Online Appendix for
Good Rents versus Bad Rents:
Firm size and R&D Misallocation

Philippe Aghion, Antonin Bergeaud
Timo Boppart, Peter J. Klenow, Huiyu Li

08/21/2025

Click here for the latest version of the paper

Contents

A Model derivations

A.1 Constrained planner's problem

First, we consider a planner who can allocate production and R&D labor across firms under the constraint the R&D needs to be done in-house, i.e., a firm needs itself to have innovated in a specific line in order to produce at a particular quality level. With homogeneity within the four type of firms $k \in \{HB, HS, LB, LS\}$, the planner's problem can be characterized as follows

$$\max_{\{C_t, Q_{t+1}, \{n_{k,t+1}, x_{k,t}\}_{\forall k}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \log(C_t), \tag{A.1}$$

where $k \in \{HB, HS, LB, LS\}$, subject to

$$C_t = Q_t \exp\left(\sum_k J\phi_k n_{k,t} \log(a_k)\right)\left(1 - \sum_k \psi_{ok} J\phi_k (n_{k,t})^2\right) L, \tag{A.2}$$

$$Z = \psi_z \sum_k J\phi_k x_{k,t}, \tag{A.3}$$

$$Q_{t+1} = Q_t \exp\left(\sum_k J\phi_k x_{k,t} \log(\gamma_k)\right), \tag{A.4}$$

$$n_{k,t+1} = n_{k,t}\left(1 - \sum_{k'} J\phi_{k'} x_{k',t}\right) + x_{k,t}, \ \forall k, \tag{A.5}$$

and a given $Q_0 = \exp\left(\int_0^1 \log(q_0(i))\, di\right)$, $n_{HB,0}$, $n_{LB,0}$, $n_{HS,0}$, and $n_{LS,0}$ and non-negativity constraints

$$n_{k,t+1} \geqslant 0, \ x_{k,t} \geqslant 0, \ \forall k, t. \tag{A.6}$$

Equation (A.2) captures the resource constraint, i.e., that consumption equals output minus overhead. Here we already have exploited the fact that it is always optimal to set $l(i) = L$ due to the Cobb-Douglas technology. Furthermore, we exploited that it is always optimal to produce by the highest quality firm in a given line.[1] Output can then be written as the product of the geometric mean of quality across lines, $Q_t \equiv \exp\left(\int_0^1 \log\left(q_t(i)\right) di\right)$, the (weighted) geometric mean of process efficiency, $\exp\left(\sum_k J\phi_k n_{k,t} \log\left(a_k\right)\right)$, and L. The term $\left(1 - \sum_k \psi_{ok} J\phi_k (n_{k,t})^2\right)$ captures output net of overhead. Equation (A.3) captures the constraint on researcher labor. Finally, (A.4) captures the law of motion of the average quality level and (A.5) gives the law of motion of the number of highest quality lines by each type of firm.

When defining $\tilde{\psi}_o = 2\psi_o$, we can write the Lagrangian associated with the planner's problem as

$$\begin{aligned}
\mathscr{L} \;=\; & \sum_{t=0}^{\infty} \beta^t \left[\log(Q_t) + \sum_k J\phi_k n_{k,t} \log(a_k) + \log\left(1 - \frac{\tilde{\psi}_o}{2} \sum_k J\phi_k n_{k,t}^2\right) + \log(L) \right] \\
+ & \sum_{t=0}^{\infty} \lambda_t^Q \left[Q_t \exp\left[\sum_k J\phi_k x_{k,t} \log\left(\gamma_k\right) \right] - Q_{t+1} \right] \\
+ & \sum_{t=0}^{\infty} \beta^t \lambda_t^z \left(Z - \sum_k J\phi_k \psi_z x_{k,t} \right) \\
+ & \sum_{t=0}^{\infty} \sum_{k'} \beta^t \lambda_t^{nk'} \left[n_{k',t} \left(1 - \sum_k J\phi_k x_{k,t}\right) + x_{k',t} - n_{k',t+1} \right] \\
+ & \sum_{t=0}^{\infty} \sum_k \beta^t \lambda_t^{xk} (x_{k,t} - 0),
\end{aligned}$$

where the last line deals with the non-negativity constraint on the R&D investment and the penultimate line is the evolution of number of products for each firm type k' given the aggregate rate of creative destruction $\sum_k J\phi_k x_{k,t}$.

We assumed $Z \leqslant \psi_z$ so that research labor is scarce in that there is not enough research labor to innovate on every line. We derive the following first-order conditions:

$$\frac{\beta^{t+1}}{Q_{t+1}} + \lambda_{t+1}^Q \frac{Q_{t+2}}{Q_{t+1}} = \lambda_t^Q,$$

$$\lambda_t^z \psi_z J\phi_k + J\phi_k \sum_{k'} \lambda_t^{nk'} n_{k',t} = \beta^{-t} \lambda_t^Q J\phi_k \log\left(\gamma_k\right) Q_{t+1} + \lambda_t^{nk} + \lambda_t^{xk},$$

[1]This follows from the restriction $\gamma_S > a_H/a_L$ we made above.

$$\lambda_t^{xk} x_{k,t} = 0, x_{k,t} \geqslant 0, \lambda_t^{xk} \geqslant 0,$$

$$\beta \left(J\phi_k \log(a_k) - \frac{\tilde{\psi}_o J\phi_k n_{k,t+1}}{1 - \tilde{\psi}_o \sum_{k'} \frac{1}{2} J\phi_{k'} (n_{k',t+1})^2} \right) + \beta \lambda_{t+1}^{nk} \left(1 - \sum_{k'} J\phi_{k'} x_{k',t+1} \right) = \lambda_t^{nk},$$

and all the remaining constraints. The LHS of the second equation is the cost of R&D while the RHS is the benefit plus the Lagrangian multiplier. The equation says that the multiplier may be positive and hence R&D is 0 when the cost of R&D exceeds the benefit.

Constrained planner's FOC along the balanced growth path

We will use $\bar{v}$ to denote the value of any variable v along the balanced growth path (BGP). Along a BGP we must have

$$\frac{C_{t+1}}{C_t} = \frac{Q_{t+1}}{Q_t} = \beta \left(\frac{\lambda_{t+1}^Q}{\lambda_t^Q} \right)^{-1} = \exp \left[\sum_k J\phi_k \bar{x}_k \log(\gamma_k) \right] \equiv 1 + \bar{g},$$

where $\bar{g}$ is some endogenous constant. Furthermore, λ^{nk} and λ^z need to be constant over time, i.e., we have $\lambda_t^{nk} = \bar{\lambda}^{nk}$, $\forall t, k$, and $\lambda_t^z = \bar{\lambda}^z$, $\forall t$. Moreover, all innovating firms have the same rate of innovation

$$\frac{\bar{x}_k}{\bar{n}_k} = \frac{Z}{\bar{\psi}_z}, \text{ where } \bar{\psi}_z \equiv \psi_z \sum_{k' \in I} J\phi_{k'} \bar{n}_{k'},$$

where I denotes the set of innovating firms. Since only innovating firms have positive number of products on the BGP, their product shares sum to one and $\bar{\psi}_z = \psi_z$. Finally, R&D labor across firms sum to total R&D labor, i.e.,

$$\sum_k J\phi_k \psi_z \bar{x}_k = Z.$$

Combining the above first-order conditions and balanced growth restrictions, we have the following conditions for the balanced growth path in the planner's solution

$$\frac{1}{Q_0} = \lambda_0^Q(\beta^{-1} - 1)(1 + \bar{g}), \tag{A.7}$$

$$J\phi_k \left(\log(a_k) - \frac{\tilde{\psi}_o \bar{n}_k}{1 - \bar{o}}\right) = \bar{\lambda}^{nk}\left(\frac{1}{\beta} - 1 + \frac{Z}{\psi_z}\right), \ \bar{o} \equiv \tilde{\psi}_o \sum_k \frac{1}{2} J\phi_k (\bar{n}_k)^2, \tag{A.8}$$

$$\bar{\lambda}^z \psi_z J\phi_k + J\phi_k \sum_{k'} \bar{\lambda}^{nk'} \bar{n}_{k'} = \lambda_0^Q J\phi_k \log(\gamma_k)(1 + \bar{g})Q_0 + \bar{\lambda}^{nk} + \bar{\lambda}^{xk}, \tag{A.9}$$

$$\bar{\lambda}^{xk}\bar{x}_k = 0, \bar{\lambda}^{xk} \geqslant 0, \bar{x}_k \geqslant 0, \tag{A.10}$$

$$\bar{n}_k \frac{Z}{\psi_z} = \bar{x}_k, \ \forall k, \tag{A.11}$$

$$\sum_k \phi_k J \bar{n}_k = 1. \tag{A.12}$$

Substituting (A.7) and (A.9) into (A.8) gives

$$\log(a_k) - \frac{\tilde{\psi}_o \bar{n}_k}{1 - \bar{o}} = \left(\frac{1}{\beta} - 1 + \frac{Z}{\psi_z}\right)\left(\bar{\lambda}^z \psi_z + \sum_{k'} \bar{\lambda}^{nk'} \bar{n}_{k'} - \frac{\log(\gamma_k)}{1/\beta - 1} - \frac{\bar{\lambda}^{xk}}{J\phi_k}\right). \tag{A.13}$$

Multiplying (A.9) by $\bar{n}_k$, summing across k and applying $\bar{\lambda}^{xk}\bar{n}_k = 0$ yield

$$\bar{\lambda}_z = \frac{1}{1/\beta - 1} \frac{\sum_k J\phi_k \bar{n}_k \log(\gamma_k)}{\psi_z}. \tag{A.14}$$

This implies for the difference between any two $\bar{n}_k$ and $\bar{n}_{k'}$ satisfies:

$$\bar{n}_k - \bar{n}_{k'} = \frac{1 - \bar{o}}{\tilde{\psi}_o}\left[\log\left(\frac{a_k}{a_{k'}}\right) + \left(\frac{1}{\beta} - 1 + \frac{Z}{\psi_z}\right)\left(\frac{\log\left(\frac{\gamma_k}{\gamma_{k'}}\right)}{1/\beta - 1} + \frac{\bar{\lambda}^{xk}}{J\phi_k} - \frac{\bar{\lambda}^{xk'}}{J\phi_{k'}}\right)\right], \tag{A.15}$$

Equations (A.10), (A.12), and (A.15) pin down all the $\bar{n}_k$. The optimal growth rate is then given by

$$1 + \bar{g} = \exp\left(\frac{Z}{\bar{\psi}_z}\left[\sum_k J\phi_k \bar{n}_k \log(\gamma_k)\right]\right) = (\bar{\gamma})^{\frac{Z}{\psi_z}}. \tag{A.16}$$

where $\bar{\gamma} = \Pi_k \gamma_k^{J\phi_k \bar{n}_k}$ is the geometric mean of the step size γ_k weighted by the share of lines $J\phi_k\bar{n}_k$.

When the solution is interior (all firms produce), (A.15) reduces to

$$\bar{n}_k - \bar{n}_{k'} = \frac{1-\bar{o}}{\tilde{\psi}_o}\left[\log\left(\frac{a_k}{a_{k'}}\right) + \left(1 + \frac{Z/\psi_z}{1/\beta - 1}\right)\log\left(\frac{\gamma_k}{\gamma_{k'}}\right)\right] \tag{A.17}$$

To solve for the equilibrium, we guess the set of producing firms and check the equilibrium conditions. Let I be the set of innovating firms. Define the share of firms among innovating firms $\phi_k^I \equiv \phi_k/(\sum_{k'\in I}\phi_{k'})$. We can combine (A.12) and (A.15) for innovating firms to derive

$$\bar{n}_k = \frac{1}{J\sum_{k'\in I}\phi_{k'}} + \frac{1-\bar{o}}{\tilde{\psi}_o}\nu_k,\ \forall k \in I, \tag{A.18}$$

where

$$\begin{aligned}\nu_k \equiv\ & \log(a_k) + \frac{1/\beta - 1 + Z/\psi_z}{1/\beta - 1}\left(\log(\gamma_k)\right) \\ & - \sum_{k'\in I}\phi_{k'}^I\left(\log(a_{k'}) + \frac{1/\beta - 1 + Z/\psi_z}{1/\beta - 1}\left(\log(\gamma_{k'})\right)\right)\end{aligned}$$

summarizes the technology level of type-k firms relative to the average innovating firm. For all $k \notin I$ we have $\bar{n}_k = 0$. The term $1/(J\sum_{k'\in I}\phi_{k'})$ is the average number of products per innovating firm. Equation (A.18) says that a firm has more products than average if its technology is "better" than average ($\nu_k > 0$).

Substituting equation (A.18) into the definition of $\bar{o}$ yields

$$\frac{1-\bar{o}}{\tilde{\psi}_o} = \frac{1}{\tilde{\psi}_o} - \frac{J\sum_{k'\in I}\phi_{k'}}{2}\sum_{k\in I}\phi_k^I\left(\frac{1}{J\sum_{k'\in I}\phi_{k'}} + \frac{1-\bar{o}}{\tilde{\psi}_o}\nu_k\right)^2.$$

Applying the quadratic formula and $\sum_{k\in I}\phi_k^I\nu_k = 0$ produces

$$\frac{1-\bar{o}}{\tilde{\psi}_o} = \frac{\sqrt{1+\left(\frac{2J}{\tilde{\psi}_o}-1\right)\left(\sum_{k'\in I}\phi_{k'}\right)\sum_{k\in I}\phi_k^I\nu_k^2}-1}{J\left(\sum_{k'\in I}\phi_{k'}\right)\sum_{k\in I}\phi_k^I\nu_k^2}. \tag{A.19}$$

Hence, a necessary condition for the existence of a solution with $\bar{o}\in(0,1)$ is

$$0 < \tilde{\psi}_o\frac{\sqrt{1+\left(\frac{2J}{\tilde{\psi}_o}-1\right)\left(\sum_{k'\in I}\phi_{k'}\right)\sum_{k\in I}\phi_k^I\nu_k^2}-1}{J\left(\sum_{k'\in I}\phi_{k'}\right)\sum_{k\in I}\phi_k^I\nu_k^2} < 1. \tag{A.20}$$

It follows from (A.19) that the weights in the growth equation are

$$\bar{S}_k = J\phi_k\bar{n}_k = \phi_k^I\left(1+\nu_k\frac{\sqrt{1+\left(\frac{2J}{\tilde{\psi}_o}-1\right)\left(\sum_{k'\in I}\phi_{k'}\right)\sum_{k\in I}\phi_k^I\nu_k^2}-1}{\sum_{k\in I}\phi_k^I\nu_k^2}\right). \tag{A.21}$$

Equation (A.21) and the definitions of $\bar{\gamma}$ and $\bar{\psi}_z$ together solve for $\bar{S}_k$.

The above solution is contingent on knowing the set of innovating firms. To check that our guess of I is correct, we can substitute $\bar{n}_k$ at the guess into (A.8) to calculate $\bar{\lambda}^{nk}$ and then check that for non-innovating firms, $\bar{n}_k = 0$ satisfies (A.13) for some non-negative $\bar{\lambda}^{xk}$.

With additional heterogeneity in the ψ_{ok} parameters the optimal market share of type k firms looks as follows: Let I again denote the set of firm types that produces. Their share of firms is $\phi^I \equiv \sum_{k'\in I}\phi_{k'}$. The firm distribution amongst producing firms is $\phi_k^I \equiv \phi_k/\phi^I$ and the harmonic mean of the overhead cost shifter for producing firms is $\bar{\tilde{\psi}}_o^I \equiv \left(\sum_{k\in I}\phi_k^I\frac{1}{\tilde{\psi}_{ok}}\right)^{-1}$. The level of $\bar{n}_k$ is determined by combining (18) with the accounting equation (6) on the aggregate number of products. This can be solved to arrive at

$$\bar{S}_k = \phi_k^I \frac{\bar{\tilde{\psi}}_o^I}{\tilde{\psi}_{ok}} \left(1 + \frac{\nu_k}{\bar{\tilde{\psi}}_o^I} \frac{\sqrt{1+\left(2J\phi^I - \sum_{k\in I}\phi_k^I \tilde{\psi}_{ok}\left(\frac{\bar{\tilde{\psi}}_o^I}{\tilde{\psi}_{ok}}\right)^2\right)\sum_{k\in I}\phi_k^I \nu_k^2/\tilde{\psi}_{ok}} - 1}{\sum_{k\in I}\phi_k^I \nu_k^2/\tilde{\psi}_{ok}} \right) \tag{A.22}$$

where

$$\begin{aligned} \nu_k \quad &\equiv \quad \log(a_k) + \frac{1/\beta - 1 + Z/\psi_z}{1/\beta - 1}\left(\log(\gamma_k)\right) \\ &- \quad \sum_{k'\in I} \phi_{k'}^I \frac{\bar{\tilde{\psi}}_o^I}{\tilde{\psi}_{ok'}} \left(\log(a_{k'}) + \frac{1/\beta - 1 + Z/\psi_z}{1/\beta - 1}\left(\log(\gamma_{k'})\right)\right). \end{aligned}$$

Again, here we assume that parameters are such that there is an interior BGP. This is guaranteed as long as the following condition holds

$$0 < \frac{\sqrt{1+\left(2J\phi^I - \sum_{k\in I}\phi_k^I \tilde{\psi}_{ok}\left(\frac{\bar{\tilde{\psi}}_o^I}{\tilde{\psi}_{ok}}\right)^2\right)\sum_{k\in I}\phi_k^I \nu_k^2/\tilde{\psi}_{ok}} - 1}{J\phi^I \sum_{k\in I}\phi_k^I \nu_k^2/\tilde{\psi}_{ok}} < 1. \tag{A.23}$$

This is a necessary and sufficient condition for the solution given by (A.22) to generate $\bar{o} \in (0, 1)$. When this condition is satisfied, we can show that the share of products $\bar{S}_k$ approaches the share of firms ϕ_k as ψ_{ok} all becomes arbitrarily large, i.e., as the overhead cost schedule shifts up. The convex overhead cost is the reason why the planner does not allocate all products to a particular type of firm with the highest combination of step size and process efficiency.

In the planner's problem, labor input is the same across all products. In each point in time the planner will set $l_t(i, j)$ to maximize static output Y_t that is given by

$$Y_t = Q_t \exp\left(\int_0^1 \log\left[a(j) l_t(i, j)\right] di\right), \tag{A.24}$$

and the planner is maximizing this subject to the constraint $\int_0^1 l_t(i, j) di = L$. Because of the Cobb-Douglas structure, it is optimal to set $l_t(i, j) = L$. Hence, the employment share of each type coincides with the product share of each type of firms. Let L_k/L denote the employment share of type k firms. The planner's solution features

$$L_k/L = \bar{S}_k.$$

A.2 Unconstrained planner's problem

The constrained planner solves the problem in two steps as production can be decoupled from innovations. It will have only the high step size firms innovate and the growth rate is

$$1 + \bar{g} = \gamma_B^{\frac{Z}{\psi_z}}.$$

Maximizing process efficiency means having the same amount of labor in each line and choosing $\bar{n}_k$ to solve

$$\max_{\bar{n}_k} \quad \sum_k J\phi_k \bar{n}_k \log(a_k) + \log\left(1 - \sum_k \frac{\tilde{\psi}_{ok}}{2} J\phi_k \bar{n}_k^2\right) \quad s.t. \quad \sum_k J\phi_k \bar{n}_k = 1.$$

The first-order condition implies that the difference in size between two producing firms depends only on the overhead costs and the process efficiency differences

$$\tilde{\psi}_{ok}\bar{n}_k - \tilde{\psi}_{ok'}\bar{n}_{k'} = (1 - \bar{o}) \log\left(\frac{a_k}{a_{k'}}\right). \tag{A.25}$$

The solution to the unconstrained planner's problem is then

$$\bar{S}_k = \phi_k^I \frac{\bar{\tilde{\psi}}_o^I}{\tilde{\psi}_{ok}} \left(1 + \frac{\nu_k}{\bar{\tilde{\psi}}_o^I} \frac{\sqrt{1 + \left(2J\phi^I - \sum_{k\in I} \phi_k^I \tilde{\psi}_{ok} \left(\frac{\bar{\tilde{\psi}}_o^I}{\tilde{\psi}_{ok}}\right)^2\right) \sum_{k\in I} \phi_k^I \nu_k^2/\tilde{\psi}_{ok} - 1}}{\sum_{k\in I} \phi_k^I \nu_k^2/\tilde{\psi}_{ok}} \right), \tag{A.26}$$

where

$$\nu_k \equiv \log(a_k) - \sum_{k'\in I} \phi_{k'}^I \frac{\bar{\tilde{\psi}}_o^I}{\tilde{\psi}_{ok'}} \log(a_{k'}).$$

The unconstrained planner may let firms with lower production efficiency to produce to save overhead cost.

A.3 Decentralized equilibrium

Along a balanced growth path $\frac{C}{Y}$, $\frac{w_z}{Y}$ and $\frac{Q}{Y}$ are constant over time and are functions of the $\bar{n}_k$. We have the following 9 equations in 9 unknowns (4 $\bar{n}_k$, 4 $\bar{S}_k$, $\left(\frac{\bar{w}_z}{Y}\right) \equiv \frac{w_z}{Y}$)

$$\psi_z\left(\frac{\bar{w}_z}{Y}\right)\left(\frac{1}{\beta}-1+\frac{Z}{\psi_z}\right) \geqslant 1-\frac{a_L}{\gamma_k a_k}\left(\bar{S}\Delta+1-\bar{S}\right)-\tilde{\psi}_o\bar{n}_k, \quad \text{hold with equality if } \bar{n}_k > 0 \tag{A.27}$$

where $\bar{S} \equiv \bar{S}_{HS} + \bar{S}_{HB}$

$$\sum_k J\phi_k\bar{n}_k = 1, \quad \bar{n}_k \geqslant 0 \tag{A.28}$$

$$\bar{S}_k = J\phi_k\bar{n}_k. \tag{A.29}$$

Let I be the set of firms that produce and $\phi_k^I \equiv \phi_k / \sum_{k\in I}\phi_k$ denote the share of innovating firms that are type k. Also let $\phi^I \equiv \sum_{k\in I}$ denote the share of innovating firms and $J^I \equiv J\phi^I$ the number of innovating firms. For $k \in I$, equations (A.27) and (A.29) imply

$$\bar{n}_k = \frac{1}{J^I} + \frac{\tilde{\omega}_k}{\tilde{\psi}_o}$$

where $\Delta = a_H/a_L$ and

$$\tilde{\omega}_k \equiv \frac{\bar{S}(\Delta-1)+1}{\gamma_S}\left(\left(\sum_{k'\in I}\phi_{k'}^I\frac{\gamma_S a_L}{\gamma_{k'}a_{k'}}\right)-\frac{\gamma_S a_L}{\gamma_k a_k}\right)$$

and

$$Z\left(\frac{\bar{w}_z}{Y}\right)\left(\frac{1}{\beta}-1+\frac{Z}{\psi_z}\right) = \frac{Z}{\psi_z}\left(1-\left(\bar{S}\Delta+1-\bar{S}\right)\left(\sum_{k\in I}\phi_k^I\frac{a_L}{\gamma_k a_k}\right)-\frac{\tilde{\psi}_o}{J^I}\right).$$

When all high process efficiency firms produce, we can show that

$$\frac{\bar{S}(\Delta-1)+1}{\tilde{\psi}_o/(J^I)} = \frac{(\Delta-1)(\phi^I_{HB}+\phi^I_{HS})+1}{\tilde{\psi}_o/(J^I) - \frac{\Delta-1}{\bar{S}(\Delta-1)+1}(\phi^I_{HB}\tilde{\omega}_{HB}+\phi^I_{HS}\tilde{\omega}_{HS})}$$

which implies that $\bar{S}_k (\equiv J\phi_k \bar{n}_k = J^I \phi^I_k \bar{n}_k)$ for innovating firms is equal to

$$\phi^I_k \left(1 + \frac{\left(\left(\sum_{k'\in I}\phi^I_{k'}\frac{\gamma_S a_L}{\gamma_{k'}a_{k'}}\right) - \frac{\gamma_S a_L}{\gamma_k a_k}\right)((\Delta-1)(\phi^I_{HB}+\phi^I_{HS})+1)}{\gamma_S\frac{\tilde{\psi}_o}{J^I} - (\Delta-1)\left((\phi^I_{HB}+\phi^I_{HS})\left(\sum_{k'\in I}\phi^I_{k'}\frac{\gamma_S a_L}{\gamma_{k'}a_{k'}}\right) - \phi^I_{HB}\frac{\gamma_S a_L}{\gamma_B a_H} - \phi^I_{HS}\frac{\gamma_S a_L}{\gamma_S a_H}\right)}\right). \tag{A.30}$$

Having solved for $\bar{n}_k$, we can substitute this into (A.27) for a firm that is producing and solve for $\left(\frac{\bar{w}_z}{Y}\right)$. Then we need to check that (A.27) holds for firms not producing. That is, for k where $\bar{n}_k = 0$, we have (A.27) holding with strict inequality.

More generally, some high process efficiency firms may not produce. Let H denote the set of high process efficiency firms. $\bar{S} = \sum_{k\in I\cap H} J^I\phi^I_k\bar{n}_k$ satisfies

$$\frac{\bar{S}(\Delta-1)+1}{\tilde{\psi}_o/J^I} = \frac{(\Delta-1)(\sum_{k\in I\cap H}\phi^I_k)+1}{\tilde{\psi}_o/J^I - \frac{\Delta-1}{\bar{S}(\Delta-1)+1}(\sum_{k\in I\cap H}\phi^I_k\tilde{\omega}_k)}$$

and $\bar{S}_k (\equiv J\phi_k \bar{n}_k = J^I \phi^I_k \bar{n}_k)$ for innovating firms is equal to

$$\bar{S}_k = \phi^I_k \left(1 + \frac{\left(\left(\sum_{k'\in I}\phi^I_{k'}\frac{\gamma_S a_L}{\gamma_{k'}a_{k'}}\right) - \frac{\gamma_S a_L}{\gamma_k a_k}\right)((\Delta-1)(\sum_{k\in I\cap H}\phi^I_k)+1)}{\gamma_S\frac{\tilde{\psi}_o}{J^I} - (\Delta-1)\left((\sum_{k\in I\cap H}\phi^I_k)\left(\sum_{k'\in I}\phi^I_{k'}\frac{\gamma_S a_L}{\gamma_{k'}a_{k'}}\right) - \sum_{k\in I\cap H}\phi^I_k\frac{\gamma_S a_L}{\gamma_k a_k}\right)}\right) \tag{A.31}$$

We solve the equilibrium by calculating $\bar{S}$ and then $\bar{n}_k$ and $\left(\frac{\bar{w}_z}{Y}\right)$ for a guess of I. We will then verify the guess of I is correct in that (A.27) holds with strict inequality for firms not in I.

In the decentralized equilibrium, the relative labor share of two producing firms k and k' facing the same share of high process efficiency competitors $\bar{S}$ is given by

$$\frac{\lambda_k(\bar{S})}{\lambda_{k'}(\bar{S})} = \frac{\gamma_{k'}a_{k'}}{\gamma_k a_k}. \tag{A.32}$$

Since production wages are the same across firms and $\bar{S}_k$ is equal to the sales share of type-k firms, the relative total employment of all type k firms, L_k relative to total employment of all type k' firms $L_{k'}$ is given by

$$\frac{L_k}{L_{k'}} = \frac{\lambda_k(\bar{S})\bar{S}_k}{\lambda_{k'}(\bar{S})\bar{S}_{k'}} = \frac{\gamma_{k'}a_{k'}}{\gamma_k a_k}\frac{\bar{S}_k}{\bar{S}_{k'}}. \tag{A.33}$$

Relative employment share is lower than relative sales share of high $\gamma_k a_k$ firms because they charge a higher markup.

For producing firms, R&D spending relative to sales is proportional to $\frac{\psi_z \bar{x}_k}{\bar{n}_k}$. Aggregate R&D spending as a share of GDP is

$$Z \cdot \left(\frac{\bar{w}_z}{Y}\right) = \frac{Z/\psi_z}{\frac{1}{\beta} - 1 + \frac{Z}{\psi_z}} \left(1 - \frac{\bar{S}\Delta + 1 - \bar{S}}{\gamma_s}\left(\sum_{k \in I}\phi_k^I \frac{\gamma_S a_L}{\gamma_k a_k}\right) - \frac{\tilde{\psi}_o}{J^I}\right).$$

A.4 Compare planner's and decentralized BGP when $\gamma_k = \gamma$

In this case, and $1 + \bar{g} = \gamma^{\frac{Z}{\psi_z}}$ in both the planner's problem and the decentralized equilibrium. Define ϕ_H as the share of firms with a_H and $\bar{S}_H$ the share of products they produce. In the following, we will compare $\bar{S}_H$ in the planner's problem and decentralized equilibrium. We will use superscripts D and P to denote decentralized and planner's BGP values, respectively.

Planner's problem First

$$\nu_H = (1 - \phi_H)\log\Delta$$

$$\nu_L = -\phi_H \log\Delta$$

$$\sum_k \phi_k \nu_k^2 = (\log\Delta)^2 (1 - \phi_H)\phi_H$$

Substituting these into (A.21), we have

$$\bar{S}_H^P = \phi_H + \frac{\sqrt{1+(\frac{2J}{\psi_o}-1)(\log\Delta)^2(1-\phi_H)\phi_H}-1}{\log\Delta} \tag{A.34}$$

$$= \phi_H + \frac{(\frac{2J}{\psi_o}-1)(\log\Delta)(1-\phi_H)\phi_H}{\sqrt{1+(\frac{2J}{\psi_o}-1)(\log\Delta)^2(1-\phi_H)\phi_H}+1} \tag{A.35}$$

Decentralized equilibrium Equation (A.31) reduces to

$$\bar{S}_H^D = \phi_H + \frac{\phi_H(1-\phi_H)[(\Delta-1)\phi_H+1]}{\frac{\psi_o}{J}\frac{\gamma\Delta}{\Delta-1}-(\Delta-1)\phi_H(1-\phi_H)}. \tag{A.36}$$

The employment share of the high efficiency firms is

$$\bar{E}_H^D = \frac{\bar{S}_H^D}{\Delta-(\Delta-1)\bar{S}_H^D} = \bar{S}_H^D - \frac{\bar{S}_H^D(1-\bar{S}_H^D)(\Delta-1)}{\Delta-(\Delta-1)\bar{S}_H^D}. \tag{A.37}$$

Combining the formulas for $\bar{S}_H^P$ and $\bar{S}_H^D$, we have

$$\frac{\bar{S}_H^P-\phi_H}{\bar{S}_H^D-\phi_H} = \frac{(\frac{2J}{\psi_o}-1)(\log\Delta)\frac{\frac{\psi_o}{J}\frac{\gamma\Delta}{\Delta-1}-(\Delta-1)\phi_H(1-\phi_H)}{(\Delta-1)\phi_H+1}}{\sqrt{1+(\frac{2J}{\psi_o}-1)(\log\Delta)^2(1-\phi_H)\phi_H}+1}. \tag{A.38}$$

Substituting in $\bar{E}_H^P = \bar{S}_H^P$ and (A.37), we have

$$\bar{E}_H^P - \bar{E}_H^D = \bar{S}_H^P - \bar{S}_H^D + \frac{\bar{S}_H^D(1-\bar{S}_H^D)(\Delta-1)}{\Delta-(\Delta-1)\bar{S}_H^D}. \tag{A.39}$$

A.5 Compare planner's and decentralized BGP when $a_k = a$

This is equivalent to setting $\Delta = 1$. Define ϕ_H as the share of firms with γ_B and $\bar{S}_B$ the share of products they produce. Also, let Γ denote the step size gap $\frac{\gamma_B}{\gamma_S}$ For both the planner and decentralized equilibrium, the mean step size is given by

$$\bar{\gamma} = \Gamma^{\bar{S}_B}\gamma_S.$$

Hence, the growth rate increases with the share of products produced by firms with the big step size. In the following, we will compare $\bar{S}_B$ in the planner's problem and decentralized equilibrium. We will use superscripts D and P to denote decentralized and planner's BGP values, respectively.

Planner's problem First

$$\nu_B = \frac{1/\beta - 1 + Z/\psi_z}{1/\beta - 1}(1-\phi_B)\log\Gamma$$

$$\nu_S = -\frac{1/\beta - 1 + Z/\psi_z}{1/\beta - 1}\phi_B\log\Gamma$$

$$\sum_k \phi_k \nu_k^2 = \left(\frac{1/\beta - 1 + Z/\psi_z}{1/\beta - 1}\log\Gamma\right)^2(1-\phi_B)\phi_B$$

Substituting these into (A.21), we have

$$\bar{S}_B^P = \phi_B + \frac{\sqrt{1+(\frac{2J}{\psi_o}-1)\left(\frac{1/\beta-1+Z/\psi_z}{1/\beta-1}\log\Gamma\right)^2(1-\phi_B)\phi_B} - 1}{\frac{1/\beta-1+Z/\psi_z}{1/\beta-1}\log\Gamma} \tag{A.40}$$

$$= \phi_B + \frac{(\frac{2J}{\psi_o}-1)\left(\frac{1/\beta-1+Z/\psi_z}{1/\beta-1}\log\Gamma\right)(1-\phi_B)\phi_B}{\sqrt{1+(\frac{2J}{\psi_o}-1)\left(\frac{1/\beta-1+Z/\psi_z}{1/\beta-1}\log\Gamma\right)^2(1-\phi_B)\phi_B} + 1} \tag{A.41}$$

Decentralized equilibrium Equation (A.31) reduces to

$$\bar{S}_B^D = \phi_B + \phi_B(1-\phi_B)\frac{\Gamma-1}{\gamma_B}\frac{J}{\psi_o}. \tag{A.42}$$

The employment share of the big step size firms is

$$\bar{E}_B^D = \frac{\bar{S}_B^D}{\Gamma-(\Gamma-1)\bar{S}_B^D} = \bar{S}_B^D - \frac{\bar{S}_B^D(1-\bar{S}_B^D)(\Gamma-1)}{\Gamma-(\Gamma-1)\bar{S}_B^D}. \tag{A.43}$$

Combining the formulas for $\bar{S}_B^P$ and $\bar{S}_B^D$, we have

$$\frac{\bar{S}_B^P - \phi_B}{\bar{S}_B^D - \phi_B} = \frac{(\frac{2J}{\psi_o} - 1)\left(\frac{1/\beta - 1 + Z/\psi_z}{1/\beta - 1}\log\Gamma\right)\frac{\gamma_B}{\Gamma - 1}\frac{\psi_o}{J}}{\sqrt{1 + (\frac{2J}{\psi_o} - 1)\left(\frac{1/\beta - 1 + Z/\psi_z}{1/\beta - 1}\log\Gamma\right)^2 (1 - \phi_B)\phi_B} + 1} \tag{A.44}$$

Substituting in $\bar{E}_H^P = \bar{S}_H^P$ and (A.43), we have

$$\bar{E}_B^P - \bar{E}_B^D = \bar{S}_B^P - \bar{S}_B^D + \frac{\bar{S}_B^D(1 - \bar{S}_B^D)(\Gamma - 1)}{\Gamma - (\Gamma - 1)\bar{S}_B^D} \tag{A.45}$$

A.6 Compare planner's and decentralized BGP when $a_k\gamma_k = a_{k'}\gamma_{k'}$, for all k, k'

This is possible when $\phi_{HB} = \phi_{LS} = 0$ and $\Delta = \Gamma$, so all firms have the same markups. In the following, we will compare $\bar{S}_{HS}$ in the planner's problem and decentralized equilibrium. We will use superscripts D and P to denote decentralized and planner's BGP values, respectively.

Planner's problem First

$$\nu_{HS} = -(1 - \phi_{HS})\frac{Z/\psi_z}{1/\beta - 1}\log\Gamma$$

$$\nu_{LB} = \phi_{HS}\frac{Z/\psi_z}{1/\beta - 1}\log\Gamma$$

$$\sum_k \phi_k \nu_k^2 = \phi_{HS}(1 - \phi_{HS})(\log\Gamma)^2\left(\frac{Z/\psi_z}{1/\beta - 1}\right)^2$$

Substituting these into (A.21), we have

$$\bar{S}^P_{HS} = \phi_{HS} - \frac{\sqrt{1+(\frac{2J}{\psi_o}-1)\phi_{HS}(1-\phi_{HS})(\log\Gamma)^2(\frac{Z/\psi_z}{1/\beta-1})^2}-1}{\frac{Z/\psi_z}{1/\beta-1}\log\Gamma} \tag{A.46}$$

$$= \phi_{HS} - \frac{(\frac{2J}{\psi_o}-1)\,(\log\Gamma)\,(\frac{Z/\psi_z}{1/\beta-1})(1-\phi_{HS})\phi_{HS}}{\sqrt{1+(\frac{2J}{\psi_o}-1)\phi_{HS}(1-\phi_{HS})(\log\Gamma)^2(\frac{Z/\psi_z}{1/\beta-1})^2}+1} \tag{A.47}$$

We can also substitute ν_{HS} and ν_{LB} into (A.18) to get

$$\bar{n}^P_{LB} = \frac{1}{J} + \frac{1-\bar{o}}{\psi_o}\phi_{HS}\frac{Z/\psi_z}{1/\beta-1}\log\Gamma$$

$$\bar{n}^P_{HS} = \frac{1}{J} - \frac{1-\bar{o}}{\psi_o}(1-\phi_{HS})\frac{Z/\psi_z}{1/\beta-1}\log\Gamma$$

Combining these yields

$$\bar{n}^P_{LB} - \bar{n}^P_{HS} = \frac{1-\bar{o}}{\psi_o}\frac{Z/\psi_z}{1/\beta-1}\log\Gamma$$

The planner makes big steps firm bigger because of the greater spillover from product innovation.

Decentralized equilibrium Note that

$$\left(\left(\sum_{k'}\phi_{k'}\frac{\gamma_S a_L}{\gamma_{k'}a_{k'}}\right) - \frac{\gamma_S a_L}{\gamma_k a_k}\right) = 0$$

So, equation (A.31) reduces to

$$\bar{S}^D_k = \phi_k, \tag{A.48}$$

and the employment share of either firm type is

$$\bar{E}^D_k = \bar{S}^D_k. \tag{A.49}$$

Since all firms have the same markups in this case, they all have the same market and employment share. So for each firm type, their market share and employment share is equal to the fraction of firms of that type. Combining the formulas for $\bar{S}^P_{HS}$ and $\bar{S}^D_{HS}$, we have

$$\bar{S}^P_{HS} - \bar{S}^D_{HS} = -\frac{(\frac{2J}{\psi_o} - 1)\,(\log\Gamma)\,(\frac{Z/\psi_z}{1/\beta-1})(1-\phi_{HS})\phi_{HS}}{\sqrt{1 + (\frac{2J}{\psi_o} - 1)\phi_{HS}(1-\phi_{HS})(\log\Gamma)^2(\frac{Z/\psi_z}{1/\beta-1})^2} + 1}. \tag{A.50}$$

Substituting in $\bar{E}^P_H = \bar{S}^P_H$ and (A.49), we also have

$$\bar{E}^P_H - \bar{E}^D_H = -\frac{(\frac{2J}{\psi_o} - 1)\,(\log\Gamma)\,(\frac{Z/\psi_z}{1/\beta-1})(1-\phi_{HS})\phi_{HS}}{\sqrt{1 + (\frac{2J}{\psi_o} - 1)\phi_{HS}(1-\phi_{HS})(\log\Gamma)^2(\frac{Z/\psi_z}{1/\beta-1})^2} + 1}. \tag{A.51}$$

A.7 Decentralized equilibrium with tax policies

The tax instruments are business income flat tax rate τ and R&D subsidy rates $\underline{\tau}_{RD}$ and $\overline{\tau}_{RD}$. Rate $\underline{\tau}_{RD}$ applies for R&D expenditures below $\underline{RD} \times Y$ while rate $\overline{\tau}_{RD}$ applies for expenditures in excess of $\underline{RD} \times Y$

Post tax period profit relative to Y for intermediate producer k is then

$$\begin{aligned}
& (1-\tau)\left(\pi_k(n,h) - \frac{\psi_z w_z}{Y}x\right) \quad \text{[post income tax]} \\
+ \quad & \underline{\tau}_{RD}\ \min\left\{\frac{\psi_z w_z}{Y}x,\ \underline{RD}\right\} \quad \text{[subsidy for R\&D below threshold]} \\
+ \quad & \overline{\tau}_{RD}\ \max\left\{\frac{\psi_z w_z}{Y}x - \underline{RD},\ 0\right\} \quad \text{[subsidy for R\&D in excess threshold]}
\end{aligned}$$

As before, we have 9 equations in 9 unknowns (4 $\bar{n}_k$, 4 $\bar{S}_k$, $(\frac{\bar{w}_z}{Y}) \equiv \frac{w_z}{Y}$) but the first-order conditions change to

$$\psi_z\left(\frac{\bar{w}_z}{Y}\right)\left(\frac{1}{\beta} - 1 + \frac{Z}{\psi_z}\right)\frac{1-\tau-\tau_{RD}(Z/\psi_z\bar{n}_k)}{1-\tau} \geqslant 1 - \frac{a_L}{\gamma_k a_k}(\bar{S}\Delta + 1 - \bar{S}) - \tilde{\psi}_o\bar{n}_k, \quad \text{with "=" if } \bar{n}_k > 0.$$

where $\tau_{RD}(x)$ denotes the marginal subsidy rate for a firm k that spends R&D to draw x lines

$$\tau_{RD}(x) = \begin{cases} \underline{\tau}_{RD} & \text{if } \frac{\psi_z w_z}{Y} x < \underline{RD} \\ \overline{\tau}_{RD} & \text{if } \frac{\psi_z w_z}{Y} x > \underline{RD}. \end{cases} \tag{A.52}$$

Aggregate R&D expenditure as a share of total output becomes

$$\left(\frac{\bar{w}_z}{Y}\right) Z = \frac{Z}{\psi_z} \frac{(1 - \frac{\bar{S}\Delta + 1 - \bar{S}}{\gamma_S} \sum_{k \in I} \phi_k \frac{\gamma_S a_L}{\gamma_k a_k} - \tilde{\psi}_o \sum_{k \in I} \phi_k \bar{n}_k)}{\left(\frac{1}{\beta} - 1 + \frac{Z}{\psi_z}\right) \sum_{k \in I} \phi_k (1 - \frac{\underline{\tau}_{RD}(Z/\psi_z \bar{n}_k)}{1-\tau})} \tag{A.53}$$

Note that if the marginal R&D subsidy is the same for all firms in the equilibrium, the subsidy does not affect the equilibrium firm size distribution. It only raises equilibrium w_z and pretax R&D expenditure. The post tax R&D expenditure is unaffected.

The net tax revenue T is the difference between revenue from the business income tax and the subsidy. We assume T is rebated lump sum to the household. The household's problem is the same as before but with a modified budget constraint

$$A_{t+1} = A_t(1 + r_t) + w_t L + w_{z,t} Z + T_t - C_t$$

Since government's net revenue is fully rebated to the household, the resource constraints are the same as before.

In the France, $\tau = 0.33, \underline{\tau}_{RD} = 0.3, \overline{\tau}_{RD} = 0.05$. The threshold 100 million Euros is 11.96 times average firm sales. Since average sales is Y/J in the model, the R&D subsidy cutoff relative to Y is

$$\underline{\text{RD}} = \frac{11.96}{J^I}$$

The J^I term appears in the denominator of the cutoff and firm-level R&D spending $\psi_z \left(\frac{\bar{w}_z}{Y}\right) x_k = \left(\frac{\bar{w}_z}{Y}\right) Z \frac{S_k}{\phi_k^I J^I}$. Therefore, marginal tax rate realized at the equilibrium and other equilibrium variables do not depend on J^I

$$\tau_{RD}(x(S_k^*)) = \begin{cases} \underline{\tau}_{RD} & \text{if } \frac{S_k^*}{\phi_k^I}\frac{w_z Z}{Y} < 11.96 \\ \overline{\tau}_{RD} & \text{if } \frac{S_k^*}{\phi_k^I}\frac{w_z Z}{Y} > 11.96. \end{cases} \tag{A.54}$$

B Data

In this section we present how we construct our main variables by merging various firm and product level data on the French manufacturing sector. We first list the datasets used and then describe the procedures we used.

B.1 Data sources

FARE FARE (“*Fichier Approché des Résultats d’ESANE*”) is a dataset of individual accounting data of French companies. It uses administrative data obtained from the annual profit declarations that companies make to the tax authorities and from the annual social data that provide information on employees. FARE covers all business firms actively participating in the productive system but exclude firms in the financial sector (except holdings and firms in the sector: Activities auxiliary to financial services and insurance activities) and farming businesses (except logging activities). FARE is an annual panel dataset which covers the period 2009-2019 and where firms are identified by their standard administrative identifier known as the “siren” number. We obtain the following variables from FARE.

Gross output: the sum of the gross value of products sold, stored production and operating subsidy

Intermediates: operating subsidy, material purchased, external expenses (including outsourcing) and change in material

Value added: the difference between gross output and intermediate inputs

Wagebill: the total wagebill declared by the firm including salary, bonuses, paid leave and other compensations received by workers during the year

Capital Stock: the stock of non-financial tangible assets at current value. The values reported in the balance sheet are not given at the current prices but at the price of acquisition. Nevertheless, firm's balance sheets provide information on both the gross acquisition value of fixed assets and on the accumulated depreciated value of each assets. A proxy for the mean age of each asset can be recovered thanks to the ratio of the accumulated depreciated of this asset over its gross book value (assuming that the assets are linearly amortized) with an assumption of the life expectancy of this asset. We use a value of 10 years for tangible capital.

Formally, let $K_{t,t_0}P_{t_0}$ denotes the gross book value reported in year t of a volume K of capital acquired in year t_0 and D_t the reported accumulated depreciation, then the average age is estimated as:

$$a_t = \frac{\min\left[D_t, K_{t,t_0}P_{t_0}\right]}{K_{t,t_0}P_{t_0}} d,$$

where d is the theoretical life expectancy. From it we can estimate t_0 and use the corresponding national price index for this asset class to calculate $K_{t,t_0}P_t$. We then deduce depreciation to estimate the stock of net capital.

For firms that rent or lease their real-estate, we use the rental cost reported in the balance sheet and a user cost of $r_K = 15\%$ to impute a value for the stock.

Industry: two digit NACE industry in manufacturing

DADS DADS ("*Déclarations Annuelles des Données Sociales*") are mandatory declarative formalities that must be completed by any company with paid employees. In these document which is shared with the social security administrator, all firms provide information about the establishment and its employees on an annual basis. Each entry of this dataset corresponds to a specific employee's position in a plant with details on wage, working condition, age, sex, occupation etc.

Hours: firm level hours by summing the hours of each position declared in the contract of the employees

Overhead: wage bill of employees in occupations not directly associated with production, management, or R&D—primarily encompassing support functions such as human resources, accounting, and legal services

EAP EAP "Enquête Annuelle de Production" is an annual survey that covers the manufacturing sector (with the exception of the food manufacturing industry) and gives details of the breakdown of production across products on about 35,000 firms. The survey is comprehensive for firms larger than 20 employees and is a stratified sample below this threshold. For each product and each firm, the EAP provides information on the quantity and total value sold. See Monin and Castillo (2020) for a discussion on the quality and reliability of this dataset.

Quantity: quantity are given in various units, we harmonized these units within a product (e.g. convert tons to kilograms, dozens to units, mega watts to giga watts etc...)

Value sold: the value is directly given in euros and correspond to the total sales of the corresponding product during the year

Product class: The product class "PRODFRA" contains more than 4000 different products and whose first 8 digit correspond to the European classification PRODCOM.

B.2 Data cleaning

FARE

We keep firms with a standard business tax scheme ("*Régimes du bénéfice réel normal*"). This essentially excludes small firms with a simplified accounting declaration procedure. We also remove firms that are not corporate businesses ("*Société commerciale*"). We then drop firms with a negative capital stock and a gross output of 0 or lower.

DADS

We remove contracts with 0 hours worked or a gross wage of 0. Note that temporary workers are allocated to their employer (the temporary agency) and not to the firm where they work.

EAP

The product classification within the EAP undergoes minor modifications over time. To establish a consistent classification framework, we adopt the procedure outlined in Monin and Castillo (2020). This involves utilizing yearly crosswalks to formulate the convex envelope for a set of products in case they are impacted by the changes in the classification (hence, if for example in 2012 a product category A disappear and is split into a product category B and another product category C in 2013, we would merge A, B and C into a single category). As a result, the distinct product count in our sample is reduced from 4,444 to 3,824.

We keep only products manufactured, designed and sold by the company and on the national territory or through subcontracting, (models 2 and 3 in the EAP nomenclature). We drop observations with a quantity of 0 (or missing information) and observations with a non positive value.

As described below, we use the EAP to calculate unit price. This unit price is then winsorized at the 2.5% threshold (both at the top and bottom of the distribution).

FARE-DADS sample

We first merge FARE and DADS at the firm-year level using the period 2012-2019. Over this period, we use the consolidated definition of a firm from the INSEE ("*entreprises profilées*"). This changes has been made to better account for the fact that large groups are typically composed of several, sometimes hundreds, of legal entities. Merging DADS and FARE leaves out a number of firms, namely those that are not in the private corporate sector but also firms with no paid workers. The resulting dataset counts 925,765 unique firms with an average of more than 500,000 observations per year.

We then proceed to clean the dataset by removing firms with extreme value for several ratios. Specifically, we remove observations that are in the top 0.5% or bottom 0.5% of the distribution of (i) wage over capital, (ii) employment over capital, (iii) hours over employment and (iv) wagebill over value added.

The cost shares of capital are calculated for each firm and year and the corresponding values of TFPR is then computed. We trim the top and bottom 1% of the distribution of TFPR and remove observations for which we cannot calculate TFPR (with missing or negative values) and then recompute the cost shares on the remaining firms.

Table B.1: Data cleaning (2012–2019)

Sample	Observations	Unique firms	Total VA
1. Manufacturing businesses in both FARE and DADS	401,577	66,675	221b
2. With TFPR	393,590	65,757	214b
3. Merged to EAP	110,711	19,792	160b
4. Regressions	75,904	13,942	147b

Notes: Total value added is taken in 2019.

FARE-DADS-EAP sample

We keep only firms in industries NACE (rev 2) 13 to 33, excluding industry 19 due to a too small number of firms. We also restrict to firms that are larger than

Table B.2: Coverage by manufacturing industries

Industry	National Accounts	FARE-DADS merged	With TFPR	Merged with EAP	Regression
10	39.4	28.1	25.7	7.4	0
11	6.7	6.7	4.7	0.7	0
12	0	0	0	0	0
13	2.0	1.8	1.7	1.5	1.4
14	1.8	1.7	1.7	1.5	1.5
15	2.1	2.6	2.6	2.4	0.5
16	3.3	3.1	3.0	2.1	2.0
17	4.5	4.8	4.7	4.5	4.5
18	3.7	2.4	2.3	0.4	0.4
20	20.9	21.2	21.0	20.6	20.6
21	12.8	11.6	11.3	7.2	7.2
22	11.5	12.9	12.8	11.9	11.8
23	8.6	8.8	8.7	8.3	8.1
24	4.7	4.6	4.5	3.9	3.9
25	20.6	18.4	17.9	10.3	10.0
26	13.3	6.6	6.5	5.6	5.6
27	7.2	9.7	9.6	9.4	9.3
28	12.3	15.3	15.1	14.2	14.0
29	13.7	15.7	15.3	14.8	14.7
30	20.6	23.8	23.8	21.0	21.0
31	2.2	1.9	1.9	1.8	1.7
32	5.1	5.0	4.9	3.2	3.2
33	24.8	12.2	12.0	5.7	5.6
Total	242	221	214	160	147

Notes: Values correspond to total value added in billion of euros and are measured in 2019. Industry code corresponds to the first two digit of the NACE classification.

5 full time equivalent employees. Table B.1 summarizes the change made to every step of the sample selection and B.2 shows coverage of manufacturing industries compared to national accounts. As shown in Table B.2, our final regression sample represents 60% of the whole manufacturing sector and 75% of the manufacturing sector excluding food, beverage, and tobacco (code NACE 10, 11, and 12).

B.3 Calculation of main variables

Table B.3 displays the raw data variables we use to construct price, TFPR and TFPQ. Throughout we use subscript j to index firms, t to index data year and $s(j)$ to denote the industry firm j belongs to.

Table B.3: Raw data variables

Notation	Definition	Source
$P^Q_{jt} \cdot Q_{jt}$	Nominal gross output	FARE
$P^K_{jt} \cdot K_{jt}$	Nominal capital stock (at current value) for non-financial tangible assets	FARE
W_{jt}	Nominal payroll	FARE
$P^X_{jt} \cdot X_{jt}$	Nominal intermediate input	FARE
H_{jt}	Hours	DADS
$P^{KLEMS}_{K,s(j),t}$	Industry level capital price deflator	KLEMS
$h^{KLEMS}_{s(j),t}$	Industry level labor composition/quality index	KLEMS
$y(i,j,t)$	Quantity of product i by firm j in year t	EAP
$p(i,j,t) \cdot y(i,j,t)$	Nominal sales of product i by firm j in year t	EAP

B.4 Variables used for cross-sectional facts

Next, we describe how we calculate the the firm-level price, TFPR, and TFPQ used for the cross-sectional facts. Define the unit price of good i sold by firm j in period t as

$$p(i,j,t) \equiv \frac{p(i,j,t) \cdot y(i,j,t)}{y(i,j,t)}$$

Also, define real capital used by firm j in period t as

$$\widetilde{K}_{jt} \equiv \frac{P^K_{jt} \cdot K_{jt}}{P^{KLEMS}_{K,s(j),t}}$$

and quality adjusted labor input

$$\widetilde{H}_{jt} \equiv h^{KLEMS}_{s(j),t} \cdot H_{jt}$$

Firm level price First we standardize unit price with a product category. Let $\mathcal{N}_{i,t}$ be the set of firms that sell product i in year t. The sales share of firm j in product category i is

$$\kappa(i,j,t) = \frac{p(i,j,t) \cdot y(i,j,t)}{\sum_{j' \in \mathcal{N}_{i,t}} p(i,j',t) \cdot y(i,j',t)}$$

The average price in product category i is sales-weighted geometric average of prices in the category across firms that sell the product

$$P(i,t) = \prod_{j \in \mathcal{N}_{i,t}} p(i,j,t)^{\kappa(i,j,t)}$$

The standardized price is unit price divided by average product category price

$$\hat{p}(i,j,t) = \frac{p(i,j,t)}{P(i,t)}$$

The standardized firm level price is the sales-weighted geometric average of prices sold by the firm

$$\hat{p}(j,t) = \prod_{i \in \mathcal{S}_{j,t}} \hat{p}(i,j,t)^{\omega(i,j,t)}$$

where $\mathcal{S}_{j,t}$ is the set of products sold by firm j in period t and

$$\omega(i,j,t) = \frac{p(i,j,t) \cdot y(i,j,t)}{\sum_{i' \in \mathcal{S}_{j,t}} p(i',j,t) \cdot y(i',j,t)}$$

Firm level TFPR and TFPQ Define factor cost shares as

$$\alpha_{s,t} = \frac{\sum_{j \in s} R \cdot (P^K_{jt} \cdot K_{jt})}{\sum_{j \in s} \left(R \cdot (P^K_{jt} \cdot K_{jt}) + W_{jt} \right)}$$

and

$$\theta_{s,t} = \frac{\sum_{j\in s}\left(R\cdot(P^K_{jt}\cdot K_{jt}) + W_{jt}\right)}{\sum_{j\in s}\left(R\cdot(P^K_{jt}\cdot K_{jt}) + W_{jt} + P^X_{jt}\cdot X_{jt}\right)}$$

where we assume $R = 15\%$.

To measure TFPR, we divide nominal gross output by a a weighted average of nominal capital, wagebill and intermediate:

$$TFPR_{jt} = \frac{P^Q_{jt}\cdot Q_{jt}}{(\tilde{K}_{jt})^{\alpha_{s,t}\theta_{s,t}}(W_{jt})^{(1-\alpha_{s,t})\theta_{s,t}}(\tilde{X}_{jt})^{(1-\theta_{s,t})}} \tag{B.1}$$

TFPQ is TFPR divided by standardized firm level price

$$TFPQ_{jt} = \frac{TFPR_{jt}}{\hat{p}(j,t)}$$

We use the wage bill to account for potential differences in labor quality across firms. We do not need to use price deflators because these are defined at the sector-year level and will be netted out when we control for industry-year fixed effects in a log/log regression. In the next section, where we calculate TFPQ growth, we will use hours to construct TFPQ growth and account for labor-quality growth.

B.5 Formula for calculating TFPQ growth

First we compute factor cost shares as Tornqvist averages of two consecutive periods

$$\bar{\alpha}_{s,t} = \frac{\alpha_{s,t}+\alpha_{s,t-1}}{2} \quad \text{and} \quad \bar{\theta}_{s,t} = \frac{\theta_{s,t}+\theta_{s,t-1}}{2}$$

where α and θ are defined above. Let $TFPR^g_{jt}$ be the measure of TFPR we use for constructing TFPQ growth.

$$TFPR^g_{jt} = \frac{P^Q_{jt}\cdot Q_{jt}}{(\tilde{K}_{jt})^{\bar{\alpha}_{s,t}\bar{\theta}_{s,t}}(\tilde{H}_{jt})^{(1-\bar{\alpha}_{s,t})\bar{\theta}_{s,t}}(\tilde{X}_{jt})^{(1-\bar{\theta}_{s,t})}} \tag{B.2}$$

This is distinct from the TFPR measure we want to use for the cross-section. Here we do not subtract industry-year means and we Tornqvist the cost shares.

The growth rate is

$$g_{jt}^{TFPR} = \ln(TFPR_{jt}^{g}) - \ln(TFPR_{j,t-1}^{g})$$

Let $\mathcal{C}_{j,t,t-1}$ be the set of products that firm j sells in both period t and $t-1$, we define the Tornqvist sales share of products within the continuing set in t and $t-1$ as:

$$\bar{\omega}(i,j,t) \equiv \frac{1}{2}\left(\frac{p(i,j,t)\cdot y(i,j,t)}{\sum_{i'\in\mathcal{C}_{j,t,t-1}} p(i',j,t)\cdot y(i',j,t)} + \frac{p(i,j,t-1)\cdot y(i,j,t-1)}{\sum_{i'\in\mathcal{C}_{j,t,t-1}} p(i',j,t-1)\cdot y(i',j,t-1)}\right)$$

And the price growth for firm j as:

$$g_{j,t}^{p} = \sum_{i\in\mathcal{C}_{j,t,t-1}} \bar{\omega}(i,j,t)\left(\ln(p(i,j,t)) - \ln(p(i,j,t-1))\right)$$

The growth of TFPQ for firm j is

$$g_{jt}^{TFPQ} = g_{jt}^{TFPR} - g_{j,t}^{p}$$

The TFPQ growth in value-added and labor-augmenting form is

$$\tilde{g}_{jt}^{TFPQ} = \frac{g_{t}^{TFPQ}}{(1-\bar{\alpha}_{s(j)t})\bar{\theta}_{s(j),t}}$$

Aggregate TFPQ growth in factor neutral, gross-output terms is given by the sales-weighted firm level TFPQ growth rates

$$g_{t}^{TFPQ} = \sum_{j} \bar{\zeta}(j,t)\cdot g_{jt}^{TFPQ}$$

where

$$\bar{\zeta}(j,t) = \frac{\zeta(j,t) + \zeta(j,t-1)}{2}, \quad \zeta(j,t) = \frac{P^Q_{jt} \cdot Q_{jt}}{\sum_{j'} P^Q_{j't} \cdot Q_{j't}}$$

Note that this is the sales share of firm j amongst all firms (pooling industries).

Aggregate value-added TFPQ growth is given by the value added-weighted firm level value added TFPQ growth rates

$$\tilde{g}^{TFPQ}_t = \sum_j \bar{\xi}(j,t) \cdot \tilde{g}^{TFPQ}_{jt}$$

where

$$\bar{\xi}(j,t) = \frac{\xi(j,t) + \xi(j,t-1)}{2}, \quad \xi(j,t) = \frac{P^Q_{jt} \cdot Q_{jt} - P^X_{jt} \cdot X_{jt}}{\sum_{j'}(P^Q_{j't} \cdot Q_{j't} - P^X_{j't} \cdot X_{j't})}$$

B.6 TFPQ growth in the data

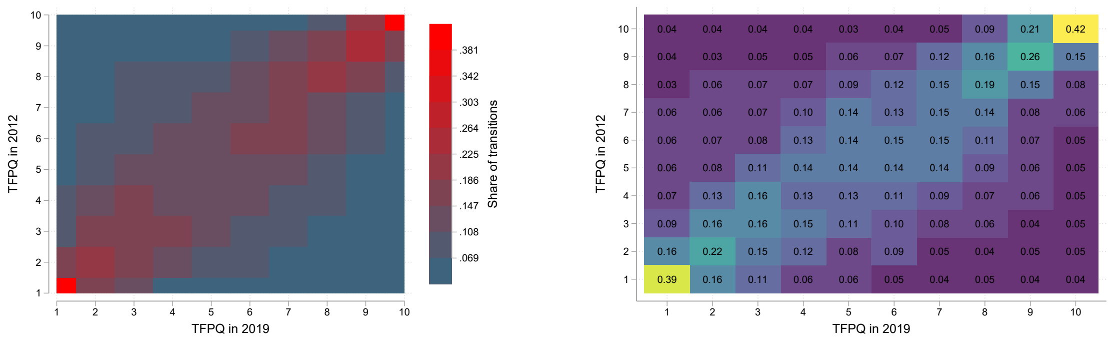
Figure B.1 displays the transition matrix where the horizontal axis is the firm's decile of TFPQ in 2019 and the vertical axis is the firm's decile in 2012. Each cell represents the share of firms in the final state given in the x-axis among all firms that started in the same initial state shown in the y-axis. Each row sums to 100%. Both groups are calculated within industry.

B.7 Additional empirical facts

Overhead not strongly related with Firm Size. While overhead costs are not directly observable in the data, we can approximate them using the detailed information available in the DADS dataset. Specifically, we consider the total wage bill of employees in occupations not directly associated with production, management, or R&D—primarily encompassing support functions such as human resources, accounting, and legal services. In addition, as many firms outsource these tasks to specialized service providers—a trend that has accelerated over time and is notably prevalent in France (Goldschmidt and

Figure B.1: Firm TFPQ transition

(a) Heatmap (b) Probabilities



Schmieder, 2017)—we augment our overhead proxy by including firms' expenditures on "external labor services," as reported in their balance sheets. Depending on whether this component is included, overhead accounts for approximately 13.5% to 15.6% of the total wage bill and around 3% of sales. With this proxy, we can look at the correlation of overhead with size as measured by total sales. Table B.4 shows the results. Columns 1 and 2 use the overhead as a share of total wagebill (not including then including outsourcing expenditures) and report a positive relationship with size. A doubling of the size increases overhead by 0.8 to 1 percentage point which is about 6-7% of the average value. Column 3 uses the ratio of total overhead over sales and report a small negative coefficient which is likely explain by a division bias. Hence, column (4) estimates the coefficient using log hours as an instrument and report a coefficient of 0.0012 which implies that a doubling of sales corresponds to a 0.008 pp increase of the ratio, or about 3% of the baseline value.

Ambiguous relationship between R&D and size. Another potential source of heterogeneity in the model is R&D efficiency (parameter ψ_Z). However, only few firms report positive R&D expenditures. For example in 2021, only 21,695 firms out of more than 3 millions did declare some R&D to receive a tax credit.

Table B.4: Regression of overhead cost share on size across firms

	(1)	(2)	(3)	(4)
Size (log sales)	0.0122 (0.0013)	0.0151 (0.0014)	-0.0009 (0.0003)	0.0012 (0.0003)
Avg. value of dep. var.	0.135	0.156	0.029	0.029

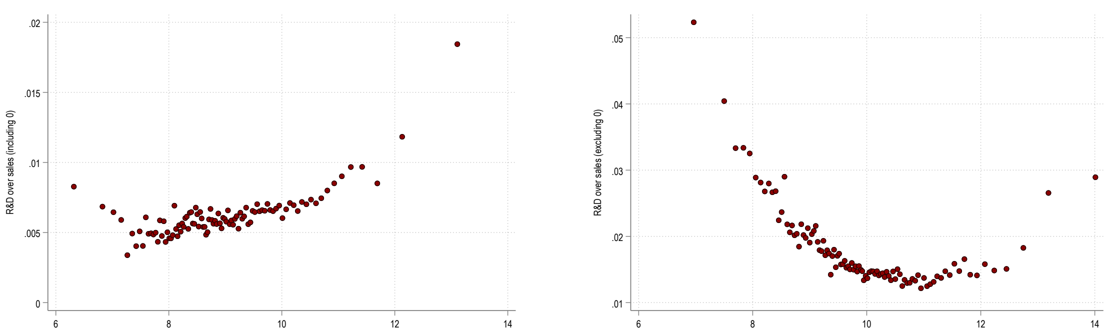
Notes: The level of observation is a firm j and year t. Columns 1, 2 and 3 estimate the parameters using an OLS estimator with sector-year fixed effects. Column 4 uses a 2SLS IV estimation instrumenting size by its lagged value. Size is measured using nominal sales $P^Q Q$ in the firm obtained directly from FARE (see Section 3.3.1). Standard errors are clustered by 2-digit sector. 84,839 observations. Columns differ by the dependent variable considered. Column 1 uses the share of wagebill acounted for by non production, non management and non R&D occupation. Column 2 includes outsourced labor serviced both in the numerator and the denominator of this ratio. Columns 3 and 4 use the same numerator as in column 1 but divide by sales.

In addition, the exact amount of R&D expenditures is not readily available in the balance sheet. We nevertheless proxy for it by using the wagebill of R&D workers following Bergeaud et al., 2022. Figure B.2 then looks at the correlation of R&D intensity (R&D divided by sales) with size, as measured by the logarithm of total hours worked. It shows that the correlation can either be positive or negative depending on whether we include firms that do not employ any R&D worker.

Figure B.2: R&D intensity vs. sales

(a) Including firms with no R&D

(b) Excluding firms with no R&D



Notes: Underlying data is French manufacturing firm-year observations from 2012–2019. R&D over sales is measured using the wagebill of R&D workers (following Bergeaud et al., 2022). Each dot aggregate over 840 observations (firm-year). Both Sales and R&D over sales have been residualized on a sector-year fixed effect.

B.8 Calibration targets

We describe how we compute the data targets used for calibration as reported in Table 2. In the following, $s(j,t)$ denotes the industry of a firm j during year t and the set of firms belonging to this industry during year t is $\int(s,t)$. There are S industries and N firms in the total economy and N_{st} firms in industry s during year t.

1. **Interquartile ratio of fitted TFPQ:** To mitigate potential measurement error in TFPQ and price, we construct fitted values based on a predictive regression using a 5-year lag and controlling for industry-year fixed effects. Specifically, we estimate the following regression:

$$\log(\text{TFPQ}_{j,t}) = \beta \log(\text{TFPQ}_{j,t-5}) + \gamma_{s(j,t),t} + \varepsilon_{j,t},$$

   where $s(j)$ denotes the 4-digit industry of firm j, and $\gamma_{s(j),t}$ are industry-year fixed effects. We defined Let $\log\widetilde{(\text{TFPQ})}_{j,t}$ as $\hat{\beta}\log(\text{TFPQ}_{j,t-5}) + \hat{\gamma}_{s(j),t}$ the fitted value from this regression. The first moment of interest, M^1, is the interquartile ratio of fitted TFPQ across firms:

$$M^{(1)} = \exp\left(Q_{0.75}(\log\widetilde{\text{TFPQ}}_{j,t}) - Q_{0.25}(\log\widetilde{\text{TFPQ}}_{j,t})\right),$$

   where $Q_{0.75}(\cdot)$ and $Q_{0.25}(\cdot)$ denote the empirical 75th and 25th percentiles, respectively.[2]

2. **Interquartile ratio of fitted price:** $M^{(2)}$ is calculated exactly like $M^{(1)}$ but using price $\hat{p}$ instead of TFPQ.

[2] According to Wooldridge (2010), one needs to correct the bias in the level of TFPQ when using the projected log TFPQ, i.e., the bias-corrected fitted TFPQ is $\widetilde{\text{TFPQ}}_{j,t} = \exp\left(\log\widetilde{(\text{TFPQ})}_{j,t} + \frac{1}{2}\widehat{\sigma}^2\right)$ where $\widehat{\sigma}^2$ is the estimated variance of the residuals. This correction does not affect the interquartile ratio since the σ^2 terms cancels out.

3. **Market share of firms above mean price:** the market share is defined as the ratio of nominal sales of a firm $P_{j,t}Q_{j,t}$ over the total of the sector in the year:

$$\mathcal{M}_{j,t} = \frac{P_{j,t}Q_{j,t}}{\sum_{j' \in \int(s(j,t),t)} P_{j',t}Q_{j',t}}$$

We identify firms that are above the average value of price defined for every pair (s, t) by the dummy variable d:

$$d_{j,t} = \mathbb{1}\left(\breve{p}_{j,t} \geqslant \frac{\sum_{j' \in \int(s(j,t),t)} \breve{p}_{j',t}}{\sum_{j' \in \int(s(j,t),t)} 1}\right)$$

We then define:

$$M^3_{s,t} = \sum_{j \in \int(s,t)} \mathcal{M}_{j,t} d_{j,t}$$

the total market share of firms above the mean ($d_{j,t} = 1$) in sector s and year t. The moment M^3 is simply defined as the unweighted average of $M^3_{s,t}$ over all pairs (s, t).

4. **Market share of firms above mean TFPQ:** $M^{(4)}$ is computed exactly like $M^{(3)}$ except that we use $\widetilde{TFPQ}$ instead of $\breve{p}$

5. **Market share of firms above mean TFPR:** $M^{(5)}$ is computed exactly like $M^{(3)}$ except that we use $\widetilde{TFPR}$ instead of $\breve{p}$

6. **Share of firm above mean price and above mean TFPQ:** we start by defining the size of firm in each sector and year:

$$\mathcal{S}_{s,t} = \sum_{j \in \int(s,t)} 1$$

and we define two binary variables $d^{(1)}$ and $d^{(2)}$ as:

$$d_{j,t}^{(1)} = \mathbb{1}\left(\breve{p}_{j,t} \geqslant \frac{\sum_{j' \in \int(s(j,t),t)} \breve{p}_{j',t}}{\sum_{j' \in \int(s(j,t),t)} 1}\right)$$

and

$$d_{j,t}^{(2)} = \mathbb{1}\left(\widetilde{TFPQ}_{j,t} \geqslant \frac{\sum_{j' \in \int(s(j,t),t)} \widetilde{TFPQ}_{j',t}}{\sum_{j' \in \int(s(j,t),t)} 1}\right)$$

We can then calculate $M_{s,t}^{6}$ for each pair (s,t) as:

$$M_{s,t}^{6} = \sum_{j \in \int(s,t)} \mathcal{S}_{j,t} d_{j,t}^{(1)} d_{j,t}^{(2)}$$

and $M^{(6)}$ is simply defined as the unweighted average of $M_{s,t}^{6}$ over all pairs (s,t).

7. **Share of firm below mean price and above mean TFPQ:** using the same notation as above

$$M_{s,t}^{7} = \sum_{j \in \int(s,t)} \mathcal{S}_{j,t} (1 - d_{j,t}^{(1)}) d_{j,t}^{(2)}$$

and again $M^{(7)}$ is defined as the unweighted average of $M_{s,t}^{7}$ over all pairs (s,t).

8. **Share of firm above mean price and below mean TFPQ:** using the same notation as above

$$M_{s,t}^{8} = \sum_{j \in \int(s,t)} \mathcal{S}_{j,t} d_{j,t}^{(1)} (1 - d_{j,t}^{(2)})$$

$M^{(8)}$ is defined as the unweighted average of $M_{s,t}^{8}$ over all pairs (s,t).

9. **Share of firm below mean price and below mean TFPQ:** using the same notation as above

$$M_{s,t}^{9} = \sum_{j \in \int(s,t)} \mathcal{S}_{j,t} (1 - d_{j,t}^{(1)})(1 - d_{j,t}^{(2)})$$

$M^{(9)}$ is defined as the unweighted average of $M^9_{s,t}$ over all pairs (s,t). Note that M^9 is by definition equal to $1 - M^6 - M^7 - M^8$

10. **Semi-elasticity of overhead cost with respect to sales:** we denote $\mathcal{O}_{j,t}$ our proxy for total overhead cost at the firm level. Section B.B.7 explain how we measure it. We then estimate the semi-elasticity β from:

$$\frac{\mathcal{O}_{j,t}}{P^Q_{j,t}Q_{j,t}} = \beta \log\left(P^Q_{j,t}Q_{j,t}\right) + \gamma_{s(j,t),t} + \varepsilon_{j,t}$$

because $P^Q_{j,t}Q_{j,t}$ appears both in the denominator of the dependent variable and as a measure of size, we instrument it by its lagged value $\log\left(P^Q_{j,t-1}Q_{j,t-1}\right)$ and measure $M^{(10)} = \hat{\beta}$ using 2SLS. This is reported in column (4) of Table B.4.

11. **R&D share of output.** $M^{(11)}$ is measured from EU-KLEMS in the manufacturing sector.

12. **Productivity growth rate.** $M^{(12)}$ is measured by taking the average growth rate of value added TFP in the manufacturing sector divided by the labor share (total compensation over value added), taken from EU-KLEMS, over the period 1995-2019.

13. **Interest rate.** $M^{(13)}$ is directly taken from Farhi and Gourio (2018).

References

Bergeaud, Antonin, Arthur Guillouzouic, Emeric Henry, and Clément Malgouyres, "From public labs to private firms: magnitude and channels of R&D Spillovers," Technical Report 1882, Centre for Economic Performance - LSE 2022.

Farhi, Emmanuel and François Gourio, "Accounting for Macro-Finance Trends: Market Power, Intangibles, and Risk Premia," *Brookings Papers on Economic Activity*, 2018, "" (147), 147.

Goldschmidt, Deborah and Johannes F Schmieder, "The rise of domestic outsourcing and the evolution of the German wage structure," *The Quarterly Journal of Economics*, 2017, *132* (3), 1165–1217.

Monin, Rémi and M Suarez Castillo, "Product switching, market power and distance to core competency," Document de Travail G2020-06, Institut National de la Statistique et des Etudes Economiques 2020.

Wooldridge, Jeffrey M, *Econometric analysis of cross section and panel data*, MIT press, 2010.