tilde{n}_k^i\right)\right)<\delta$ for all $\tilde{n}_k^i\in[0,n_k^i]$, we have that

$$\Pr\left(d_{i,k}^*\left(h_t^i\right)=1\right)=\chi\left(n_k^i,\tilde{n}_k^i,\varepsilon,Q\left(\varepsilon,n_k^i\right)\right).$$

By definition,

$$\frac{d\Pr\left(d_{i,k}^*\left(h_t^i\right)=1\right)}{d\varepsilon}=\left.\frac{\partial\chi\left(n_k^i,\tilde{n}_k^i,\varepsilon,Q\left(\varepsilon,n_k^i\right)\right)}{\partial\varepsilon}\right|_{Q\left(\varepsilon,n_k^i\right)=\bar{Q}}+\frac{\partial\chi\left(n_k^i,\tilde{n}_k^i,\varepsilon,Q\left(\varepsilon,n_k^i\right)\right)}{\partial Q\left(\varepsilon,n_k^i\right)}\frac{dQ\left(\varepsilon,n_k^i\right)}{d\varepsilon}.$$

Lemma 6 implies that the first term is non-positive and that $\partial\chi\left(n_k^i,\tilde{n}_k^i,\varepsilon,\bar{Q}\right)/\partial\bar{Q}\geq0$, while Lemma 5 establishes that $dQ\left(\varepsilon,n_k^i\right)/d\varepsilon\leq0$. Finally, if $\pi\left(0;\left(n_k^i,\tilde{n}_k^i\right)\right)<\delta$ for all $\tilde{n}_k^i\in[0,n_k^i]$, a change in ε has no impact on $\Pr\left(d_{i,k}^*\left(h_t^i\right)=1\right)$, which is equal to 0. This establishes the proposition.

Proof of Lemma 4: The equality follows from the assumption that $q_0=1/2$. More formally, equations (11) and (12) imply that for all $\tilde{n}_k^i\leq\left(n_k^i-1\right)/2$,

$$\begin{aligned}\pi\left(0;\left(n_k^i,\tilde{n}_k^i\right)\right)&=\varepsilon+\left(1-\frac{1}{1+\left(\frac{1-\varepsilon}{\varepsilon}\right)^{n_k^i-2\tilde{n}_k^i}}\right)(1-2\varepsilon)\\&=\varepsilon+\frac{1}{1+\left(\frac{1-\varepsilon}{\varepsilon}\right)^{-\left(n_k^i-2\tilde{n}_k^i\right)}}(1-2\varepsilon)=\pi\left(0;\left(n_k^i,n_k^i-\tilde{n}_k^i\right)\right).\end{aligned}$$

The same conclusion follows from applying (11) to the case in which $\tilde{n}_k^i>\left(n_k^i-1\right)/2$.■

Proof of Lemma 5: Suppose that there exists $\tilde{n}_k^i\in[0,n_k^i]$ such that $\pi\left(0;\left(n_k^i,\tilde{n}_k^i\right)\right)\geq\delta$. Let $X\left(\varepsilon,n_k^i\right)\in\mathbb{R}_+$ be the unique value of X that solves the following equation:

$$\tilde{\pi}\left(0;\left(n_k^i,X\right)\right)\equiv\varepsilon+\left(1-\frac{1}{1+\left(\frac{1-\varepsilon}{\varepsilon}\right)^{n_k^i-2X}}\right)(1-2\varepsilon)=\delta.$$

This equation has a unique solution since, by hypothesis, $\pi\left(0;\left(n_k^i,\tilde{n}_k^i\right)\right)\geq\delta$ for some $\tilde{n}_k^i\in[0,n_k^i]$; moreover, $\tilde{\pi}\left(0;\left(n_k^i,n_k^i/2\right)\right)=1-\varepsilon<\delta$, and the left-hand side is continuous and monotonically decreasing in both ε and X in the range where $X\in\left(0,n_k^i/2\right)$. [To see that $\tilde{\pi}\left(0;\left(n_k^i,X\right)\right)$ is decreasing in ε, let $\zeta\left(n_k^i,X\right)\equiv\left[1+\left(\frac{1-\varepsilon}{\varepsilon}\right)^{n_k^i-2X}\right]^{-1}$. In the range $X\in\left(0,n_k^i/2\right)$, we have $\frac{\partial}{\partial\varepsilon}\tilde{\pi}\left(0;\left(n_k^i,X\right)\right)=1-2\left(1-\zeta\left(n_k^i,X\right)\right)-(1-2\varepsilon)\frac{\partial\zeta\left(n_k^i,X\right)}{\partial\varepsilon}$. Since $\zeta\left(n_k^i,X\right)<1/2$ and $\partial\zeta\left(n_k^i,X\right)/\partial\varepsilon>0$, $\partial\tilde{\pi}\left(0;\left(n_k^i,X\right)\right)/\partial\varepsilon<0$ follows]. Therefore, there exists a unique $X\left(\varepsilon,n_k^i\right)$ such that $\tilde{\pi}\left(0;\left(n_k^i,X\left(\varepsilon,n_k^i\right)\right)\right)=\delta$. Lemma 4 implies that $\tilde{\pi}\left(0;\left(n_k^i,X\left(\varepsilon,n_k^i\right)\right)\right)=\delta$ if and only if $\tilde{\pi}\left(0;\left(n_k^i,n_k^i-X\left(\varepsilon,n_k^i\right)\right)\right)=\delta$. Let $Q\left(\varepsilon,n_k^i\right)\in\mathbb{Z}_+$ be the largest integer smaller than X, which will be the threshold number of realization of either L or R such that $\pi\left(0;\left(n_k^i,\tilde{n}_k^i\right)\right)>\delta$. More formally, $Q\left(\varepsilon,n_k^i\right)\equiv\max_{z\in\mathbb{Z}_+}\left\{z\leq X\left(\varepsilon,n_k^i\right)\right\}$. The properties of $X\left(\varepsilon,n_k^i\right)$ immediately imply that $Q\left(\varepsilon,n_k^i\right)$ is no larger than $\left(n_k^i-1\right)/2$ and is non-increasing in ε.■

Proof of Lemma 6: Since $\tilde{n}_k^i$ is the number of successes out of n_k^i in a Bernoulli trial, then holding

Q constant at $\bar{Q}$, we have that

$$\begin{aligned}
\Pr\left[\tilde{n}_k^i \leq \bar{Q}\right] &= \sum_{\tilde{n}_k^i=0}^{\bar{Q}} \begin{pmatrix} n_k^i \\ \tilde{n}_k^i \end{pmatrix} (1-\varepsilon)^{\tilde{n}_k^i} \varepsilon^{n_k^i-\tilde{n}_k^i}, \text{ and} \\
\Pr\left[\tilde{n}_k^i \geq 1-\bar{Q}\right] &= \sum_{\tilde{n}_k^i=n_k^i-\bar{Q}}^{n_k^i} \begin{pmatrix} n_k^i \\ \tilde{n}_k^i \end{pmatrix} (1-\varepsilon)^{\tilde{n}_k^i} \varepsilon^{n_k^i-\tilde{n}_k^i}, \\
&= \sum_{\tilde{n}_k^i=0}^{\bar{Q}} \begin{pmatrix} n_k^i \\ \tilde{n}_k^i \end{pmatrix} \varepsilon^{\tilde{n}_k^i} (1-\varepsilon)^{n_k^i-\tilde{n}_k^i}.
\end{aligned}$$

The latter equation implies

$$\begin{aligned}
\frac{\partial \Pr\left[\tilde{n}_k^i \geq 1-\bar{Q}\right]}{\partial \varepsilon} &= \sum_{\tilde{n}_k^i=0}^{\bar{Q}} \begin{pmatrix} n_k^i \\ \tilde{n}_k^i \end{pmatrix} \frac{\partial}{\partial \varepsilon}\left(\varepsilon^{\tilde{n}_k^i} (1-\varepsilon)^{n_k^i-\tilde{n}_k^i}\right) \\
&= \begin{pmatrix} n_k^i \\ 0 \end{pmatrix} \left(-n_k^i (1-\varepsilon)^{n_k^i-1}\right) + \\
&\quad \begin{pmatrix} n_k^i \\ 1 \end{pmatrix} \left((1-\varepsilon)^{n_k^i-1} - \left(n_k^i-1\right) \varepsilon (1-\varepsilon)^{n_k^i-2}\right) + \\
&\quad \begin{pmatrix} n_k^i \\ 2 \end{pmatrix} \left(2\varepsilon (1-\varepsilon)^{n_k^i-2} - \left(n_k^i-2\right) \varepsilon^2 (1-\varepsilon)^{n_k^i-3}\right) + \\
&\quad \ldots\ldots + \\
&\quad \begin{pmatrix} n_k^i \\ \bar{Q}-1 \end{pmatrix} \left(\left(\bar{Q}-1\right) \varepsilon^{\bar{Q}-2} (1-\varepsilon)^{n_k^i-\bar{Q}-1} - \left(n-\left(\bar{Q}-1\right)\right) \varepsilon^{\left(\bar{Q}-1\right)} (1-\varepsilon)^{n_k^i-\bar{Q}}\right) + \\
&\quad \begin{pmatrix} n_k^i \\ \bar{Q} \end{pmatrix} \left(\bar{Q}\varepsilon^{\bar{Q}-1} (1-\varepsilon)^{n_k^i-\bar{Q}} - \left(n-\bar{Q}\right) \varepsilon^{\bar{Q}} (1-\varepsilon)^{n_k^i-\bar{Q}-1}\right).
\end{aligned}$$

Evaluating these terms and canceling them pairwise, we obtain

$$\frac{\partial \Pr\left[\tilde{n}_k^i \geq 1-\bar{Q}\right]}{\partial \varepsilon} = -\begin{pmatrix} n_k^i \\ \bar{Q} \end{pmatrix} \left(n-\bar{Q}\right) \varepsilon^{\bar{Q}} (1-\varepsilon)^{n_k^i-\bar{Q}-1}.$$

A similar argument establishes:

$$\frac{\partial \Pr\left[\tilde{n}_k^i \leq \bar{Q}\right]}{\partial \varepsilon} = \begin{pmatrix} n_k^i \\ \bar{Q} \end{pmatrix} \left(n-\bar{Q}\right) (1-\varepsilon)^{\bar{Q}} \varepsilon^{n_k^i-\bar{Q}-1}.$$

Combining these two expressions, we have

$$\begin{aligned}
\frac{\partial \chi\left(n_k^i, \tilde{n}_k^i, \varepsilon, \bar{Q}\right)}{\partial \varepsilon} &= \frac{\partial \Pr\left[\tilde{n}_k^i \geq 1-\bar{Q}\right]}{\partial \varepsilon} + \frac{\partial \Pr\left[\tilde{n}_k^i \leq \bar{Q}\right]}{\partial \varepsilon}, \\
&= \begin{pmatrix} n_k^i \\ \bar{Q} \end{pmatrix} \left(n-\bar{Q}\right) \left((1-\varepsilon)^{\bar{Q}} \varepsilon^{n_k^i-\bar{Q}-1} - \varepsilon^{\bar{Q}} (1-\varepsilon)^{n_k^i-\bar{Q}-1}\right), \\
&= \begin{pmatrix} n_k^i \\ \bar{Q} \end{pmatrix} \left(n-\bar{Q}\right) (1-\varepsilon)^{n_k^i-1} \left(\left(\frac{1-\varepsilon}{\varepsilon}\right)^{\bar{Q}} - \left(\frac{1-\varepsilon}{\varepsilon}\right)^{n_k^i-\bar{Q}-1}\right) < 0,
\end{aligned}$$

where the last inequality follows using the facts that $\varepsilon < 1/2$ and $\bar{Q} \leq \left(n_k^i - 1\right)/2$.■

Intuitively, when ε is small, the performance of firms that have implemented the same technology in the past reveals more information about the reference action. Thus, firms' posterior beliefs are more responsive to public information. Note that this applies to both "correct" and "incorrect" beliefs. For instance, suppose that $x_k^* = L$, but in the sample available to the firm R has been successful more than half of the time; then, when ε is small, the firm will assign higher probability to R being the correct action (i.e., $\pi\left(0; h_k^i \mid x_{i,k} = R\right)$ will take on a larger value). The complication in establishing Proposition 2 comes from the fact that a change in ε affects the likelihood of different histories. Nevertheless, the proof establishes that a greater ε changes the ex ante distribution of different histories so as to also increase $\Pr\left(d_{i,k}^*\left(h_t^i\right) = 1\right)$.

Proposition 2 provides the most interesting testable implication of our approach; it suggests that there should be more decentralization in industries with greater dispersion of performance across firms and also for firms that are more dissimilar to others.

Finally, we briefly extend our basic model to derive a relationship between firm age and organizational structure. In the model analyzed so far the deviation between the reference action and the correct action for each firm was independently and identically distributed across technologies, firms and time. Consequently, a firm's information on how to implement technology k was independent from that firm's actions and performance on previous technologies $k' < k$. More generally, one could assume that there is a positive correlation between the correct actions that a firm should take across successive technologies, for example, because the specific skills of the employees or the culture of the organization differ across firms. In this case, the firm could learn from its own past experience as well as from the experiences of other firms.

Since solving the signal extraction problem with multiple sources of uncertainty is complicated and not our main focus here, we assume in this subsection that firms cannot learn from other firms. This enables us to focus instead on firms' learning from their own performance. The analogue to equation (1) in the text is

$$x_{i,k}^* = \begin{cases} x_i^* & \text{with probability } 1 - \varepsilon_i \\ \sim x_i^* & \text{with probability } \varepsilon_i \end{cases},$$

for any $I_i(k;t) = 1$. This equation implies that although the reference action for firm i is $x_i^* \in \{L, R\}$, the correct action for technology k may differ from this with some probability $\varepsilon_i < 1/2$.

This equation implies that the updating problem is now identical to that discussed previously, up to a reinterpretation of the information set. In particular, what used to be history $h_k^i = \left\{n_k^i, \tilde{n}_k^i\right\}$ is now replaced by $h_k^i = \left\{n_i^k, \tilde{n}_i^k\right\}$, where, given $I_i(k;t) = 1$, n_i^k denotes the number of technologies that firm i has implemented before technology k, and $\tilde{n}_i^k$ denotes the number of times in which action L turned out to be the correct choice in this firm's own experience in the past. Given this reinterpretation, our previous analysis implies (proof omitted):

Proposition 3 ***(Age)*** Consider the adoption decision of technology k by firm i, and suppose that Assumption 1 holds and $x_i^* = L$. Then:
(i) For the youngest firm with $n_i^k = 0$, we have $q\left(h_t^i = \{0,0\}\right) = \pi\left(0, h_t^i = \{0,0\}\right) = 1/2 < \delta$, and the principal chooses delegation, $d_{i,k}^*\left(h_t^i\right) = 1$;
(ii) For a sufficiently old firm,, i.e., $n_i^k \to \infty$, we have $p\lim_{n_i^k \longrightarrow \infty} q\left(h_t^i = \left\{n_i^k, \tilde{n}_i^k\right\}\right) = 1$, $p\lim_{n_k^i \longrightarrow \infty} \pi\left(0, h_t^i = \left\{n_i^k, \tilde{n}_i^k\right\}\right) = 1 - \varepsilon > \delta$, and the principal almost surely chooses centralization, i.e., $p\lim_{n_k^i \longrightarrow \infty} d_{i,k}^*\left(h_t^i = \left\{n_i^k, \tilde{n}_i^k\right\}\right) = 0$.

We next briefly discuss the possibility that firms may use incentive contracts to induce managers to choose the right action. Let the private benefit on the manager when the implements is preferred action be By_{t-1}. We will show that when B (the benefit accruing to the manager when she chooses her preferred action) is sufficiently large, such incentive contracts will not be optimal. The intuition is that because managers are credit constrained, incentive contracts give the right incentive to managers only by transferring rents to them. If B is large, this is not profitable for the principal.

Let us assume that the principal decides whether to hire the manager before knowing whether that $I_i(k,t) = 1$. Let us also normalize the outside option of the manager to zero, and recall that the manager is also risk neutral. Given the credit constraints of the manager , the optimal contract takes a simple form: the principal will pay the manager $By_{i,t-1}$ in case of success. Both when the manager is unsuccessful in the implementation of the new technology and when there is no new technology to be implemented, it is optimal for the principal to pay him zero. This contract will induce the manager to choose the right action. It will also meet his participation constraint, since the manager will receive $By_{i,t-1}$ with probability $p_i\delta > 0$.

The alternative is to pay the manager zero irrespective of success, and let him choose his preferred action. This contract also meets the manager's participation constraint, since he derives private benefits from the implementation of the project. This option was the one analyzed in the text. Since the issue of whether there is delegation or not is only interesting in the case when there is a technology to be implemented, let us focus on time t such . We then have:

Proposition 4 Suppose that $B > \frac{(\gamma-1)(1-\delta)}{1-\beta p_i\gamma}$ and that $I_i(k,t) = 1$. Then, for any history $h_k^i \in \mathcal{H}_k^t$, the optimal strategy for the principal of firm i is not to offer an incentive contract to the manager.

Proof. Let $d_{i,k} = 2$ denote firm i's decision to delegate control with full compensation to the manager for choosing the profit-maximizing action at date t on technology k when $I_i(k,t) = 1$. The value of the decentralized firm offering the manager an incentive contract is

$$\begin{aligned} V_i\left(y_i, h_k^i \mid 2\right) &= p_i[(\gamma - B)\, y_i + \beta\mathbb{E}\left[V_i\left(\gamma y_i, h_{k+1}^i\right) \mid h_k^i\right]] \\ &\quad + (1-p_i)\{y_i + \beta\mathbb{E}\left[V_i\left(y_i, h_{k+1}^i\right) \mid h_k^i\right]\}. \end{aligned}$$

Solving this functional equation leads to $V_i\left(y_i, h_k^i \mid 2\right) = A_i\left(h_k^i \mid 2\right) y_i$, where:

$$\begin{aligned} A_i\left(h_k^i \mid 2\right) &= [p_i(\gamma - B) + 1 - p_i] + \\ &\quad \beta[p_i\gamma + 1 - p_i] \max_{d_{i,k+1}\in\{0,1,2\}} \mathbb{E}\left[A_i\left(h_{k+1}^i \mid d_{i,k+1}\right) \mid h_k^i\right]. \end{aligned} \tag{18}$$

Instead, the value of a decentralized firm offering the manager a flat wage is $V_i\left(y_i, h_k^t \mid 1\right) = A_i\left(h_k^t \mid 1\right) y_i$, where

$$\begin{aligned} A_i\left(h_k^i \mid 1\right) &= [p_i(\delta\gamma + (1-\delta)) + 1 - p_i] \\ &\quad + \beta[p_i(\delta\gamma + (1-\delta)) + 1 - p_i] \max_{d_{i,k+1}\in\{0,1,2\}} \mathbb{E}\left[A_i\left(h_{k+1}^i \mid d_{i,k+1}\right) \mid h_k^i\right]. \end{aligned} \tag{19}$$

Which of the two regimes yields a larger value to the firm depends on whether $A_i\left(h_k^i \mid 1\right)$ is larger or smaller than $A_i\left(h_k^i \mid 2\right)$. To establish when this is the case, note that

$$\begin{aligned} A_i\left(h_k^i \mid 1\right) &> A_i\left(h_k^i \mid 2\right) \Leftrightarrow \\ B &> (\gamma-1)(1-\delta)\left(1 + \beta \max_{d_{i,k+1}\in\{0,1,2\}} \mathbb{E}\left[A_i\left(h_{k+1}^i \mid d_{i,k+1}\right) \mid h_k^i\right]\right). \end{aligned} \tag{20}$$

An upper bound to the future value of the firm can be calculated by assuming that, from period $(t+1)$ onwards, the firm will innovate successfully whenever a new technology opportunity arises, which takes place with probability p_i, and will pay no managerial wage. This yields

$$\max_{d_{i,k+1}\in\{0,1,2\}} \mathbb{E}\left[A_i\left(h_{k+1}^i \mid d_{i,k+1}\right) \mid h_k^i\right] < \bar{A} \equiv \frac{p_i\gamma}{1-\beta p_i\gamma}.$$

Thus, substituting $\max_{d_{i,k+1}\in\{0,1,2\}} \mathbb{E}\left[A_i\left(h_{k+1}^i \mid d_{i,k+1}\right) \mid h_k^i\right]$ by $\bar{A}$ we obtain the sufficient condition

$$B > \frac{(\gamma-1)(1-\delta)}{1-\beta p_i\gamma} \Rightarrow A_i\left(h_k^i \mid 1\right) > A_i\left(h_k^i \mid 2\right)$$

for incentive contracts not to be profitable for the principal. ■

Appendix B

B.1 French Data

COI ("Changements Organisationnels et Informatisation," SESSI)

This is a firm level survey providing information on organization and other firm characteristics conducted in 1997. It covers manufacturing sectors only (4,153 firms). There are several questions on organizational design.

ER (Enquête Reponse 1998; "Relations Professionnelles et Négociations d'Entreprise," DARES)

The Enquête Reponse is an establishment level survey. This contains information about organizational change between 1996 and 1998. It covers both manufacturing and non-manufacturing sectors and is an updated version of the Reponse 1992 survey used by Caroli and Van Reenen (2001). 2,943 establishments of manufacturing and non-manufacturing sectors were surveyed with senior managers being asked questions about industrial relations, organization and other aspects of performance in 1998.

FUTE files ("Format Unifié Total d'Entreprises," INSEE)

The FUTE dataset is the key data we use to construct many of the variables used in the paper. FUTE is constructed from the merging of two datasets, the BRN ("Bénéfices Réels Normaux") and the EAE ("Enquêtes Annuelles d'Entreprises"), that are then checked rigorously for consistency at INSEE.

The BRN files consist of firms' balance sheets collected annually by the Direction Générale des Impôts (Fiscal Administration) and provides firm-level accounting information (value added, capital investment, wage bills,employment, etc.). This tax regime is mandatory for the companies that have a level of sales higher than 3.8 million Francs, but can also be also disclosed by smaller firms. These files include around 600,000 firms,[32] in the private non-financial, non-agricultural sectors each year and covers around 80% of total output in the French economy. The EAE survey is conducted by SESSI (production industries), INSEE (Services and Trade), the Ministry of Agriculture and Ministry of Equipment (Transportation and Construction). The annual survey is mandatory and exhaustive for firms hiring more than 20 workers. It includes a detailed sectoral description of the various activities of each firm surveyed (the amount of each kind of output).[33]

DADS files ("Déclarations Annuelles de Données Sociales")

The DADS files consists of yearly mandatory employer reports of each worker's gross earnings subject to payroll taxes. Hours are also reported since 1993 (but of good quality only since 1994). These files include around 27 million workers each year (27,535,562 in 1996 after some basic cleaning), which we aggregate at the plant (1,587,157 plants in 1996) or firm level (1,379,032 firms in 1996) to get information on the workforce structure (age, gender, skill group in terms of hours worked). We also use the total hours series necessary for the measures of productivity underlying the heterogeneity and proximity to the frontier measures (see below).

[32] 630,593 firms in 1996 of which 489,783 report a strictly positive number of workers.

[33] In French, the question was :
"Répartir le chiffre d'affaires net hors taxes et les exportations directes de votre entreprise selon les différentes activités conformément aux nomenclatures officielles d'activités et de produits. Le total du chiffre d'affaires net doit correspondre au montant du poste du compte de résultat. Les reventes en l'état de marchandises ou de produits doivent être déclarées dans une ou plusieurs rubriques *négoce*."

LIFI Surveys ("Liaisons Financières," INSEE)

Yearly survey describing the structure of ownership of French firms of the private sector whose financial investments in other firms (participations) are higher than 8 million Francs or having sales above 400 million Francs or a number of workers above 500.

Even after keeping only firms who are in the COI, BRN, DADS and EAE we are still left with over 90% of the original COI sample (3,751 observations). The firms who we lose tend to be the smallest firms. We loose a few more observations in our regressions due to missing values on the some of the questions in COI (final sample for regressions is 3,570 observations). For the Reponse sample, we only keep firms that are part of a larger French or foreign group, but that are not the corporate head quarters (final sample for regressions is 1,258 observations).

B.2 Variable Definitions

The firm and industry level quantitative variables introduced in the regressions are averaged over four years (COI) if available (three years for Reponse, "Delayering"). Unless otherwise indicated all industry variables are at the four-digit NACE level.

B.2.1 Decentralization into Profit Centers

Our main measure of decentralization is from the COI. Managers were asked:

Is your firm organized into profit centers ?

The translation of the French definition used in COI is "*Organization in profit centers. A profit center is an enterprise unit that has a margin of budgetary manoeuvre, and therefore some relative autonomy in their choices (usually it has its own accounting system to measure their profit).*"

We coded the measure of decentralization to be unity if the manager answered "yes" to this question and zero if the answer was "no".

Using the COI we build a measure of decentralization based on the organization of its business units into profit centers.[34] In practice, once a firm gets beyond a minimal size it faces the choice of retaining central control or allowing some decentralization. Firms are generally organized into business units and different firms make decisions about what degree of responsibility to devolve to the managers of these units. Some firms retain complete command and control at the center, but most create some form of "responsibility centers" for business unit managers.[35] Business scholars delineate three broad types of responsibility centers (from the most to the least decentralized): profit centers, cost centers and revenue centers. Our key indicator for decentralization is whether the firm is organized primarily into profit centers. As its name suggests, when a firm organizes into profit centers a manager is responsible for the profits of the unit she manages. In general the profit center manager is given considerable autonomy to make decisions on the purchase of assets, hiring of personnel, setting salary and promotion schedules and managing inventories. A manager of a profit center is concerned with all aspects of the business that contribute to profitability. Such a manager keeps track of both revenues and costs with the aim of maximizing profit. As one management specialist puts it:

"The profit center managers frequently know their business better than top management does because they can devote much more of their time to following up developments in their specialized areas. Hence, top level managers usually do not have detailed knowledge of the actions they want particular profit center managers to take, and even direct monitoring of the actions taken, if it were feasible would not ensure profit center managers were acting appropriately." (*Motivating Profit Center Managers*, Merchant, 1989, p.10)

In contrast to a profit center manager, a cost center manager will have the quantity or quality of output set by someone higher up in the organization. The manager is delegated with some power,

[34]This follows Janod (2002) and Janod and Saint-Martin (2004).

[35]For an introduction to responsibility centers in general and profit centers in particular see, for example: http://smccd.net/accounts/nurre/online/chtr12a.htm.

however, in order to try and reduce costs. He will be able to decide on some short-run (but not long-term) asset purchases, hire temporary and contract staff (but not permanent employees) and manage inventories. A revenue center manager has the least autonomy of all.[36] She is told to spend a certain amount of resource and account for revenues but has no (or little) discretion to exceed spending limits. Inventories are managed but staff and investments are not acquired unless he is authorized explicitly to do so.

There are numerous examples from the business literature on the greater autonomy of profit centers. It is well recognized that organizing divisions into profit centers delegates more power to managers, and it is generally agreed that a characteristic of companies that organize divisions into profit centers is that it "allows decision making and power to be delegated effectively". Similarly, the first disadvantage of profit centers is viewed as "loss of overall central control of the company." [37]

Although it is possible in principle for a profit center manager to be monitored on profits and yet not be given any powers to affect these profits would seem sub-optimal for the firm (Dearden, 1987, Merchant, 1989, Bouwens and van Lent, 2004). A profit center manager would be held responsible for outcomes that he cannot affect, so this would de-motivate such managers. Some organizations like this probably exist - the only way to know more would be to have subjective questions on the degree to which different profit center managers have greater decision making powers. The advantage of our profit center variable is that it is an objective feature of the firm and does not rely on a manager's subjective statement of his power relative to a senior manager.

In short, we have an indicator equal to unity if the firm is organized into profit centers and a zero otherwise. So the base group contains firms who are organized primarily into responsibility centers with less autonomy (i.e. cost and revenue centers) and those firms who have no responsibility centers at all and maintain command and control. Unfortunately, the data does not allow us to distinguish the latter groups more finely.

B.2.2 Managerial Autonomy to Make Investment Decisions

In the Enquête Reponse 1998 the establishment's senior manager was asked how much autonomy from headquarters she had to make decisions over investment.[38] Answers were coded to be one if she answered that she had "full" autonomy or "important"' autonomy and coded to zero if she had "limited" or "no" autonomy. This is used as the dependent variable in the first three columns of Table 4. We consider only establishments that are part of a wider group (as a single site establishment will not have a separate headquarters).

B.2.3 Delayering

While COI provides data about the current organization of the firm, the ER dataset provides information on organizational "changes" (i.e., whether a firm became more or less hierarchical) rather than "levels".

Our preferred measure of delayering is from the Enquête Reponse 1998 where we use the following question:

For any of the following technologies and methods, would you tell us whether it is implemented in your establishment ?

- Shortening of the hierarchical line (delayering of an intermediate hierarchical level).

The indicator used is a dummy variable coded to one if the respondent answered "yes" to this question.

[36] In fact "revenue center" is rather a misnomer because a notional revenue is assigned by the organization's controller based on activities and transfer prices. "Expense center" is sometimes used as the manager accounts mainly for the expenses incurred.

[37] These quotes are taken from the educational web-site:
http://www.aloa.co.uk/members/downloads/PDF%20Output/costcentres.pdf
See also Janod (2002).

[38] In French: "Par rapport au siège ou à la maison mère de l'entreprise ou du groupe, quelle est l'autonomie de votre établissement en matière d'investissement?
Totale / Importante / Limitée / Nulle."

Case study and econometric evidence suggest that delayering is associated with decentralization (Rajan and Wulf, 2005, Caroli and Van Reenen, 2001).

B.2.4 Proximity to Technological Frontier

Value-added (VA_{ilt}, FUTE) is defined as sales minus purchases of materials. It is deflated with the value added price index at the two-digit level (NAF36) available from National Accounts (value added at 1995 prices)
Total hours ($HOURS_{ilt}$, DADS)
Capital Stock (K_{ilt}, FUTE) is computed from firm level fixed assets. This information is registered at historical cost in the balance sheets. We recover volumes by deflating the initial measure by the investment price index (National Accounts) at the date considered minus an estimated age of capital. This age is calculated as the ratio of depreciated assets over fixed assets multiplied by an average equipment length of life (16 years).

Labour Productivity and TFP are defined as:

$$\begin{aligned} y_{ilt} &= \ln(VA_{ilt}) - \ln(HOURS_{ilt}) \\ TFP_{ilt} &= \ln(VA_{ilt}) - \alpha_l \ \ln(HOURS_{ilt}) - (1-\alpha_l) \ \ \ln(K_{ilt}) \end{aligned}$$

Where α_l is the wage bill share of value added in the four-digit NACE (we also considered an economy-wide weight of 0.7). We drop firms reporting divergent values of total number of employees in the FUTE and in the DADS (values greater than double one way or the other). Industries represented by less than ten firms in the FUTE are also dropped. For each firm (and like other firm level variables introduced in the regressions) the labour productivity and TFP values are averaged over four years if available (three years respectively Delayering in Enquete Reponse).

The industry "frontier" ($y_{F_l t}$ or $TFP_{F_l t}$) is defined as the 99$^{\text{th}}$ percentile (or 95$^{\text{th}}$, or 90$^{\text{th}}$ when specified) of the obtained series at the NACE four-digit level. The constrained term defined as $GAP^y_{ilt} = \ln(y_{ilt}) - \ln(y_{F_l t})$ is a firm level measure of proximity to the technological frontier.

Another alternative measure of distance to frontier is the rank of firms in their industry (the firms are ranked according to their labour productivity in the regressions presented).

B.2.5 Heterogeneity Measures

We use the FUTE to construct $\Delta \ln y_{ilt}$, the firm specific annual productivity growth rate (value added per hour) for all firms. We average this growth rate for up to three years. We then construct the percentiles of the inter-firm productivity growth distribution within each four-digit NACE sector. The 90-10 is $(\Delta \ln y_{ilt})^{90} - (\Delta \ln y_{ilt})^{10}$ where $(\Delta \ln y_{ilt})^{90}$ is the productivity growth at the 90$^{\text{th}}$ percentile and $(\Delta \ln y_{ilt})^{10}$ is productivity growth at the 10$^{\text{th}}$ percentile. Alternative measures of heterogeneity are based on other indicators of dispersion of the same series of firm level labour productivity growth rates: the 95-5 is $(\Delta \ln y_{ilt})^{95} - (\Delta \ln y_{ilt})^{5}$, the standard deviation, the standard deviation after trimming bottom and top 5 % of values in each four-digit industry.

The levels-based measure of industry it originates is defined similarly as $(\ln y_{ilt})^{90} - (\ln y_{ilt})^{10}$.

In addition, we use the FUTE files to construct two additional firm - level measures of heterogeneity A firm i is characterized by its vector of kind of productions (sold, l being one of its markets):

$$S_i = (S_{i1}, \ldots, S_{il}, \ldots, S_{iL}), \qquad \text{or in shares :} \quad s_i = \left(\underbrace{\frac{S_{i1}}{\sum_{h\in\mathcal{L}} S_{ih}}}_{s_{i1}}, \ldots, \underbrace{\frac{S_{il}}{\sum_{h\in\mathcal{L}} S_{ih}}}_{s_{il}}, \ldots, \underbrace{\frac{S_{iL}}{\sum_{h\in\mathcal{L}} S_{ih}}}_{s_{iL}} \right)$$

where $\mathcal{L}$ refers to the set of industries. Our main index of firm-level heterogeneity is constructed as

$$H_i^F = \log\left(\frac{\sum_{i'\in\mathcal{N},i'\neq i} c_{ii'}\cdot IT_{i'}}{\sum_{i'\in\mathcal{N},i'\neq i} IT_{i'}}\right)^{-1}$$

for firm i at time t, with N referring to the sample of firms in the FUTE, and IT_i refers to the level of IT investment of firm i,[39] and the closeness measure is

$$c_{ii'} = \frac{\sum_{l\in\mathcal{L}} s_{il}.s_{i'l}}{(\sum_{l\in\mathcal{L}} s_{il}^2)^{\frac{1}{2}}.(\sum_{l\in\mathcal{L}} s_{i'l}^2)^{\frac{1}{2}}},$$

The alternative measure is an unweighted version of our main measure calculated as:

$$H_i^{FA} = \log\left(\frac{\sum_{i'\in\mathcal{N},i'\neq i} c_{ii'}}{N-1}\right)^{-1},$$

where N is the total number of firms in the set $\mathcal{N}$.

B.2.6 Other Firm Level Variables

All firm level variables are averaged over four years if available.

Lerner Index: We calculate gross profits as value added minus labor costs and then divide by sales. This proxies for the average profit margin. All these variables are sourced from FUTE. We considered also making a control for total capital costs using the stock of capital and the average user-cost (this is also deducted from gross profits to allow a normal rate of return on fixed capital). This lead to very similar results.

Capital Intensity: Fixed capital stock divided by value added. Sourced from FUTE.

Firm / Plant age: Information available from the SIRENE dataset (reproduced in the DADS) or from files reporting the yearly creations of firms (Firm Demography Department). Plant age is available in the Reponse survey.

Joint Stock Firms: Indicator of a firm being a Joint Stock Company (as opposed to smaller and less anonymous structures, e.g. limited liability firms). Sourced from FUTE.

Foreign Ownership: Indicator of whether a firm is part of a larger foreign group. Sourced from LIFI.

Number of Plants: Number of plants belonging to each firm (and their region of localization). Sourced from DADS.

Size: The *number of workers* at the plant level for Reponse and at the firm level for COI. Sourced from DADS.

Skills: Share of hours worked by skilled workers at the firm level. We consider as unskilled: *Industrial blue collar workers* (CS 67, Ouvriers non qualifiés de type industriel); *Craftsmen* (CS 68, Ouvriers non qualifiés de type artisanal), *Foremen and Supervisors* (CS 53, Agents de surveillance), *Clerical* (CS 55, Employés de commerce), *Personnel of the direct services to the private individuals sectors* (CS56, Personnels des services directs aux particuliers). Others are considered as "skilled". Sourced from DADS.

Worker age: Average age of workers at the firm level (weighted by hours worked). Sourced from DADS.

[39] Note that the IT investment series (EAE/FUTE) are available in 1996 and 1997.

Technology: A pseudo-continuous variable of proportion of workers using micro-computers is constructed from information available in both Reponse 98 (relating to 1998) and COI (1997).

We use the FUTE dataset and the decomposition of the various activities of firms i in terms of amount of each kind of product l produced and sold (S_{il}) to construct the following indicators:

Firm-level Market Share:

$$MS_i = \sum_l \frac{S_{il}}{S_i} . \frac{S_{il}}{S_l}$$

Herfindahl Index:

$$HR_i = \sum_l \frac{S_{il}}{S_i} .HR_l, \quad HR_l = \sum_i \left(\frac{S_{il}}{\sum_{i'} S_{i'l}} \right)^2$$

This is constructed in the standard way at the industry level (H_l), but note that we weight this measure if a firm operates in more than one market (by a firm's market share in that sector ($\frac{S_{il}}{S_i}$), so it has a firm specific component).

Firm-level Diversification Indicator:

This simply indicates (the inverse of) the degree to which a firm operates across separate sectors

$$SPE_i = \sum_l \left(\frac{S_{il}}{S_i} \right)^2 .$$

B.2.7 Other Industry Level Information

All industry level variables are averaged over three years.

Sector Capital Intensity total capital stock in the four-digit industry divided by total number of workers in the industry. Sourced from FUTE.

Sector IT Investment: total IT investment in the four-digit industry divided by total number of workers in industry. Sourced from FUTE.

B.3 UK Data

B.3.1 Workplace Employee Relations Survey (WERS)

WERS 1998 is a survey of establishments in Britain conducted in 1998 (there were also surveys in 1980, 1984 and 1990).[40] It is described in detail in Cully *et al.* (1999). In one part of the questionnaire the establishment's senior manager is asked whether she "is able to make decisions without consulting" Head Quarters. Some of these decisions are relatively minor (such as staff appraisal). We focus on whether decisions over staff recruitment can be made by establishment's management without consulting someone higher in the corporate hierarchy as this is a key aspect of decentralized decision making (unfortunately the question on investment decisions used in France was not asked).

This question was only asked if the establishment was part of a larger multi-plant firm. Although the question was asked to all establishments we have to focus on manufacturing because the ABI data is only available for services from 1997 onwards so we would not be able to construct a robust measure of heterogeneity. The WERS data cannot be matched at the establishment-level to Census data so we are unable to condition on as rich a set of covariates as we can in France. In particular, we do

[40] The 1984 and 1990 panels were used by Caroli and Van Reenen (2001). There was also a WERS conducted in 2004, but the four-digit industry codes for this data set have not yet been released (there are only 12 industry divisions).

not have information on value added, profits or capital. Consequently we cannot include measures of the establishment's own productivity or Lerner Index in the regression. WERS does contain basic information on workers demographics (skill, age, female and part-timers) and we condition on these in the regressions (see Table notes). As a proxy for market power we used the question asked to managers whether the establishment faces no competitors, some competitors or many competitors (this is the same as Nickell, 1996).

B.3.2 ABI "Census" Data

We constructed heterogeneity and Frontier productivity terms for each UK four-digit industry. The UK and France share the European Unions' NACE classification system so this was straightforward. The only restriction was that industry averages with cell sizes below 25 are not allowed out of the Office of National Statistics (ONS).

Our base dataset is a panel of establishments covering almost all sectors of the UK private sector called the ABI (Annual Business Inquiry). This underlies many of the UK national statistics and is similar in structure to the US Longitudinal Research Database (LRD) being a population sample of large plants and a stratified random sample of smaller plants. The response rates to the ABI are high because it is illegal not to return the forms to the Office of National Statistics (ONS). The ABI contains all the basic information needed to estimate production functions (gross output, labour, materials, investment, etc.). For each firm we constructed value per worker and followed the same rules described for France to calculate heterogeneity (e.g. 90-10 of productivity growth rates) and the Frontier productivity (99th percentile). The UK data does not contain information on hours so the UK productivity measure is cruder than in France.

C Robustness Checks

The robustness checks discussed in Section 5.4 are presented in Tables B1 and B2. Table B1 includes the results with alternative estimation strategies, additional controls and alternative samples. Table B2 includes the results of the instrumental variables strategy discussed in the text together with the corresponding first stages.

Table B1: Robustness checks on the probability of firm being decentralized
(Enquête COI)

Dependent variable (mean=0.304)	Firm decentralized into Profit Centers						
	Baseline (marginal effects) (1)	Logit ML (marginal effects) (2)	Linear probability model (3)	Frontier: TFP (4)	Firm rank in productivity distribution (5)	Frontier: 95^{th} percentile (6)	Frontier: 90^{th} percentile (7)
Heterogeneity	0.296	0.293	0.261	0.243	0.239	0.221	0.181
	(0.127)	(0.121)	(0.119)	(0.095)	(0.105)	(0.121)	(0.117)
Frontier term	-0.225	-0.228	-0.191	-0.066	-	-0.179	-0.104
(99^{th} percentile)	(0.045)	(0.044)	(0.040)	(0.023)		(0.077)	(0.091)
Labour productivity	0.141	0.144	0.128	0.202	-	0.137	0.132
(firm level)	(0.033)	(0.033)	(0.030)	(0.060)		(0.033)	(0.033)
Firm rank	-	-	-	-	-0.038	-	-
					(0.012)		
Firm age<5 years	0.172	0.172	0.151	0.157	0.164	0.170	0.170
	(0.041)	(0.044)	(0.034)	(0.042)	(0.041)	(0.041)	(0.041)
5≤ Firm age<10 years	0.066	0.061	0.057	0.075	0.02	0.065	0.065
	(0.022)	(0.023)	(0.017)	(0.023)	(0.022)	(0.022)	(0.022)
10≤Firm age<20 years	0.039	0.035	0.035	0.041	0.039	0.040	0.04
	(0.019)	(0.020)	(0.016)	(0.020)	(0.019)	(0.019)	(0.019)
Lerner index	-0.660	-0.672	-0.586	-0.543	-0.521	-0.636	-0.634
	(0.144)	(0.144)	(0.124)	(0.146)	(0.123)	(0.144)	(0.144)
Industry dummies	yes (73)	yes (73)	yes (73)	yes (73)	yes (73)	yes (73)	yes (73)
Observations	3,570	3,570	3,570	3,479	3,570	3,570	3,570

NOTES: All coefficients are marginal effects from probit maximum likelihood estimation except column (2) which is estimated by by Logit (marginal effects at the sample mean reported) and column (3) which is estimated by by OLS (linear probability model). Robust standard errors corrected for arbitrary variance-covariance matrix at the four-digit industry level in parentheses. Heterogeneity is defined as the dispersion of productivity growth rates within a four digit industry (the 90^{th} percentile less the 10^{th} percentile). Full set of firm and industry level controls included; see text for variable definitions. Exact definition of each experiment is described in text. In column (4), the proximity to frontier terms are expressed in TFP instead of labor productivity. In column (5), the indicator of proximity to frontier is the firm's rank (in terms of labor productivity) in its four-digit industry. In columns (6) and (7), the frontier term is defined as the 95^{th} or 90^{th} percentile of the four-digit industry labor productivity (instead of the 99^{th} in the main specification).

Table B1: (-cont.) Robustness checks on the probability of firm being decentralized

(Enquête COI)

Dependent variable (mean=0.304)	Firm decentralized into Profit Centers				
	Weighted: log(employment)	Part of a larger group	Joint stock firms	Specialized firms	French owned
	(8)	(9)	(10)	(11)	(12)
Heterogeneity	0.368	0.461	0.403	0.324	0.309
	(0.138)	(0.140)	(0.116)	(0.159)	(0.120)
Frontier term	-0.251	-0.303	-0.192	-0.179	-0.152
(99^{th} percentile)	(0.051)	(0.056)	(0.041)	(0.053)	(0.041)
Labour productivity	0.152	0.146	0.144	0.110	0.155
(firm level)	(0.040)	(0.052)	(0.042)	(0.034)	(0.032)
Firm age<5 years	0.151	0.125	0.224	0.184	0.158
	(0.045)	(0.053)	(0.042)	(0.050)	(0.048)
5≤ Firm age<10 years	0.060	0.059	0.084	0.073	0.067
	(0.025)	(0.035)	(0.025)	(0.024)	(0.024)
10≤Firm age<20 years	0.044	0.061	0.048	0.026	0.032
	(0.022)	(0.029)	(0.022)	(0.021)	(0.021)
Lerner index	-0.688	-0.827	-0.658	-0.533	-0.675
	(0.173)	(0.244)	(0.174)	(0.156)	(0.139)
Industry dummies	yes (73)	yes (67)	yes (67)	yes (73)	yes (73)
Observations	3,570	1,793	2,990	2,555	2,951

NOTES: (-cont.) In column (8), all firms are weighted (in the regression) by the log of their employment. In column (9) “Part of a larger group” means that the firm is part of a larger group (either French or foreign). In column (10) “Joint stock firm” means that the firm is a joint stock fim, as opposed to limited liability firms, etc. In column (11) “Specialized” indicates that the firm has at least 80% of its sales in one four-digit industry. In column (12) “French owned” means that the firm is part of a French group.

Table B2: Allowing for Potential Endogeneity of Heterogeneity and Proximity to Frontier
(Enquête COI)

Dependent variable	Profit centers	Heterogeneity (France)	Profit centers	Proximity (France)	Profit centers	Heterogeneity (France)	Proximity (France)	Profit centers
Assumptions over the exogeneity of heterogeneity and proximity	Both exogenous (1)	(First stage) (2)	Endogenous heterogeneity (3)	(First stage) (4)	Endogenous proximity (5)	(First stage) (6)	(First stage) (7)	Both endogenous (8)
UK Heterogeneity	-	0.156	-	-	-	0.157	-0.093	-
		(0.048)				(0.054)	(0.099)	
UK Frontier	-	-	-	-0.218	-	0.005	-0.217	-
(99^{th} percentile)				(0.053)		(0.027)	(0.057)	
Heterogeneity	0.230	-	1.185	-0.419	0.310	-	-	1.572
(France)	(0.120)		(0.681)	(0.230)	(0.130)			(0.699)
Proximity to frontier	0.167	-0.017	0.187	-	0.341	-	-	0.456
(France)	(0.029)	(0.008)	(0.185)		(0.123)			(0.194)
Firm age<5 years	0.166	0.007	0.156	-0.054	0.175	0.008	-0.057	0.165
	(0.041)	(0.003)	(0.042)	(0.027)	(0.042)	(0.003)	(0.027)	(0.042)
5≤ Firm age<10 years	0.071	0.003	0.066	-0.030	0.077	0.004	-0.031	0.073
	(0.022)	(0.003)	(0.022)	(0.015)	(0.023)	(0.003)	(0.015)	(0.022)
10≤Firm age<20 years	0.043	0.005	0.038	-0.006	0.044	0.005	-0.008	0.037
	(0.020)	(0.002)	(0.020)	(0.011)	(0.020)	(0.002)	(0.011)	(0.020)
Lerner index	-0.723	0.035	-0.767	2.896	-1.228	-0.015	2.903	-1.542
	(0.137)	(0.025)	(0.545)	(0.109)	(0.387)	(0.015)	(0.110)	(0.570)
Industry dummies	yes (71)	yes (71)	yes (71)	yes (71)	yes (71)	yes (71)	yes (71)	yes (71)
Observations	3,518	3,518	3,518	3,518	3,518	3,518	3,518	3,518

NOTES: All coefficients are marginal effects from probit (multivariate normal) maximum likelihood estimation. Robust standard errors corrected for arbitrary variance-covariance matrix at the four-digit industry level in parentheses. Heterogeneity is defined as the dispersion of productivity growth rates within a four digit industry (the 90^{th} percentile less the 10^{th} percentile). Full set of firm and industry level controls included; see text for variable definitions. Observations only kept when match between UK and French four-digit industry is possible. In column (1), heterogeneity and proximity to frontier are both assumed exogenous in the "profit centers" equation. In columns (2) and (3), heterogeneity is assumed endogenous, while proximity to frontier is assumed exogenous. In columns (4) and (5), proximity to frontier is assumed endogenous, while heterogeneity is assumed exogenous. In columns (6) to (8), heterogeneity and proximity to frontier are both assumed endogenous.