Web Appendix

This web appendix includes some technical material for the paper "From imitation to innovation: Where is all that Chinese R&D going?" by Michael König, Zheng Song, Kjetil Storesletten, and Fabrizio Zilibotti.

The following sections are included to this Webpage Appendix:

A. Theory: Proof of Proposition 2.

B. Theory: A particular case that admits a full analytical solution.

C. Numerical implementation: Details.

D. Data and descriptive statistics: Potential mismeasurement of TFP.

E. Data and descriptive statistics: Details on calculation of TFP following Brandt et al. (2017).

F. Multiple regressions similar to Table 2 with a more stringent criterion for innovative firms.

G. Multiple regressions similar to Table 2 without province and age dummies.

A Proof of Proposition 2

In this section, we provide a formal proof of Proposition 2.

Proof of Proposition 2. The evolution of the log-productivity distribution $\Gamma_a(t)$ for large t is given by the following system of integro-difference Equations (cf. Equation (10))

$$\begin{aligned}\Gamma_a(t+1)-\Gamma_a(t) = \sum_{j\in\{l,h\}} \bar{\omega}_{\tau_j} \int_{[0,\overline{p}]} \Big[&\big(\chi^{\text{im}}(a-1,p,\tau_j;\Gamma) \\ &+\delta\chi^{\text{in}}(a-1,p,\tau_j;\Gamma)(1-p)\big)\, q(1-F_{a-1}(t))\Gamma_{a-1}(t) \\ &-\big(\chi^{\text{im}}(a,p,\tau_j;\Gamma)+\delta\chi^{\text{in}}(a,p,\tau_j;\Gamma)(1-p)\big)\, q(1-F_a(t))\Gamma_a(t) \\ &-\chi^{\text{in}}(a,p,\tau_j;\Gamma)p\Gamma_a(t)+\chi^{\text{in}}(a-1,p,\tau_j;\Gamma)p\Gamma_{a-1}(t)\Big]\, dG(p), \end{aligned} \tag{A17}$$

where $\omega_{\tau_l}, \omega_{\tau_h}$ denote the proportion of for low- and high-wedge firms, respectively, $G : [0, \overline{p}] \to [0, 1]$ is the density function of a random variable over the interval $[0, \overline{p}]$, δ is the passive imitation probability and

$$\chi^{\text{im}}(a,p,\tau;\Gamma) = 1-\chi^{\text{in}}(a,p,\tau;\Gamma) = \begin{cases} 1 & \text{if } p \le Q(a,\tau;\Gamma), \\ 0 & \text{otherwise,} \end{cases}$$

with $F_a = \sum_{b=1}^{a} \Gamma_b = \sum_{b=1}^{a} \Gamma_b$, and the average log-productivity given by $\bar{a} = \sum_{b=1}^{\infty} F_b$. We can write Equation (6) as follows

$$\begin{aligned}\Gamma_a(t+1)-\Gamma_a(t) = \sum_{j\in\{l,h\}} \bar{\omega}_{\tau_j} \int_{[0,\overline{p}]} \Big[&\big(\chi^{\text{im}}(a-1,p,\tau_j;\Gamma)+(1-p)\delta\chi^{\text{in}}(a-1,p,\tau_j;\Gamma)\big)\, q(1-F_{a-1}(t))\Gamma_{a-1}(t) \\ &-\big(\chi^{\text{im}}(a,p,\tau_j;\Gamma)+(1-p)\delta\chi^{\text{in}}(a,p,\tau_j;\Gamma)\big)\, q(1-F_a(t))\Gamma_a(t) \\ &-\chi^{\text{in}}(a,p,\tau_j;\Gamma)p\Gamma_a(t)+\chi^{\text{in}}(a-1,p,\tau_j;\Gamma)p\Gamma_{a-1}(t)\Big]\, dG(p). \end{aligned}$$

Note that $Q(a, \tau; \Gamma) > 0$ and

$$\int_{[0,\overline{p}]} \chi^{\text{im}}(a,p,\tau;\Gamma)dG(p) = \int_{[0,\overline{p}]} \mathbf{1}_{\{p<Q(a,\tau;\Gamma)\}} dG(p) = G([0,\min\{Q(a,\tau;\Gamma),\overline{p}\}]).$$

Moreover, we have that

$$\int_{[0,\overline{p}]} p\chi^{\text{in}}(a,p,\tau;\Gamma)dG(p) = \int_{[0,\overline{p}]} p\mathbf{1}_{\{p>Q(a,\tau;\Gamma)\}} dG(p) = p^*_G([Q(a,\tau;\Gamma),\overline{p}]),$$

where $p^*_G([Q(a,\tau;\Gamma),\overline{p}]) = \int_{[0,\overline{p}]} p\mathbf{1}_{\{p>Q(a,\tau;\Gamma)\}} dG(p) = \int_{[Q(a,\tau;\Gamma),\overline{p}]} p dG(p)$ denotes the average of p over the interval $[Q(a, \tau; \Gamma), \overline{p}]$. Further, we have that

$$\int_{[0,\overline{p}]} (1-p)\chi^{\text{in}}(a,p,\tau;\Gamma)dG(p) = \int_{[0,\overline{p}]} (1-p)\mathbf{1}_{\{p>Q(a,\tau;\Gamma)\}} G(dp) = q^*_G((([Q(a,\tau;\Gamma),\overline{p}]),$$

where $q^*_G((([Q(a,\tau;\Gamma),\overline{p}]) = \int_{[0,\overline{p}]}(1-p)\mathbf{1}_{\{p>Q(a,\tau;\Gamma)\}} G(dp) = \int_{[Q(a,\tau;\Gamma),\overline{p}]}(1-p)dG(p)$ denotes the average of $1-p$ over the interval $[Q(a,\tau;\Gamma)\},\overline{p}]$. Equation (6) can then be written as follows

$$\begin{aligned}\Gamma_a(t+1)-\Gamma_a(t) = \Bigg[& q(1-F_{a-1}(t))\Gamma_{a-1}(t)\Bigg(\sum_{j\in\{l,h\}}\bar{\omega}_{\tau_j}G([0,\min\{Q(a-1,\tau_j;\Gamma),\overline{p}\}]) \\ & + \delta\sum_{j\in\{l,h\}}\bar{\omega}_{\tau_j}q_G^*([Q(a-1,\tau_j;\Gamma),\overline{p}])\Bigg) \\ & -q(1-F_a(t))\Gamma_a(t)\Bigg(\sum_{j\in\{l,h\}}\bar{\omega}_{\tau_j}G([0,\min\{Q(a,\tau_j;\Gamma),\overline{p}\}])\delta\sum_{j\in\{l,h\}}\bar{\omega}_{\tau_j}q_G^*([Q(a,\tau_j;\Gamma),\overline{p}])\Bigg) \\ & -\Gamma_a(t)\sum_{j\in\{l,h\}}\bar{\omega}_{\tau_j}p_G^*([Q(a,\tau_j;\Gamma),\overline{p}])+\Gamma_{a-1}(t)\sum_{j\in\{l,h\}}\bar{\omega}_{\tau_j}p_G^*([Q(a-1,\tau_j;\Gamma),\overline{p}])\Bigg].\end{aligned} \tag{A18}$$

Writing out the two possible values for $\tau\in\{\tau_l,\tau_h\}$ which are realized with probabilities $\bar{\omega}_{\tau_l}$ and $\bar{\omega}_{\tau_h}=1-\bar{\omega}_{\tau_l}$, respectively, yields

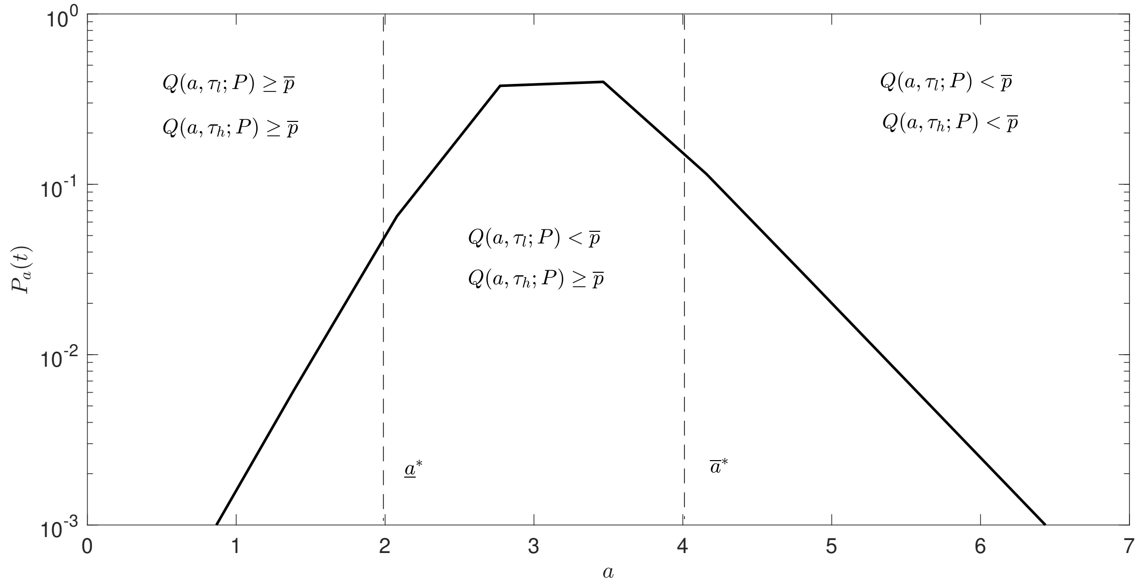
$$\begin{aligned}&\Gamma_a(t+1)-\Gamma_a(t) = \\ &[q(1-F_{a-1}(t))\Gamma_{a-1}(t)\,(\omega_{\tau_l}G([0,\min\{Q(a-1,\tau_l;\Gamma),\overline{p}\}])+(1-\omega_{\tau_l})G([0,\min\{Q(a-1,\Gamma,\tau_h),\overline{p}\}]) \\ &+\ \delta\,(\bar{\omega}_{\tau_l}q_G^*([Q(a-1,\tau_l;\Gamma),\overline{p}])+(1-\bar{\omega}_{\tau_l})q_G^*([Q(a-1,\tau_h;\Gamma),\overline{p}]))) \\ &-q(1-F_a(t))\Gamma_a(t)\,(\bar{\omega}_{\tau_l}G([0,\min\{Q(a,\tau_l;\Gamma),\overline{p}\}])+(1-\bar{\omega}_{\tau_l})G([0,\min\{Q(a,\tau_h;\Gamma),\overline{p}\}]) \\ &+\ \delta\,(\bar{\omega}_{\tau_l}q_G^*([Q(a,\tau_l;\Gamma),\overline{p}])+(1-\bar{\omega}_{\tau_l})q_G^*([Q(a,\tau_h;\Gamma),\overline{p}]))) \\ &-\Gamma_a(t)\,(\bar{\omega}_{\tau_l}p_G^*([Q(a,\tau_l;\Gamma),\overline{p}])+(1-\bar{\omega}_{\tau_l})p_G^*([Q(a,\tau_h;\Gamma),\overline{p}])) \\ &+\Gamma_{a-1}(t)\,(\bar{\omega}_{\tau_l}p_G^*([Q(a-1,\tau_l;\Gamma),\overline{p}])+(1-\bar{\omega}_{\tau_l})p_G^*([Q(a-1,\tau_h;\Gamma),\overline{p}]))]\,.\end{aligned} \tag{A19}$$

Next, note that due to the monotonicity of $Q(a,\tau;\Gamma)$, there exist two unique thresholds, $\underline{a}^*(t)<\overline{a}^*(t)$, such that

$$\begin{aligned}&Q(a,\tau_l;\Gamma)\geq\overline{p}\wedge Q(a,\tau_h;\Gamma)\geq\overline{p}\text{ if }a\leq\underline{a}^*,\\&Q(a,\tau_l;\Gamma)<\overline{p}\wedge Q(a,\tau_h;\Gamma)\geq\overline{p}\text{ if }\underline{a}^*<a\leq\overline{a}^*,\\&Q(a,\tau_l;\Gamma)<\overline{p}\wedge Q(a,\tau_h;\Gamma)<\overline{p}\text{ if }\overline{a}^*<a.\end{aligned}$$

Figure WA1 shows an illustration of the two thresholds, $\underline{a}^*(t)<\overline{a}^*(t)$, and the inequalities of Equation (A1). Considering all possible cases, we then can write Equation (A19) as follows

Figure WA1: Two Thresholds and Inequalities in Equation (A1)



Note: An illustration of the two thresholds, $\underline{a}^*(t) < \overline{a}^*(t)$, and the inequalities of Equation (A1).

$$
\begin{aligned}
&\Gamma_a(t+1) - \Gamma_a(t) = \\
&\begin{cases}
q(1-F_{a-1}(t))\Gamma_{a-1}(t)G([0,\overline{p}]) - q(1-F_a(t))\Gamma_a(t)G([0,\overline{p}]), & \text{if } a \leq \underline{a}^*(t), \\
q(1-F_{a-1}(t))\Gamma_{a-1}(t)G([0,\overline{p}]) - q(1-F_a(t))\Gamma_a(t)\left(\bar{\omega}_{\tau_l}G([0,Q(a,\tau_l;\Gamma)])\right. & \\
\quad \left.+(1-\bar{\omega}_{\tau_l})G([0,\overline{p}]) + \delta\bar{\omega}_{\tau_l}q_G^*([Q(a,\tau_l;\Gamma),\overline{p}])\right) - \Gamma_a(t)\bar{\omega}_{\tau_l}p_G^*([Q(a,\tau_l;\Gamma),\overline{p}]), & \text{if } a = \underline{a}^*(t)+1 < \overline{a}^*(t), \\
q(1-F_{a-1}(t))\Gamma_{a-1}(t)\left(\bar{\omega}_{\tau_l}G([0,Q(a-1,\Gamma,\tau_l)]) + (1-\bar{\omega}_{\tau_l})G([0,\overline{p}])\right. & \\
\quad \left.+\delta\bar{\omega}_{\tau_l}p_G^*([Q(a-1,\Gamma,\tau_l),\overline{p}])\right) & \\
\quad -q(1-F_a(t))\Gamma_a(t)\left(\bar{\omega}_{\tau_l}G([0,Q(a,\tau_l;\Gamma)]) + (1-\bar{\omega}_{\tau_l})G([0,\overline{p}])\right. & \\
\quad \left.+\delta\bar{\omega}_{\tau_l}p_G^*([Q(a,\tau_l;\Gamma),\overline{p}])\right) & \\
\quad -\Gamma_a(t)\bar{\omega}_{\tau_l}p_G^*([Q(a,\tau_l;\Gamma),\overline{p}]) + \Gamma_{a-1}(t)\bar{\omega}_{\tau_l}p_G^*([Q(a-1,\tau_l;\Gamma),\overline{p}]), & \text{if } \underline{a}^*(t)+1 < a \leq \overline{a}^*(t), \\
q(1-F_{a-1}(t))\Gamma_{a-1}(t)\left(\bar{\omega}_{\tau_l}G([0,Q(a-1,\tau_l;\Gamma)]) + (1-\omega_{\tau_l})G([0,\overline{p}])\right. & \\
\quad \left.+\delta\bar{\omega}_{\tau_l}q_G^*([Q(a-1,\tau_l;\Gamma),\overline{p}])\right) & \\
\quad -q(1-F_a(t))\Gamma_a(t)\left(\bar{\omega}_{\tau_l}G([0,Q(a,\tau_l;\Gamma)]) + (1-\bar{\omega}_{\tau_l})G([0,Q(a,\tau_h;\Gamma)])\right. & \\
\quad \left.+\delta\omega_{\tau_l}q_G^*(([Q(a,\tau_l;\Gamma),\overline{p}]) + \delta(1-\bar{\omega}_{\tau_l})q_G^*([Q(a,\tau_h;\Gamma),\overline{p}])\right) & \\
\quad -\Gamma_a(t)\left(\bar{\omega}_{\tau_l}p_G^*([Q(a,\tau_l;\Gamma),\overline{p}]) + (1-\bar{\omega}_{\tau_l})p_G^*([Q(a,\tau_h;\Gamma),\overline{p}])\right) & \\
\quad +\Gamma_{a-1}(t)\bar{\omega}_{\tau_l}p_G^*([Q(a-1,\tau_l;\Gamma),\overline{p}]), & \text{if } a = \overline{a}^*(t)+1, \\
q(1-F_{a-1}(t))\Gamma_{a-1}(t)\left(\bar{\omega}_{\tau_l}G([0,Q(a-1,\tau_l;\Gamma)])\right. & \\
\quad + (1-\bar{\omega}_{\tau_l})G([0,Q(a-1,\tau_h;\Gamma)]) + \delta\bar{\omega}_{\tau_l}q_G^*([Q(a-1,\tau_l;\Gamma),\overline{p}]) & \\
\quad \left.+ \delta(1-\bar{\omega}_{\tau_l})q_G^*([Q(a-1,\tau_h;\Gamma),\overline{p}])\right) & \\
\quad -q(1-F_a(t))\Gamma_a(t)\left(\bar{\omega}_{\tau_l}G([0,Q(a,\tau_l;\Gamma)]) + (1-\bar{\omega}_{\tau_l})G([0,Q(a,\tau_h;\Gamma)])\right. & \\
\quad \left.+ \delta\bar{\omega}_{\tau_l}q_G^*([Q(a,\tau_l;\Gamma),\overline{p}]) + \delta(1-\bar{\omega}_{\tau_l})q_G^*([Q(a,\tau_h;\Gamma),\overline{p}])\right) & \\
\quad -\Gamma_a(t)\left(\bar{\omega}_{\tau_l}p_G^*([Q(a,\tau_l;\Gamma),\overline{p}]) + (1-\bar{\omega}_{\tau_l})p_G^*([Q(a,\tau_h;\Gamma),\overline{p}])\right) & \\
\quad +\Gamma_{a-1}(t)\left(\bar{\omega}_{\tau_l}p_G^*([Q(a-1,\tau_l;\Gamma),\overline{p}]) + (1-\bar{\omega}_{\tau_l})p_G^*([Q(a,\tau_h;\Gamma),\overline{p}])\right), & \text{if } a > \overline{a}^*(t)+1.
\end{cases}
\end{aligned}
$$

After some tedious calculations similar to the proof of Proposition 1 we can derive from the dynamics of the p.m.f the corresponding dynamics of the c.d.f.:

$$F_a(t+1) - F_a(t) = \sum_{b=1}^{a} \left(\Gamma_b(t+1) - \Gamma_b(t)\right) =$$

$$\begin{cases} -q(1-F_a(t))(F_a(t)-F_{a-1}(t)), & \text{if } a \le \underline{a}^*(t), \\ -q(1-F_a(t))(F_a(t)-F_{a-1}(t))\left(\bar{\omega}_{\tau_l} G([0,Q(a,\tau_l;\Gamma)]) + (1-\bar{\omega}_{\tau_l})G([0,Q(a,\tau_h;\Gamma)])\right. & \\ \quad \left. + \, \delta\bar{\omega}_{\tau_l} q_G^*([Q(a,\tau_l;\Gamma),\overline{p}]) + \delta(1-\bar{\omega}_{\tau_l})q_G^*([Q(a,\tau_h;\Gamma),\overline{p}])\right) & \\ \quad -(F_a(t)-F_{a-1}(t))\left(\bar{\omega}_{\tau_l} p_G^*([Q(a,\tau_l;\Gamma),\overline{p}]) + (1-\bar{\omega}_{\tau_l})p_G^*([Q(a,\tau_h;\Gamma),\overline{p}])\right), & \text{if } a > \overline{a}^*(t)+1. \end{cases}$$

Similarly for the complementary cumulative distribution function, $H_a(t) = 1 - F_a(t)$, we get

$$H_a(t+1) - H_a(t) =$$

$$\begin{cases} qH_a(t)(H_{a-1}(t)-H_a(t)), & \text{if } a \le \underline{a}^*(t), \\ qH_a(t)(H_{a-1}(t)-H_a(t))\left(\bar{\omega}_{\tau_l} G([0,Q(a,\tau_l;\Gamma)]) + (1-\bar{\omega}_{\tau_l})G([0,Q(a,\tau_h;\Gamma)])\right. & \\ \quad + \, \delta\bar{\omega}_{\tau_l} q_G^*([Q(a,\tau_l;\Gamma),\overline{p}]) & \\ \quad \left. + \, \delta(1-\bar{\omega}_{\tau_l})q_G^*([Q(a,\tau_h;\Gamma),\overline{p}])\right) & \\ \quad +(H_{a-1}(t)-H_a(t))\left(\bar{\omega}_{\tau_l} p_G^*([Q(a,\tau_l;\Gamma),\overline{p}]) + (1-\bar{\omega}_{\tau_l})p_G^*([Q(a,\tau_h;\Gamma),\overline{p}])\right), & \text{if } a > \overline{a}^*(t)+1. \end{cases}$$

Note that $F_a(t+1) \le F_a(t)$ (and conversely, $H_a(t+1) \ge H_a(t)$). Since the probability mass is conserved to one (and $\lim_{a\to+\infty} F_a = 1$), the fact that F_a is decreasing over time t for every a implies that the distribution must shift to the right (i.e. towards higher values of a). As in the proof of Proposition 1 there exists a travelling wave solution of the form $F_a(t) = \tilde{f}(a-\nu t)$ ($H_a(t) = h(a-\nu t)$) with velocity $\nu > 0$, so that $F_a(t+1) - F_a(t) = -\nu \tilde{f}'$, and therefore

$$-\nu \tilde{f}'(x) = \begin{cases} -q(1-\tilde{f}(x))(\tilde{f}(x)-\tilde{f}(x-1)), & \text{if } x \le \underline{x}^*, \\ -q(1-\tilde{f}(x))(\tilde{f}(x)-\tilde{f}(x-1))\left(\bar{\omega}_{\tau_l} G([0,Q(x,\tau_l;\Gamma)]) + (1-\bar{\omega}_{\tau_l})G([0,Q(x,\tau_h;\Gamma)])\right. & \\ \quad \left. + \, \delta\bar{\omega}_{\tau_l} q_G^*([Q(x,\tau_l;\Gamma),\overline{p}]) + \delta(1-\bar{\omega}_{\tau_l})q_G^*([Q(x,\tau_h;\Gamma),\overline{p}])\right) & \\ \quad -(\tilde{f}(x)-\tilde{f}(x-1))\left(\bar{\omega}_{\tau_l} p_G^*([Q(x,\tau_l;\Gamma),\overline{p}]) + (1-\bar{\omega}_{\tau_l})p_G^*([Q(x,\tau_h;\Gamma),\overline{p}])\right), & \text{if } x > \overline{x}^*+1, \end{cases}$$

where we have denoted by $x = a - \nu t$. Similarly

$$-\nu \tilde{h}'(x) = \begin{cases} q\tilde{h}(x)(\tilde{h}(x-1)-\tilde{h}(x)), & \text{if } x \le \underline{x}^*, \\ qh(x)(h(x)-h(x-1))\left(\bar{\omega}_{\tau_l} G([0,Q(a,\tau_l;\Gamma)]) + (1-\bar{\omega}_{\tau_l})G([0,Q(a,\tau_h;\Gamma)])\right. & \\ \left. + \, \delta\bar{\omega}_{\tau_l} q_G^*([Q(a,\tau_l;\Gamma),\overline{p}]) + \delta(1-\bar{\omega}_{\tau_l})q_G^*([Q(a,\tau_h;\Gamma),\overline{p}])\right) & \\ +(h(x)-h(x-1))\left(\bar{\omega}_{\tau_l} p_G^*([Q(a,\tau_l;\Gamma),\overline{p}]) + (1-\bar{\omega}_{\tau_l})p_G^*([Q(a,\tau_h;\Gamma),\overline{p}])\right), & \text{if } x > \overline{x}^*+1. \end{cases}$$

For the case of $x \le \underline{x}^*$ we obtain the delay difference equation

$$-\nu f'(x) = -q(1-\tilde{f}(x))(\tilde{f}(x)-\tilde{f}(x-1)),$$

or equivalently

$$\nu f'(x) = q(\tilde{f}(x) - \tilde{f}(x-1) - \tilde{f}(x)^2 + \tilde{f}(x)\tilde{f}(x-1)).$$

For $x \to -\infty$, the terms $\tilde{f}(x)^2$ and $\tilde{f}(x)\tilde{f}(x-1)$ can be neglected (compared to higher order terms), and we can write

$$\nu f'(x) = q(\tilde{f}(x) - \tilde{f}(x-1)).$$

We guess a solution of the form $\tilde{f}(x) = c_1 e^{\lambda x}$ for $x \to -\infty$. Inserting gives

$$c_1 \lambda \nu e^{\lambda x} \approx q(c_1 e^{\lambda x} - c_1 e^{\lambda(x-1)}).$$

This can be written as $\lambda\nu \approx q(1 - e^{-\lambda})$. The solution is given by

$$\lambda = \frac{\nu W\left(-\frac{q e^{-\frac{q}{\nu}}}{\nu}\right) + q}{\nu},$$

where W denotes the Lambert W-function, and we require that $\frac{q e^{-\frac{q}{\nu}}}{\nu} \leq \frac{1}{e}$. Similarly, for the case of $x > \overline{x}^* + 1$ we obtain the DDE

$$\begin{aligned}
\nu h'(x) = q\left(h(x)^2 - h(x)h(x-1)\right)&\left(\omega_{\tau_l} G([0, Q(x, \tau_l; \Gamma)]) + (1 - \bar{\omega}_{\tau_l}) G([0, Q(x, \tau_h; \Gamma)])\right.\\
&+ \delta \bar{\omega}_{\tau_l} q_G^*(([Q(x, \tau_l; \Gamma), \overline{p}]) + \delta(1 - \bar{\omega}_{\tau_l}) q_G^*((([Q(x, \tau_h; \Gamma), \overline{p}]))\\
&+ (h(x) - h(x-1))\left(\bar{\omega}_{\tau_l} p_G^*([Q(x, \tau_l; \Gamma), \overline{p}]) + (1 - \bar{\omega}_{\tau_l}) p_G^*([Q(x, \tau_h; \Gamma), \overline{p}])\right).
\end{aligned}$$

For $x \to +\infty$, the terms $h(x)^2$ and $h(x)h(x-1)$ can be neglected, and $Q(x, \tau; \Gamma)$ will tend to zero, so that we can write

$$\nu h'(x) = (h(x) - h(x-1))\left(\bar{\omega}_{\tau_l} p_G^*([0, \overline{p}]) + (1 - \bar{\omega}_{\tau_l}) p_G^*([0, \overline{p}])\right).$$

We guess a solution of the form $h(x) = c_2 e^{-\rho x}$ for large x. Inserting gives

$$-c_2 \nu \rho e^{-\rho x} = (c_2 e^{-\rho x} - c_2 e^{-\rho(x-1)})\left(\bar{\omega}_{\tau_l} p_G^*([0, \overline{p}]) + (1 - \bar{\omega}_{\tau_l}) p_G^*([0, \overline{p}])\right).$$

The exponent ρ is then the root of the following transcendental equation

$$\nu \rho = p_G^*([0, \overline{p}])(e^{\rho} - 1).$$

The solution can be written as

$$\rho = \frac{-\nu W\left(-\frac{p_G^*([0,\overline{p}]) e^{-\frac{p_G^*([0,\overline{p}])}{\nu}}}{\nu}\right) - p_G^*([0, \overline{p}])}{\nu},$$

where W denotes the Lambert W-function, and we require that $\frac{p_G^*([0,\overline{p}]) e^{-\frac{\rho}{\nu}}}{\nu} \leq \frac{1}{e}$. This concludes the proof of the proposition. ■

B A particular case that admits an analytical solution.

The following remark deals with a particular case in which the growth rate ν admits an analytical solution.

Remark 1 (Triangular Approximation) *Assume $p_i = p$ for all i (no ex-ante heterogeneity in in-house R&D capability), $\delta = 0$ (firms pursuing innovation cannot imitate) and assume that $q > p$. Making a triangular approximation*[42] *of the probability mass function (p.m.f.), there exists a traveling wave solution of the form $P_a(t) = f(a - \nu t)$ with velocity $\nu > 0$, with left and right Pareto tails, characterized as follows: (i) The right power tail exponent ρ can be approximated by the root of the equation $(q-p)e^{\frac{q\rho}{p-q}} + pe^{\rho} - q = 0$ (where a real root is guaranteed to exists if $q > p$), while the left tail exponent λ can be approximated by $\lambda \simeq q\rho/(q-p)$; (ii) the traveling wave velocity is approximately given by $\nu \simeq p(e^{\rho} - 1)/\rho$; (iii) the tail exponents, λ, ρ, and the traveling wave velocity, ν, satisfy the following comparative statics with respect to p and q:*

$$\frac{\partial \rho}{\partial p} \simeq -\frac{e^{\rho\left(1-\frac{q}{p-q}\right)} - (p - q(\rho+1))/(p-q)}{pe^{\rho\left(1-\frac{q}{p-q}\right)} - q} \leq 0,$$

$$\frac{\partial \rho}{\partial q} \simeq \frac{1 + e^{\frac{q\rho}{p-q}}(p(\rho-1)+q)/(p-q)}{pe^{\rho} - qe^{\frac{q\rho}{p-q}}} \geq 0,$$

$$\frac{\partial \lambda}{\partial p} \simeq \frac{\rho q}{(q-p)^2} - \frac{q}{(q-p)}\frac{e^{\rho\left(1-\frac{q}{p-q}\right)} - (p - q(\rho+1))/(p-q)}{pe^{\rho\left(1-\frac{q}{p-q}\right)} - q} \leq 0,$$

$$\frac{\partial \lambda}{\partial q} \simeq -\frac{p\rho}{(q-p)^2} + \frac{q}{q-p}\frac{e^{\frac{q\rho}{p-q}}(p(\rho-1)+q)/(p-q)+1}{pe^{\rho} - qe^{\frac{q\rho}{p-q}}} \geq 0,$$

$$\frac{\partial \nu}{\partial p} \simeq \frac{e^{\rho}-1}{\rho} - p\frac{1+e^{\rho}(\rho-1)}{\rho^2}\frac{e^{\rho} - (p-q(\rho+1))e^{\frac{q\rho}{p-q}}/(p-q)}{pe^{\rho} - qe^{\frac{q\rho}{p-q}}} \geq 0,$$

$$\frac{\partial \nu}{\partial q} \simeq p\frac{1+e^{\rho}(\rho-1)}{\rho^2}\frac{e^{\frac{q\rho}{p-q}}(p(\rho-1)+q)/(p-q)+1}{pe^{\rho} - qe^{\frac{q\rho}{p-q}}} \geq 0,$$

and, in particular, $\frac{\partial \nu}{\partial p} \geq \frac{\partial \nu}{\partial q}$.

From the comparative statics analysis we find that an increase in p increases the productivity dispersion (because $\frac{\partial \rho}{\partial p} \leq 0$ and $\frac{\partial \lambda}{\partial p} \leq 0$) while an increase in q decreases the dispersion (because $\frac{\partial \rho}{\partial q} \geq 0$ and $\frac{\partial \lambda}{\partial q} \geq 0$). Moreover, both, an increase in p and an increase in q increase the traveling wave velocity ν (growth rate), but the marginal effect is higher for p than for q (since $\frac{\partial \nu}{\partial p} \geq \frac{\partial \nu}{\partial q}$).

Proof of Remark 1. When $\delta = 0$ and $p_i = p$ for all i, we can write the DDEs in the following compact form:

$$-\nu\tilde{h}'(x) = \begin{cases} q\tilde{h}(x)(\tilde{h}(x-1) - \tilde{h}(x)), & \text{if } x \leq x^*, \\ p(\tilde{h}(x-1) - \tilde{h}(x)), & \text{if } x > x^*. \end{cases} \tag{A20}$$

[42] A triangular function is a function whose graph takes the shape of a triangle. See the right panels in Figures WA2 and WA4.

Consider, first, the case $x \leq x^*$. The definition of $\tilde{h}$ as a complementary c.d.f implies the following boundary condition: $\lim_{x\to-\infty} \tilde{h}(x) = 1$. The solution to this DDE can be expressed as (cf. Asl and Ulsoy, 2003):

$$\tilde{h}(x) = \sum_{k=-\infty}^{\infty} \alpha_k e^{\lambda_k x}.$$

The boundary condition implies that $\lambda_0 = 0$ and $\alpha_0 = 1$, so that we can write

$$\tilde{h}(x) = 1 - \sum_{k\neq 0} \tilde{\alpha}_k e^{\lambda_k x}.$$

Taking only the dominant term (and denoting it by $\lambda_1 = \lambda$; we also set $\tilde{\alpha}_1 = \alpha$), we can write the following approximation

$$\tilde{h}(x) \sim 1 - \alpha e^{\lambda x}, \tag{A21}$$

Denote by $\tilde{\pi}$ the associated probability mass function (p.m.f.). Then,

$$\tilde{\pi}(x) = \tilde{h}(x-1) - \tilde{h}(x) \sim 1 - \alpha e^{\lambda(x-1)} - (1 - \alpha e^{\lambda x}) = \alpha e^{\lambda x}(1 - e^{-\lambda}).$$

Next, consider the range $x > x^*$. There,

$$\nu \tilde{h}'((x) = p(\tilde{h}(x) - \tilde{h}(x-1)),$$

with the solution

$$\tilde{h}(x) = \sum_{k=-\infty}^{\infty} \beta_k e^{-\rho_k x}.$$

The definition of $\tilde{h}$ as a complementary c.d.f. implies the following boundary condition: $\lim_{x\to\infty} \tilde{h}(x) = 0$. By the same procedure as in the other case, we can write the following approximation:

$$\tilde{h}(x) \sim \beta e^{-\rho x}, \tag{A22}$$

The corresponding p.m.f. $\tilde{\pi}$ is given by

$$\tilde{\pi}(x) = \tilde{h}(x-1) - \tilde{h}(x) \sim \beta e^{-\rho(x-1)} - \beta e^{-\rho x} = \beta e^{-\rho x}(e^{\rho} - 1).$$

Next, requiring continuity of the p.m.f. at the threshold $x = x^*$ yields

$$\alpha(1 - e^{-\lambda}) = \beta(e^{\rho} - 1).$$

Solving for β yields $\beta = \frac{1-e^{-\lambda}}{e^{\rho}-1}\alpha$. Inserting it into the equation for π yields $\tilde{\pi}(x) = (1 - e^{-\lambda})\alpha e^{-\rho x}$ for $x > x^*$.

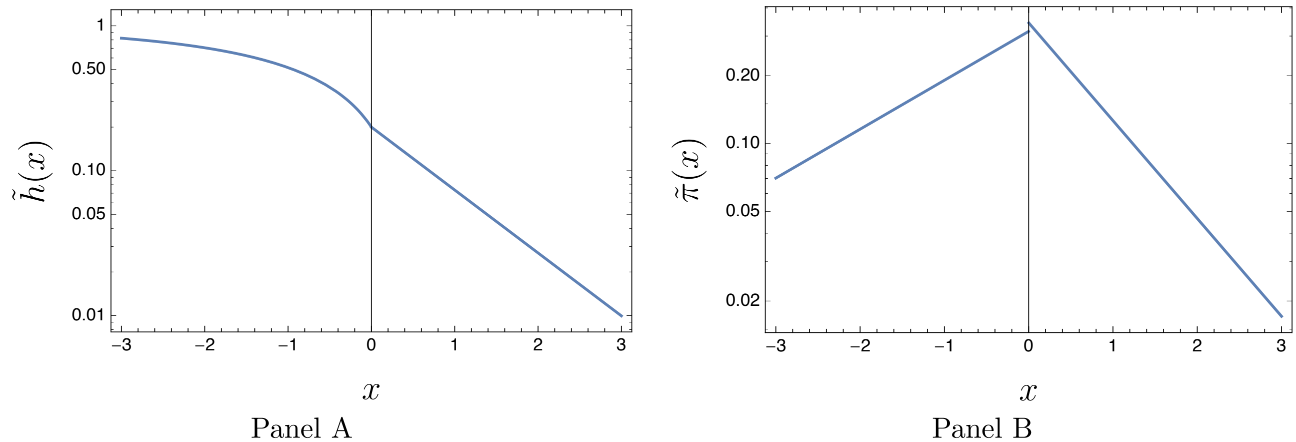
At the threshold $x = x^*$, the expected profits from innovation and imitation must be the same. Thus setting $\chi^{\text{im}}(a, P) = \chi^{\text{in}}(a, P)$ implies that

$$p = q\tilde{h}(0) = q(1 - \alpha).$$

Solving for α yields

$$\alpha = \frac{q - p}{q}.$$

Figure WA2: Functions $\tilde{h}(x)$ and $\tilde{\pi}(x)$.



Note: Panel A plots the complementary c.d.f. $\tilde{h}(x)$ of Equation (A23). Panel B plots the p.m.f. $\tilde{\pi}(x)$ of Equation (A24).

Hence, we can write

$$\tilde{h}(x) \sim \begin{cases} 1 - \frac{q-p}{q} e^{\lambda x}, & \text{if } x \leq x^*, \\ \frac{q-p}{q} \frac{1-e^{-\lambda}}{e^{\rho}-1} e^{-\rho x}, & \text{if } x > x^*, \end{cases} \tag{A23}$$

and

$$\tilde{\pi}(x) \sim \begin{cases} \frac{q-p}{q}(1-e^{-\lambda})e^{\lambda x}, & \text{if } x \leq x^*, \\ \frac{q-p}{q}(1-e^{-\lambda})e^{-\rho x}, & \text{if } x > x^*. \end{cases} \tag{A24}$$

An illustration can be seen in Figure WA2. The figure shows that the first order approximations of Equations (A21) and (A22) correspond to a triangular approximation of the true p.m.f. solving Equation (A20).[43] Furthermore, the properties of the p.m.f. function $\tilde{\pi}$ require that

$$1 = \sum_{x=-\infty}^{\infty} \tilde{\pi}(x) = \frac{q-p}{q}(1-e^{-\lambda}) \left(\sum_{x=-\infty}^{0} e^{\lambda x} + \sum_{x=1}^{\infty} e^{-\rho x} \right) = \frac{q-p}{q} \frac{(e^{\rho} - e^{-\lambda})}{e^{\rho} - 1}.$$

Next, we know that the right-tail exponent satisfies $\nu\rho = p(e^{\rho} - 1)$, while the left-tail exponent satisfies $\lambda\nu = q(1 - e^{-\lambda})$. We then can write

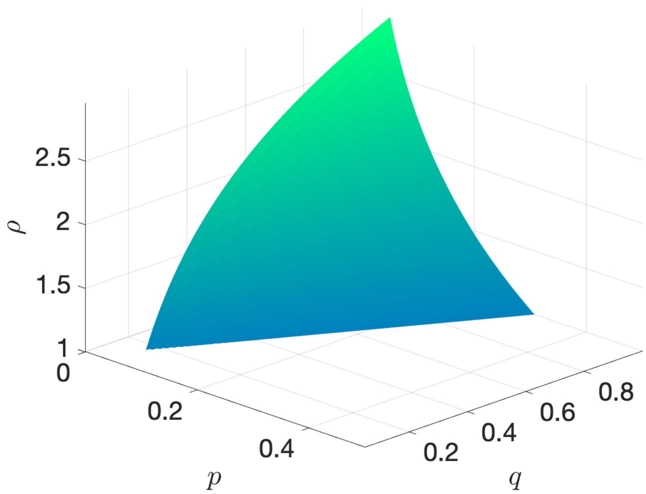
$$1 = \frac{q-p}{q} \frac{(e^{\rho} - 1 + 1 - e^{-\lambda})}{e^{\rho} - 1} = \frac{q-p}{q} \frac{\left(\frac{\nu\rho}{p} + \frac{\lambda\nu}{q}\right)}{\frac{\nu\rho}{p}} = \frac{q-p}{q} \left(1 + \frac{p\lambda}{q\rho}\right).$$

Further, we have that

$$\frac{p(e^{\rho} - 1)}{\rho} = \frac{q(1 - e^{-\lambda})}{\lambda}.$$

[43]See also the right panel in Figure WA4.

Figure WA3: The solution ρ of Equation (A25) as a function of p and q.



Hence, we have to solve a system of two unknowns, λ, ρ, given by

$$\frac{\lambda p}{q\rho} = \frac{1 - e^{-\lambda}}{e^{\rho} - 1},$$
$$1 + \frac{\lambda p}{q\rho} = \frac{q}{q - p}.$$

Solving the second equation for λ gives

$$\lambda = \frac{q\rho}{q - p},$$

and inserting into the first yields a nonlinear equation for ρ given by

$$(q - p)e^{\frac{q\rho}{p-q}} = q - pe^{\rho}. \tag{A25}$$

An illustration of the LHS and RHS of Equation (A25) can be seen in Figure WA4 for $q = 0.5$ and $p = 0.1$. The figure illustrates that there exists a unique non-trivial root for ρ. A solution exists as long as $p < q$. This root is shown in Figure WA3 for different values of p and q (with $p < q$).

We next turn to the comparative statics analysis. Recall from Panel A in Figure WA4 that ρ is the intersection of the two curves $f_1(p, q, \rho) \equiv f_2(p, q, \rho)$ where

$$f_1(p, q, \rho) = (p - q)e^{\frac{q\rho}{p-q}}$$

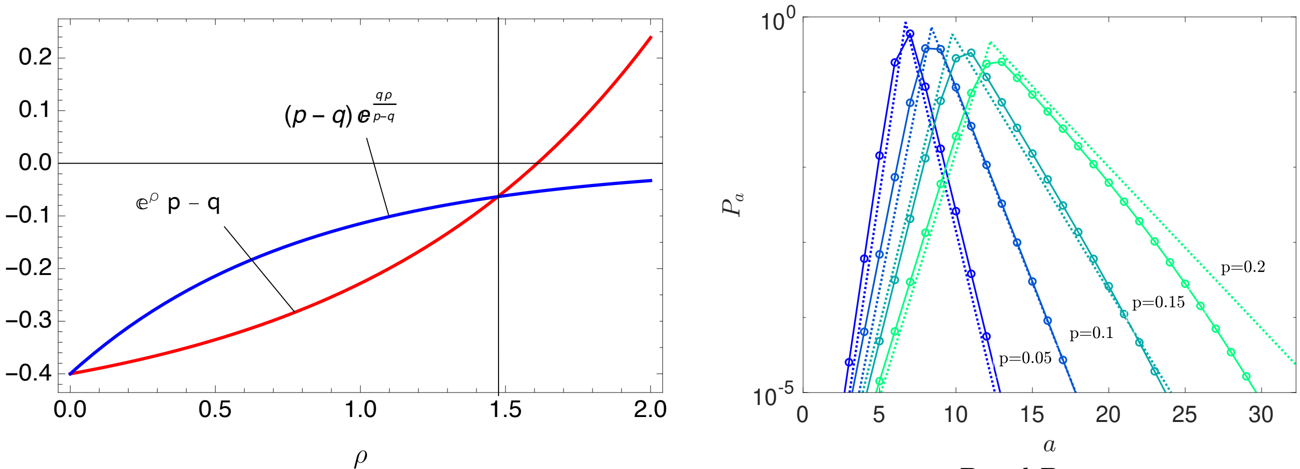
and

$$f_2(p, q, \rho) \equiv pe^{\rho} - q.$$

Increasing q shifts both curves down, while increasing p shifts both curves up. More precisely, the marginal effect from an increase in p is

$$\frac{\partial f_1}{\partial p} = \frac{e^{\frac{q\rho}{p-q}}(p - q(\rho + 1))}{p - q} \geq 0,$$
$$\frac{\partial f_2}{\partial p} = e^{\rho} \geq 0.$$

Figure WA4: Illustration of Solutions to Equations (A24)–(A25).



Panel A

Panel B

Note: Panel A plots the left-hand side and right-hand side of Equation (A25) for $q = 0.5$ and $p = 0.1$. Panel B plots the numerical solution to Equation (A20) and the approximation of Equation (A24) for $q = 0.5$ and varying values of p. As p becomes smaller, the distribution becomes more peaked and the triangular approximation becomes more accurate.

We get that $\frac{\partial f_1}{\partial p} \leq \frac{\partial f_2}{\partial p}$ for $q > p$. Hence, an increase in p shifts f_2 further up than f_1 and therefore, the intersection in ρ where $f_1 = f_2$ gets moved to the left. This means that $\frac{\partial \rho}{\partial p} \leq 0$. Similarly, the marginal effect from an increase in q is

$$\frac{\partial f_1}{\partial q} = \frac{e^{\frac{q\rho}{p-q}}(p(\rho - 1) + q)}{p - q} \leq 0$$
$$\frac{\partial f_2}{\partial q} = -1 < 0.$$

We have that $|\frac{\partial f_1}{\partial q}| \leq |\frac{\partial f_2}{\partial q}|$ for $q > p$. Hence, an increase in q shifts f_2 further down than f_1 and therefore, the intersection $f_1 = f_2$ gets moved to the right. This means that $\frac{\partial \rho}{\partial q} \geq 0$.

In order to obtain explicit algebraic expressions for the marginal effects, we can apply the implicit function theorem. Using Equation (A25) we define $f(p, q, \rho(p, q)) \equiv (q - p)e^{\frac{q\rho}{p-q}} + pe^{\rho} - q$. The right tail exponent ρ is then the solution to $f(p, q, \rho(p, q)) = 0$. We have shown graphically that $\frac{\partial \rho}{\partial p} \leq 0$ and $\frac{\partial \rho}{\partial q} \geq 0$. Taking the total derivative of f with respect to p gives $\frac{\partial f}{\partial p} + \frac{\partial f}{\partial \rho}\frac{\partial \rho}{\partial p} = 0$, and solving for $\frac{\partial \rho}{\partial p}$ yields

$$\frac{\partial \rho}{\partial p} = -\frac{\frac{\partial f}{\partial p}}{\frac{\partial f}{\partial \rho}}.$$

Similarly, from the total derivative with respect to q we get

$$\frac{\partial \rho}{\partial q} = -\frac{\frac{\partial f}{\partial q}}{\frac{\partial f}{\partial \rho}}.$$

We have for the right tail exponent ρ that

$$\frac{\partial f}{\partial p} = e^{\rho} - \frac{p - q(1+\rho)}{p-q} e^{\frac{q\rho}{p-q}},$$
$$\frac{\partial f}{\partial q} = -\frac{e^{\frac{q\rho}{p-q}} (p(\rho-1)+q)}{p-q} - 1,$$
$$\frac{\partial f}{\partial \rho} = pe^{\rho} - qe^{\frac{q\rho}{p-q}},$$

and therefore

$$\frac{\partial \rho}{\partial p} = -\frac{e^{\rho} - \frac{(p-q(\rho+1))e^{\frac{q\rho}{p-q}}}{p-q}}{pe^{\rho} - qe^{\frac{q\rho}{p-q}}} = -\frac{e^{\rho\left(1-\frac{q}{p-q}\right)} - \frac{p-q(\rho+1)}{p-q}}{pe^{\rho\left(1-\frac{q}{p-q}\right)} - q} \leq 0.$$

See also Figure WA3. Moreover, we have that

$$\frac{\partial \rho}{\partial q} = \frac{\frac{e^{\frac{q\rho}{p-q}}(p(\rho-1)+q)}{p-q} + 1}{pe^{\rho} - qe^{\frac{q\rho}{p-q}}} \geq 0,$$

for $p < q$. See Figure WA3. For the left tail exponent $\lambda = q\rho/(q-p)$ we get

$$\frac{\partial \lambda}{\partial p} = \frac{q\rho}{(q-p)^2} + \frac{q}{q-p}\frac{\partial \rho}{\partial p} = \frac{q}{(q-p)}\left(\frac{\rho}{(q-p)} - \frac{e^{\rho\left(1-\frac{q}{p-q}\right)} - \frac{p-q(\rho+1)}{p-q}}{pe^{\rho\left(1-\frac{q}{p-q}\right)} - q}\right) \leq 0.$$

Similarly,

$$\frac{\partial \lambda}{\partial q} = -\frac{p\rho}{(q-p)^2} + \frac{q}{q-p}\frac{\partial \rho}{\partial q} = -\frac{p\rho}{(q-p)^2} + \frac{q}{q-p}\frac{\frac{e^{\frac{q\rho}{p-q}}(p(\rho-1)+q)}{p-q} + 1}{pe^{\rho} - qe^{\frac{q\rho}{p-q}}} \geq 0.$$

Further, for the traveling wave velocity $\nu = p(e^{\rho}-1)/\rho$ we have that

$$\frac{\partial \nu}{\partial p} = \frac{e^{\rho}-1}{\rho} + p\frac{1+e^{\rho}(\rho-1)}{\rho^2}\frac{\partial \rho}{\partial p} = \frac{e^{\rho}-1}{\rho} - p\frac{1+e^{\rho}(\rho-1)}{\rho^2}\frac{e^{\rho} - \frac{(p-q(\rho+1))e^{\frac{q\rho}{p-q}}}{p-q}}{pe^{\rho} - qe^{\frac{q\rho}{p-q}}} \geq 0.$$

Similarly,

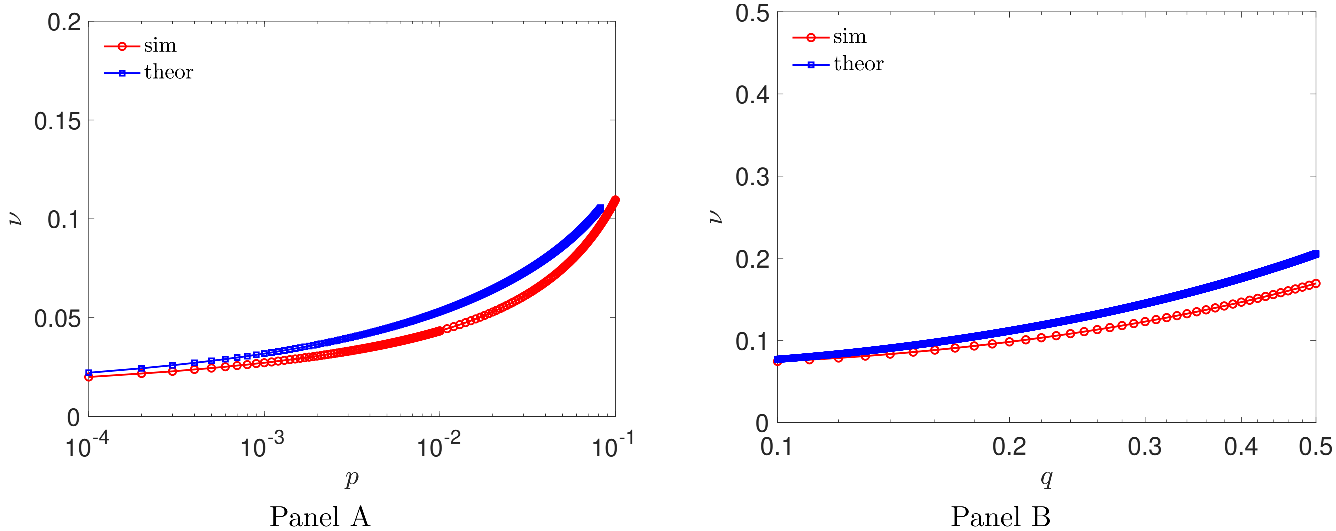
$$\frac{\partial \nu}{\partial q} = p\frac{1+e^{\rho}(\rho-1)}{\rho^2}\frac{\partial \rho}{\partial q} = p\frac{1+e^{\rho}(\rho-1)}{\rho^2}\frac{\frac{e^{\frac{q\rho}{p-q}}(p(\rho-1)+q)}{p-q} + 1}{pe^{\rho} - qe^{\frac{q\rho}{p-q}}} \geq 0,$$

and from comparing both marginal effects we find that $\frac{\partial \nu}{\partial p} \geq \frac{\partial \nu}{\partial q}$.

Figure WA5 shows a comparison of the growth rate ν from a direct numerical simulation of Equation (A20) and the theoretical prediction under the triangular approximation for varying values of p and q. The growth rate ν is decreasing with both, p and q, and the triangular approximation becomes more accurate as p and q become smaller.

■

Figure WA5: A comparison of the growth rate ν from a direct numerical simulation of Equation (A20) and the theoretical prediction of the triangular approximation.



Note: Panel A plots the change in the growth rate ν as a function of the in-house R&D success probability p while setting $q = 0.164$. Panel B plots the change in the growth rate ν as a function of the imitation success probability q while setting $p = 0.071$.

.

C Numerical implementation

In the theory, there are countably infinite many rungs of the productivity (log TFP) ladder, $\{..., a_n, a_{n+1}, a_{n+2}, ...\}$. The distance between each rung is $\tilde{a}$. In the numerical implementation we make two adjustments. First, we truncate the distribution. Second, we place the rungs at a closer distance than $\tilde{a}$ to allow a finer grid. However, we keep the step size of productivity growth following successful imitation or innovation constant at the calibrated level $\tilde{a}$ (whose calibration is discussed in the text). Such a finer grid has two advantages. First, it allows us to better approximate a smooth continuous distribution. Second, and more importantly, the finer grid substantially reduces the occurrence of cycles in which the numerical routine to compute stationary equilibria occasionally gets stuck. This speeds up the estimation procedure.

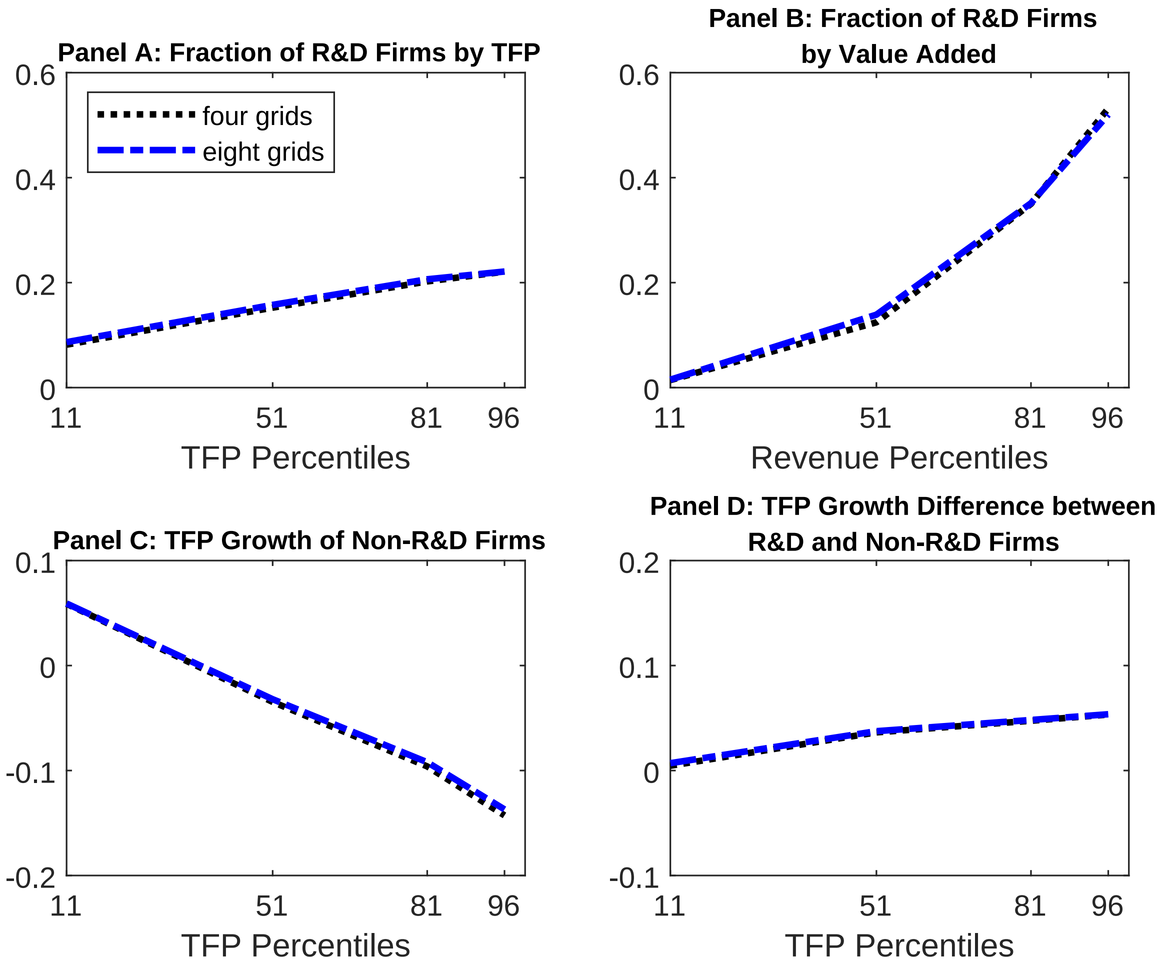
We assume that successful imitation can only occur if the imitating firm meets a firm whose TFP is larger than its own TFP by at least $\tilde{a}$. More formally, the TFP increment $\tilde{a}$ accrues only when a firm with log(TFP) a_n meets another firm whose log(TFP) is larger than or equal to $a_n + \tilde{a}$. Otherwise, an imitating firm would become more productive than the firm from which it learns, which would be illogical.

More formally, we construct a bounded discrete state space of log TFP, $\{a_1, a_2, ..., a_N\}$ that approximates the theoretically unbounded state space. Let $\mathcal{A}_{n,t}$ denote the proportion of firms at the nth grid point at the end of period t. At the beginning of period $t+1$, we eliminate the $\underline{X_t}$ grid nodes at the bottom of the ladder with mass $\mathcal{A}_{n,t} < 10^{-10}$. Similarly, we eliminate the $\bar{X}_t$ grid nodes at the top of the ladder with mass $\mathcal{A}_{n,t} < 10^{-10}$. We then add back the deleted mass (for rungs $n \leq \underline{X_t}$ and $n \geq N - \bar{X}_t$) proportionally to the surviving rungs. We then shift every grid point $n > \underline{X_t}$ down $\underline{X_t}$ rungs so that $\mathcal{A}_{n,t+1} = \mathcal{A}_{n+\underline{X_t},t}$ for all $n \geq 1$. Finally, we set $\mathcal{A}_{n,t+1} = 0$ for all "new" grid points at the top, i.e., for $n = N - \underline{X_t}$ to $n = N$. The minimum log TFP in period $t+1$ is therefore a_1. This procedure is repeated for every period t. To understand why this step of adjusting the grid for growth is appropriate, note that the costs and profits are invariant to scale. Therefore, the only relevant aspect for the optimal R&D decisions is the distribution above the current rung (as this determines the likelihood for successful imitation). Therefore, when the grid for the distribution is sufficiently fine and the distribution has converged to a stationary one, the number of rungs $\underline{X_t}$ dropped at the bottom and the number of rungs $\bar{X}_t$ dropped at the top remain constant over time.

Note that the rescaling of TFP does not affect firm TFP growth and the evolution of TFP distribution. The size of the state space is set to be sufficiently large to accommodate the simulated TFP distribution for all practical purposes. Specifically, we set N to 200 and the difference between rungs to $a_n - a_{n-1} = \tilde{a}/N_{\tilde{a}}$, where $N_{\tilde{a}} = 4$. Note that the larger is $N_{\tilde{a}}$ the finer is the state space grid.

We have verified that the results are robust to changing the grid size. Figure WA6 provides an illustrative example of this robustness exercise. The dotted lines replicate the simulated moments in our estimated PAM with $N_{\tilde{a}} = 4$. The dashed lines show the simulated moments in the same model with $N_{\tilde{a}} = 8$. The moments are essentially on the top of each other. The steady state TFP growth and variance of log TFP is 3.56% and 0.73, respectively, in the benchmark case ($N_{\tilde{a}} = 4$). They are 3.43% and 0.74 in the implementation where we assume $N_{\tilde{a}} = 8$.

Figure WA6: Theoretical moments for $N_{\tilde{a}} = 4$ and $N_{\tilde{a}} = 8$.



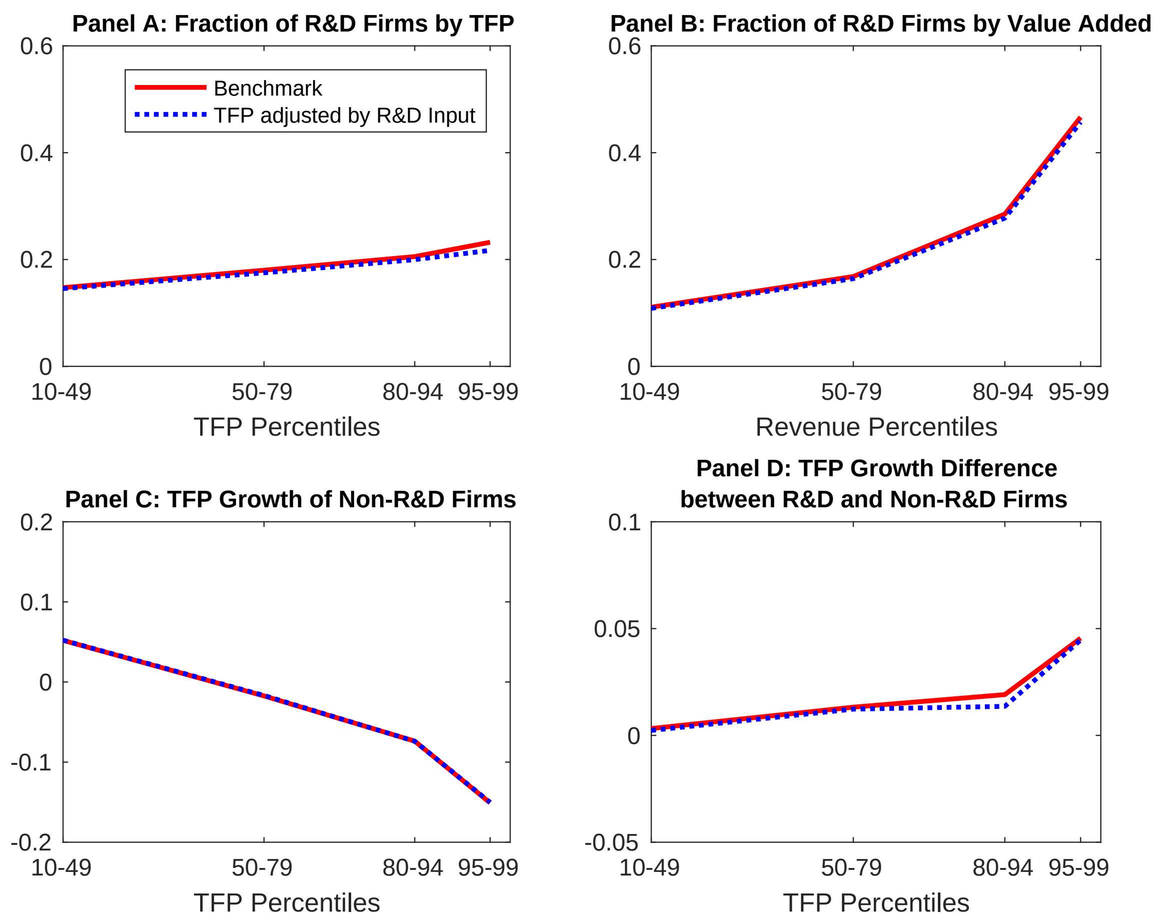
Note: The figure shows theoretical moments for the stationary distribution under two grid sizes—the benchmark $N_{\tilde{a}} = 4$ and the finer $N_{\tilde{a}} = 8$.

.

D TFP Mismeasurement due to R&D

In this section, we address a potential concern about mismeasurement of TFP in the presence of R&D investments. In the main analysis, we do not explicitly separate R&D expenditure when estimating TFP. This could bias the TFP estimates for R&D firms. The problem has no perfect solution because we do not have R&D data after 2007. To assess the importance of this bias, we adjust TFP in the earlier period 2001–07 by subtracting R&D expenditure from labor costs. Then, we plot a version of Figure 2 based on the adjusted data. See Figure WA7. The empirical moments are almost indistinguishable from those in Figure 2. We conclude that the problem is likely quantitatively small.

Figure WA7: Moments with Adjusted TFP Measurement, China 2001-07



Note: See description in the text.

E Details of the alternative TFP estimation based on Brandt et al. (2017)

The value added production function in the model is

$$P_{i,t}Y_{i,t} \propto A_{i,t}^{1-\frac{1}{\eta}} K_{i,t}^{\alpha\left(1-\frac{1}{\eta}\right)} L_{i,t}^{(1-\alpha)\left(1-\frac{1}{\eta}\right)}.$$

Our goal is to estimate α. Specifically, we estimate $\tilde{\beta}_k$ and $\tilde{\beta}_\ell$ in the following equation for each industry:

$$q_{i,t} = \tilde{\beta}_k k_{i,t} + \tilde{\beta}_\ell \ell_{i,t} + \omega_{i,t} + \epsilon_{i,t},$$

where $q_{i,t} \equiv \log(P_{i,t}Y_{i,t})$ is log value added, $k_{i,t} \equiv \log(k_{i,t})$, $\ell_{i,t} \equiv \log(L_{i,t}), \omega_{i,t} \equiv \left(1-\frac{1}{\eta}\right)\log\left(A_{i,t}\right), \epsilon_{i,t}$ is an i.i.d. shock to value added or simply measurement errors. Note that $\tilde{\beta}_k = \alpha\left(1-\frac{1}{\eta}\right)$ and $\tilde{\beta}_\ell = (1-\alpha)\left(1-\frac{1}{\eta}\right)$. We infer α from

$$\alpha = \frac{\tilde{\beta}_k}{\tilde{\beta}_k + \tilde{\beta}_\ell}$$

The methodology of Brandt et al. (2017)

Our approach follows closely Brandt et al. (2017) who estimate the following a gross output (as opposed to value added) production function for each manufacturing industry in China. Their method is based on Ackerberg et al. (2015), henceforth, ACF, and De Loecker and Warzynski (2012). They propose the following production function:

$$q_{i,t}^G = \tilde{\gamma}_k k_{i,t} + \tilde{\gamma}_\ell \ell_{i,t} + \tilde{\gamma}_m m_{i,t} + \omega_{i,t} + \epsilon_{i,t},$$

where $q_{i,t}^G$ is logarithm of gross output and $m_{i,t}$ is intermediate input. The simplest way of computing TFP by ACF estimates of output elasticities is to convert the elasticities in Brandt et al. (2017) to α in our model. Since $1-\tilde{\gamma}_m$ represents the value added share in the gross output, we have $\tilde{\gamma}_k = \alpha\left(1-\frac{1}{\eta}\right)(1-\tilde{\gamma}_m)$ and $\tilde{\gamma}_\ell = (1-\alpha)\left(1-\frac{1}{\eta}\right)(1-\tilde{\gamma}_m)$. Recall that α is the capital output elasticity in the value added production function. We then infer α from

$$\alpha = \frac{\tilde{\gamma}_k}{\tilde{\gamma}_k + \tilde{\gamma}_\ell}.$$

We refer to α converted from the estimates in Brandt et al. (2017) as BVWZ estimates.

Our estimation

There are two issues when we use BVWZ estimates. The first is that $\tilde{\gamma}_k$ and $\tilde{\gamma}_\ell$ are not directly estimated from the value added production function. Moreover, two non-trivial industries (ferrous metals and non-ferrous metals) are found to have negative $\tilde{\gamma}_k$ in Brandt et al. (2017). So, we directly estimate $\tilde{\beta}_k$ and $\tilde{\beta}_\ell$ in the value added production function. Our estimation procedure follows Brandt et al. (2017). In particular, we include tariffs as additional controls to identify TFP shocks, which Brandt et al. (2017) find vital for obtaining reasonable estimates. The only deviation from Brandt et al. (2017) is that instead of using intermediate inputs, we use firms' investment, $i_{i,t}$, as a proxy for their TFP.

$$i_{i,t} = i_t \left(\omega_{i,t}, \ell_{i,t}, k_{i,t}, w_{i,t}, e_{i,t}, \tau^I_{i,t-1}, \tau^O_{i,t-1}\right).$$

Here, we include log wage $w_{i,t}$ since we need to control for the serially correlated firm-specific shocks to the price of labour, which may affect the firm's optimal labour input and, in turn, investment choice.[44] The time-invariant function, $i_t(\cdot)$, will capture some fixed effects in the first-stage estimation below. Following Brandt et al. (2017), we also include firm's export status, $e_{i,t}$, input and output tariff, $\tau^I_{i,t-1}$ and $\tau^O_{i,t-1}$ (assuming that firms make production decision based on tariffs in the previous period). Under the assumptions that firm TFP, $\omega_{i,t}$, is the only unobserved firm-specific factor and that there exists a conditionally monotonic relationship between $\omega_{i,t}$ and $i_{i,t}$, we can rewrite the above equation as

$$\omega_{i,t} = h_t \left(i_{i,t}, \ell_{i,t}, k_{i,t}, w_{i,t}, e_{i,t}, \tau^I_{i,t-1}, \tau^O_{i,t-1}\right).$$

We further assume that the law of motion of firm TFP can be expressed as a function on the lagged TFP, output and input tariffs and firms' current export status:

$$\omega_{i,t} = g_t \left(\omega_{i,t-1}, e_{i,t}, \tau^I_{i,t-1}, \tau^O_{i,t-1}\right) + \xi_{i,t},$$

where $\xi_{i,t}$ is the innovation to firm TFP over and above what can be forecasted based on the variables $\left(\omega_{i,t-1}, e_{i,t}, \tau^I_{i,t-1}, \tau^O_{i,t-1}\right)$.

We can now write the procedure of estimating $\tilde{\beta}_k$ and $\tilde{\beta}_\ell$ in two steps. In the first stage, we estimate

$$q_{i,t} = \phi \left(\ell_{i,t}, k_{i,t}, i_{i,t}, w_{i,t}, e_{i,t}, \tau^I_{i,t-1}, \tau^O_{i,t-1}, \mathbf{Z}_{i,t}\right) + \epsilon_{i,t}.$$

Analogous to Brandt et al. (2017), we proxy the function ϕ by a third-order polynomial of capital, labor, and investment, and include the interactions of the polynomial terms with lagged industry-level input and output tariffs and firm-level wage and export status dummies. We also control for year, industry, and province fixed effects, which are all in the vector $\mathbf{Z}_{i,t}$, to obtain the estimate of expected output, denoted by $\hat{\phi}_{i,t}$. Given any $\tilde{\beta}_k$ and $\tilde{\beta}_\ell$, firm TFP, $\omega_{i,t}$, can be obtained by $\hat{\omega}_{i,t}\left(\tilde{\beta}_k, \tilde{\beta}_\ell\right) = \hat{\phi}_{i,t} - \left(\tilde{\beta}_k k_{i,t} + \tilde{\beta}_\ell \ell_{i,t}\right)$.

In the second stage, we assume that the law of motion of firm TFP, $g(\omega_{i,t-1}, e_{i,t}, \tau^I_{i,t-1}, \tau^O_{i,t-1}$, is linear:

$$\omega_{i,t} = c_0 + c_1 \omega_{i,t-1} + c_e e_{i,t} + c_O \tau^O_{i,t-1} + c_I \tau^I_{i,t-1} + \xi_{i,t}.$$

We regress the above equation by using $\hat{\omega}_{i,t}\left(\tilde{\beta}_k, \tilde{\beta}_\ell\right)$. The residual is $\hat{\xi}_{i,t}\left(\tilde{\beta}_k, \tilde{\beta}_\ell\right)$. We then use the following orthogonality condition

$$E\left(\hat{\xi}_{i,t}\left(\tilde{\beta}_k, \tilde{\beta}_\ell\right)\begin{pmatrix} k_{i,t} \\ \ell_{i,t-1} \end{pmatrix}\right) = 0$$

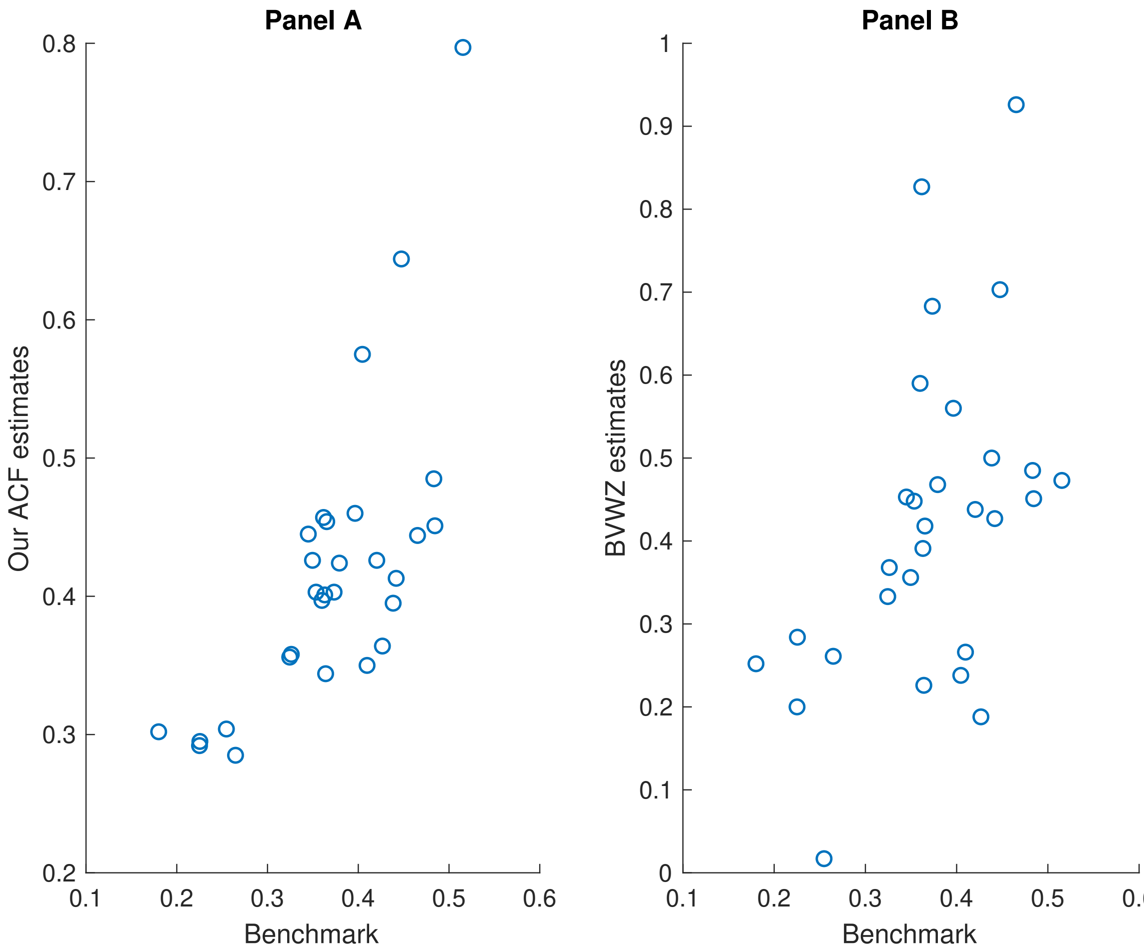
to estimate $\tilde{\beta}_k$ and $\tilde{\beta}_\ell$, applying the GMM algorithm in De Loecker and Warzynski (2012) and Brandt et al. (2017).

[44] The existence of the serially correlated firm-specific shocks also ensures that lagged labor is a valid instrument in the orthogonality condition below.

Results

Our ACF estimates of α are highly correlated with the benchmark estimates (see panel A of Figure WA8). The correlation coefficient is 0.71. As a further robustness check, we plot the BVWZ estimates of α in Panel B of Figure WA8. The correlation coefficient is 0.50.

Figure WA8: Estimates of α TFP Measurement



Note: See description in the text. For the two industries where BVWZ finds negative α, we use our estimates of α based on ACF.

We next use our ACF and BVWZ estimates of α to compute TFP. The empirical moments we target in the structural estimates are plotted in the supplementary appendix. The differences are very small.

F Multiple regressions with a more stringent criterion for innovative firms.

In this section, we present the results in Table 2 when one adopts the more stringent classification criterion for innovative firms in columns (5)-(6) of Table 3. The results are presented in Table 6.

Table 6: Robustness: Regressions of Table 2 with Stringent Classification of R&D Firms

PANEL A

Dependent variable: R&D decision in 2007.

	(1) R&D$_d$*high*	(2) R&D$_d$*high*	(3) R&D$_d$*high*	(4) R&D$_d$*high*
log(TFP)	0.0215***	0.160***	0.146***	0.121***
	(0.00503)	(0.0249)	(0.0224)	(0.0183)
wedge		-0.186***	-0.167***	-0.138***
		(0.0297)	(0.0263)	(0.0218)
export$_d$			0.0308***	0.0318***
			(0.00757)	(0.00744)
SOE$_d$				0.131***
				(0.0269)
Industry effects	+	+	+	+
Age effects	+	+	+	+
Province effects	+	+	+	+
Observations	109,799	109,799	109,799	109,799
R-squared	0.101	0.125	0.127	0.137

PANEL B

Dependent variable: TFP growth.

	(1) TFP growth	(2) TFP growth	(3) TFP growth
log(TFP)	-0.0604***	-0.0603***	-0.0608***
	(0.00368)	(0.00367)	(0.00359)
R&D$_d$*high*	0.0380***	0.0385***	0.0350***
	(0.00534)	(0.00509)	(0.00428)
export$_d$		-0.00351	-0.00452
		(0.00385)	(0.00356)
SOE$_d$			0.0326**
			(0.0112)
Industry effects	+	+	+
Age effects	+	+	+
Province effects	+	+	+
Observations	109,799	109,799	109,799
R-squared	0.0120	0.0120	0.121

Note: Panels A and B correspond to the regressions in Table 2 under a more stringent classification of R&D firms: only firms with R&D intensity above the median intensity (i.e., an R&D expenditure to value-added ratio of 1.73%) are labeled as R&D firms and firms with R&D intensity below the median are classified as nonR&D firms. The sample is the benchmark sample of fiems in China for the 2007-2012 period. See Table 2 for details.

G Regressions without Province and Age Dummies

In this section, we present regressions similar to those in Table 2 but without including controls (dummies) for firm age and for the province where the firm is located. All other aspects of the regressions are similar to those of Table 2. Tables 7 and 7 present, respectively, regressions without province dummies and without province and age dummies.

Table 7: Robustness: Regressions of Table 2 without Province Dummies

PANEL A

Dependent variable: R&D decision in 2007.

	(1)	(2)	(3)	(4)
	$R\&D_d$	$R\&D_d$	$R\&D_d$	$R\&D_d$
log(TFP)	0.0571***	0.357***	0.331***	0.289***
	(0.00699)	(0.0287)	(0.0261)	(0.0228)
wedge		-0.401***	-0.368***	-0.315***
		(0.0365)	(0.0328)	(0.0294)
export_d			0.0565***	0.0616***
			(0.0134)	(0.0133)
SOE_d				0.214***
				(0.0259)
Industry effects	+	+	+	+
Age effects	+	+	+	+
Observations	109,799	109,799	109,799	109,799
R-squared	0.130	0.193	0.196	0.211

PANEL B

Dependent variable: TFP growth.

	(1)	(2)	(3)	(4)	(5)
	TFP growth	TFP growth	TFP growth	TFP growth	TFP growth
log(TFP)	-0.0594***	-0.0591***	-0.0594***	-0.0590***	-0.0592***
	(0.00351)	(0.00361)	(0.00352)	(0.00367)	(0.00357)
$R\&D_d$	0.0305***	0.0345***	0.0302***		
	(0.00491)	(0.00450)	(0.00370)		
export_d		-0.0225***	-0.0227***	-0.0226***	-0.0229***
		(0.00388)	(0.00379)	(0.00384)	(0.00376)
SOE_d			0.0361***		0.0359***
			(0.0125)		(0.0124)
R&D intensity_h				0.0371***	0.0331***
				(0.00636)	(0.00609)
R&D intensity_m				0.0413***	0.0364***
				(0.00775)	(0.00620)
R&D intensity_l				0.0255***	0.0215***
				(0.00350)	(0.00306)
Industry effects	+	+	+	+	+
Age effects	+	+	+	+	+
Observations	109,799	109,799	109,799	109,799	109,799
R-squared	0.083	0.086	0.087	0.086	0.088

Note: Panels A and B correspond to a version of the regressions in Table 2 in which there are no dummies for the province where the firm is located.

Table 7: Robustness: Regressions of Table 2 without Province and Age Dummies

PANEL A

Dependent variable: R&D decision in 2007.

	(1) $R\&D_d$	(2) $R\&D_d$	(3) $R\&D_d$	(4) $R\&D_d$
log(TFP)	0.0579***	0.404***	0.377***	0.319***
	(0.00768)	(0.0299)	(0.0274)	(0.0228)
wedge		-0.463***	-0.429***	-0.356***
		(0.0384)	(0.0348)	(0.0296)
$export_d$			0.0561***	0.0633***
			(0.0135)	(0.0133)
SOE_d				0.243***
				(0.0287)
Industry effects	+	+	+	+
Observations	109,799	109,799	109,799	109,799
R-squared	0.081	0.180	0.183	0.204

PANEL B

Dependent variable: TFP growth.

	(1) TFP growth	(2) TFP growth	(3) TFP growth	(4) TFP growth	(5) TFP growth
log(TFP)	-0.0605***	-0.0601***	-0.0603***	-0.0600***	-0.0602***
	(0.00358)	(0.00369)	(0.00362)	(0.00375)	(0.00367)
$R\&D_d$	0.0282***	0.0341***	0.0281***		
	(0.00612)	(0.00559)	(0.00401)		
$export_d$		-0.0282***	-0.0288***	-0.0284***	-0.0290***
		(0.00397)	(0.00377)	(0.00392)	(0.00374)
SOE_d			0.0396***		0.0394***
			(0.0144)		(0.0142)
R&D $intensity_h$				0.0353***	0.0300***
				(0.00691)	(0.00626)
R&D $intensity_m$				0.0417***	0.0349***
				(0.00895)	(0.00653)
R&D $intensity_l$				0.0258***	0.0200***
				(0.00461)	(0.00340)
Industry effects	+	+	+	+	+
Observations	109,799	109,799	109,799	109,799	109,799
R-squared	0.070	0.074	0.076	0.074	0.077

Note: Panels A and B correspond to a version of the regressions in Table 2 in which there are no age and province dummies.