Supplement to "Innovation vs. imitation and the evolution of productivity distributions"

Michael D. König[a], Jan Lorenz[b], Fabrizio Zilibotti[a]

[a]Department of Economics, University of Zurich, Schönberggasse 1, CH-8001 Zurich, Switzerland.

[b]Department of Psychology and Methods, Focus Area Diversity, Jacobs University Bremen, Campus Ring 1, D-28759 Bremen, Germany.

B. Online Technical Appendix

B.1. Model Extensions

In this appendix we sketch three possible extensions of our model. First, in Appendix B.1.1 we allow for productivity shocks that can also lead to a decline in the productivity of a firm [cf. Klette and Kortum, 2004]. Next, in Appendix B.1.2 we provide a basic mechanism for firm entry and exit. Finally, Appendix B.1.3 introduces an absorptive capacity limit with an upper cutoff which bounds the relative productivity a firm can imitate from above.

B.1.1. Evolution of the Productivity Distribution with Decay

In this section we extend the model in the sense that firms not only exhibit productivity increases due to their innovation and imitation strategies but they are also exposed to possible productivity shocks, if e.g. a skilled worker leaves the company or one of their patents expires, leading to a decline in productivity. Specifically, we assume that in each period t a firm exhibits a productivity shock with probability $r \in [0,1]$ and this leads to a productivity decay of δ [cf. Klette and Kortum, 2004]. Otherwise, the firm tries to increase its productivity through innovation or imitation as discussed in the previous sections. If firm i with log-productivity $a_i(t)$ experiences a productivity decay in a small interval $\delta t = 1/N$ then her log-productivity at time $t + \Delta t$ is given by $a_i(t + \Delta t) = a_i(t) - \delta$, where $\delta \geq 0$ is a non-negative discrete random variable. Denoting by $\mathbb{P}(\delta = 1) = \delta_1$, $\mathbb{P}(\delta = 2) = \delta_2, \ldots$, we can introduce the matrix

$$\mathbf{T}^{\text{dec}} = \begin{pmatrix} 0 & 0 & \ldots & & \\ \delta_1 & -\delta_1 & 0 & \ldots & \\ \delta_2 & \delta_1 & -\delta_1 - \delta_2 & 0 & \ldots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}.$$

The evolution of the log-productivity distribution in the limit of large N is then given by

$$\frac{\partial P(t)}{\partial t} = P(t)\left((1-r)\left((\mathbf{I} - \mathbf{D})\mathbf{T}^{\text{in}} + \mathbf{D}\mathbf{T}^{\text{im}}(P(t))\right) + r\mathbf{T}^{\text{dec}} - \mathbf{I}\right). \quad (1)$$

B.1.2. Firm Entry and Exit

We assume that at a given rate $\gamma \geq 0$, new firms enter the economy with an initial productivity $A_0(t) = A_0 e^{\theta t}$, $A_0, \theta > 0$. The productivity $A_0(t)$ corresponds to the knowledge that is in the public domain and is freely accessible.[1] A higher value of θ corresponds to a weaker intellectual property right protection. $A_0(t)$ can also represent the technological level achieved through public R&D. New firms can start with this level of productivity when entering. Moreover, we assume that incumbent firms cannot have a productivity level below $A_0(t)$. Finally, we assume that incumbent firms exit the market at the same rate γ as new firms enter, keeping a balanced in- and outflow of firms. This means that a monopolist in sector i at time t is replaced with a new firm in that sector that starts with productivity $A_0(t)$.[2]

We assume that in each period, first, a randomly selected firm either decides to conduct in-house R&D or imitate other firms' technologies and, second, entry and exit takes place. Both events happen within a small time interval $[t, t + \Delta t)$. We then have to modify Equation (8) accordingly. In the case of $A_0 = 1$ we can write in the limit of large N

$$\frac{\partial P(t)}{\partial t} = (1 - \gamma - \theta t) P(t) \Big((\mathbf{I} - \mathbf{D}) \mathbf{T}^{\mathrm{in}} + \mathbf{D} \mathbf{T}^{\mathrm{im}}(P(t)) - \mathbf{I} \Big) + (\gamma - \theta t - 1) Q,$$

where $Q = (1 \quad 0 \quad 0 \quad \ldots)$.

B.1.3. Absorptive Capacity Limits with Cutoff

We assume that imitation is imperfect and a firm i is only able to imitate a fraction $D \in (0, 1)$ of the productivity of firm j.

$$A_i(t + \Delta t) = \begin{cases} A_j(t) & \text{if } A_j / A_i \in \;]1, 1 + D], \\ A_i(t) & \text{otherwise.} \end{cases} \tag{2}$$

Thus, the productivity of j is copied only if it is better than the current productivity A_i of firm i, but not better than $(1 + D)A_i$. We call the variable D the *relative absorptive capacity limit*. Taking logs of Equation (2) governing the imitation process reads as

$$a_i(t + \Delta t) = \begin{cases} a_j(t) & \text{if } a_j - a_i \in]0, d], \\ a_i(t) & \text{otherwise.} \end{cases} \tag{3}$$

We have introduced the variables $d = \log(1 + D)$. For small D it holds that $d \approx D$. The variable d is called the *absorptive capacity limit*.

We now consider the potential increase in productivity due to imitation and the associated transition matrix $\mathbf{T}^{\mathrm{im}}$. Following equation (2) we assume that a firm with a log-productivity of $a(t)$ can only imitate those other firms with log-productivities in the

[1]In contrast, any technology corresponding to a productivity level above $A_0(t)$ embodied in a firm is protected through a patent and is not accessible by any other firm. Firms can imitate other technologies, but only if they are within their absorptive capacity limits.

[2]Similarly, Melitz [2003] assumes that firms can be hit with a bad productivity shock at random and then are forced to leave the market.

interval $[a(t), a(t)+d]$. In this case $\mathbf{T}^{\text{im}}$ depends only on the current distribution of log-productivity $P(t)$ and simplifies to

$$\mathbf{T}^{\text{im}} = \begin{pmatrix} S_1(P) & P_2 & \dots & P_{1+d} & 0 & & \dots \\ 0 & S_2(P) & P_3 & \dots & P_{2+d} & 0 & \dots \\ & 0 & S_3(P) & P_4 & \dots & P_{3+d} & \dots \\ & & \ddots & \ddots & \ddots & \dots & \ddots \end{pmatrix},$$

with $P_b = P(b,t)$ and $S_b(P) = -P_{b+1} - \dots - P_{b+d}$. For the initial distribution of log-productivity $P(0)$, the evolution of the distribution is governed by

$$\frac{\partial P(t)}{\partial t} = P(t)\left((\mathbf{I}-\mathbf{D})\mathbf{T}^{\text{in}} + \mathbf{D}\mathbf{T}^{\text{im}}(P(t)) - \mathbf{I}\right),$$

where similar to the previous sections we have assumed that $\Delta t = 1/N$ and taken the limit $N \to \infty$.

B.2. Empirical Productivity Distributions

In this section, we present some empirical results about the productivity distribution across firms. We emphasize three features that are consistent with our theory. First, the distribution of high-productivity firms is well described by a power-law. Second, the distribution of low-productivity firms is also well approximated by a power-law, although this approximation is less accurate, arguably due to noisy data at low productivity levels. Third, the distribution is characterized by a constant growth rate over time, where both the right and the left power-law tails are fairly stable. This implies that the evolution over time of the productivity distribution can be described as a "traveling wave" (cf. Definition 1). While the first property is well known [see e.g. Corcos et al., 2007], the second and the third have not been emphasized in the literature.

We computed the empirical productivity levels of firms using the Amadeus database provided by Bureau van Dijk. We extracted a data set which contains a total of 5, 216, 989 entries from European firms in the years from 1992 to 2005. These were the firms for which data was available for all the variables from the following list: value added, operating revenue, fixed assets, number of employees, cost of materials and cost of employees. These were data points from 1, 413, 487 firms. As the model does not include entry and exit of firms we used a balanced subsample of all firms for which data exists in the years 1995 to 2003. We chose the time span 1995 to 2003, because these were the years with a substantial number of firms for which data exists in all years. In this balanced panel were 52, 837 firms but the coverage of firms across countries is quite heterogeneous, therefore we refrained our analysis to France where the coverage was is sufficiently adequate (with a total of 17, 404 French firms).

The productivity A_{it} for each firm i was estimated following the method introduced by Levinsohn and Petrin [2003]. We use the `STATA` implementation `levpet` explained in Petrin et al. [2004] to predict productivity values. The variables for this estimate from the Amadeus database were value added, fixed assets, number of employees and costs of materials. The variables operating revenue and cost of employees were used for robustness checks only. We follow Petrin et al. [2004] (the "value-added case"; see Section

year	λ	>mean(A)	$R^2(\lambda)$	ρ	<geomean(A)	$R^2(\rho)$
1995	3.80	35.2%	0.99	2.13	51.7%	0.97
1996	3.85	35.0%	0.99	2.50	51.8%	0.99
1997	3.77	34.6%	1.00	2.52	52.4%	0.98
1998	3.79	35.0%	0.99	2.54	52.3%	0.98
1999	3.77	34.7%	0.99	2.55	52.4%	0.99
2000	3.72	34.0%	0.99	2.31	52.9%	0.97
2001	3.71	34.2%	1.00	2.43	52.4%	0.98
2002	3.67	33.5%	0.99	2.26	52.3%	0.97
2003	3.53	33.0%	0.99	1.99	52.1%	0.96
average	**3.73**			**2.36**		

Table 1: The estimated power-law exponents for the right and left tail of the probability density function λ and ρ. The percentage of firms on which the regression is computed is shown as well as the corresponding coefficient of determination R^2.

2.1 in Petrin et al. [2004]) and Corcos et al. [2007] in our selection of these quantities for the estimation procedure. The production function underlying the method of Levinsohn and Petrin [2003] is more general than the one in our simple model introduced in Section 2. We decided to use this simple model to keep our theoretical analysis tractable, and to focus on the main driving forces underlying the innovation and imitation process. In this empirical section we consider a more general production function in order to make full use of the available data and to obtain unbiased estimates for productivity from this data. From our balanced sample of 17,404 French firms we obtained an average productivity of 53.82 in the year 1995 to an average productivity of 65.30 in the year 2003 (see also Figure 1, right panel). Further, we find that the standard deviation of log-productivity is quite stable, ranging from 1.61 to 1.67. Moreover, as Figure 1 illustrates, we observe that the left and right tails of the distributions are well approximated by power-laws, $P(A) \propto e^{\rho A}$ for small A and $P(A) \propto e^{-\lambda A}$ for large A. Table 1 shows the estimated values for ρ and λ. We observe that the exponents remain relatively stable over the years of observation. The estimated right tail exponent is around $\lambda = 3.73$ while the left tail exponent is around $\rho = 2.36$.

Moreover, the rightward shift in empirical distributions over the years of observation show a yearly increase in the average productivity (cf. Figure 1). We find that average productivity grows exponentially with time at a rate ν. We then compute the growth rate of average productivity ν from the data by estimating the parameters of an exponential growth function of the mean of productivity. Exponential growth of productivity corresponds to linear growth of log-productivity, that is, $\text{mean}(\log(A))(t) = \nu t + \text{const.}$. From our sample we estimate $\nu = 0.0271$, by linear regression on the logarithms of the values of the right panel in Figure 1.

B.3. Calibration of the Model's Parameters

The goal of this section is to calibrate the model's parameters given by the innovation success probability p and the imitation success probability q, such that the empirically observed right tail exponent λ and the growth rate of the traveling wave ν can be reproduced.

Our theoretical results on the computation of λ, ν and ρ cover only parts of the (p, q)-parameter space. Further on, the interdependence we know is quite complex and non-linear. Thus, a simple regression estimation procedure is ruled out.

We developed a hands on method to estimate λ, ρ and ν for computed trajectories of Equation (17) with parameters p and q, based on some heuristics which we derived from thorough observations. The method works as follows: Start with initial distribution $P_0 = (1, 0, \dots)$ on a long enough vector (we used length 30). All distributions mentioned here are handled as pdf's. Decide on an appropriate $T_{\max}$ and compute the distributions numerically (with `Matlab`'s ODE solver `ode45`) along the trajectory at time steps $t = 0, 1, 2, \dots, T_{\max}$. Heuristics for the choice of $T_{\max}$ where experimentally quantified such that the peak of the distribution at $T_{\max}$ lies well in the center of the support of P_0.

We then compute the arithmetic mean of productivity and the geometric mean of productivity for the distribution in each time step t. The arithmetic means build the lower bounds for the support of the distribution where λ is fitted by linear regression on the logarithm of productivity and the logarithm of the distribution function. The geometric means build the upper bounds for the support of the distribution where ρ is fitted by linear regression on the logarithm of productivity and the logarithm of the distribution function. Support for fitting was further restricted to the region where the distribution function was larger than a certain accuracy to avoid distortion from border effects which appear when floating point precision achieves its limits. Based on this we are able to fit λ and ρ for each time step t. We compute an estimate for ν for each time step t by looking at the differences in average log-productivity for time step t and $t - 1$.

We observed that for large enough $T_{\max}$ the fitted values stabilize, but some regular fluctuations remained due to the discreteness of the support of the distribution. To minimize the effect we averaged several values of λ and ρ along an interval of values of t of a certain length until $T_{\max}$. We found reasonable heuristics for assigning such a "wavelength" that the slight fluctuations could be averaged out well.

Based on this calibration method we computed values of λ, ρ and ν for the theoretical distributions of the ODE as a function of p and q on the grid $p = 0.001, \overset{+0.0002}{\dots}, 0.014$ and $q = 0.04, \overset{+0.002}{\dots}, 0.16$. After computation of the field we improved accuracy of the grid (using `Matlab`'s function `interp2`). We improved the accuracy of p to steps of length 0.000025 and the accuracy of q to steps of length 0.00025. Within this grid we computed the values of p and q which minimized the quadratic difference of empirical and theoretical λ plus the quadratic difference of the empirical and theoretical ν. This procedure yields the calibrated values for (p, q) of $(0.0049, 0.106)$.

B.4. Growth, Inequality and Policy Implications

Our model is parsimoniously parameterized by the in-house innovation probability $p \in [0, 1]$ and the parameter $q \in [0, 1]$ measuring the absorptive capacity of the firms in the

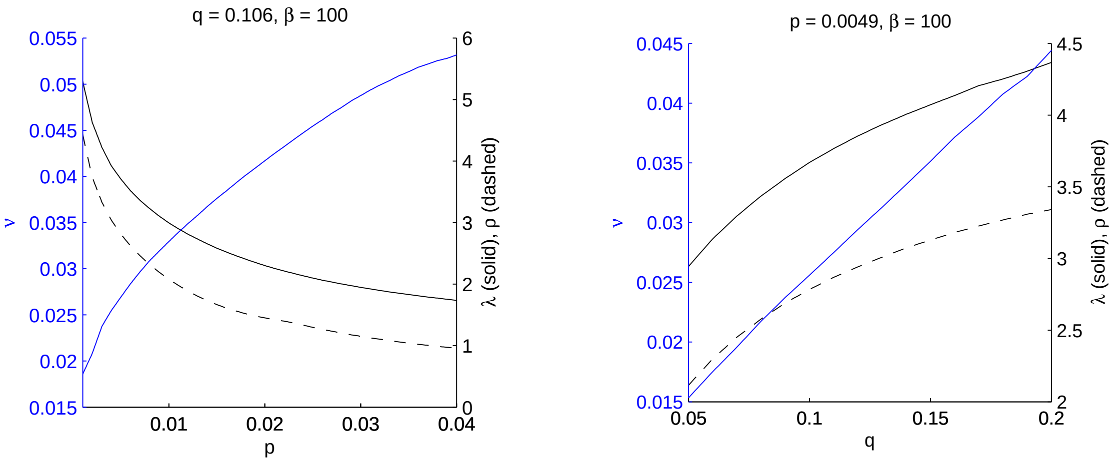

Figure B.1: Plots of λ, ρ and ν for p (resp. q) when q (resp. p) is fixed to the value from the calibrated parameters for France illustrated in Figure B.2 (color online).

economy. In this section we study the effects of each of the three parameters (p, q) on (i) the speed of growth and (ii) the inequality implied by the productivity distribution. This will allow us to analyze the effects of R&D policies that impact the innovation success probability p and the imitation success probability q. Examples for the first are R&D subsidy programs that foster the development of in-house innovations while policies that weaken the intellectual property protection regime (and hence make it easier to imitate others' technologies) are examples for the latter.

We first turn to the analysis of industry performance and efficiency. An industry has a higher performance, measured in aggregate intermediate goods and final good production, if it has a higher average log-productivity.[3] Equivalently, this corresponds to a higher average log-productivity per unit of time, as measured by the growth rate ν. We do this for two possible cases: (a) we keep the value of the absorptive capacity parameter q at its calibrated value of 0.106 and analyze the impact of changes in the innovation success probability p, or (b) we set p to its calibrated value of 0.0049 and study the effects of a change in q (see also Appendix B.3 for the calibration of these parameters). The results are shown in Figure B.1 (color online). In case (a) in Figure B.1 (left panels) we find that an increase in the innovation success probability p increases ν and hence accelerates growth. Thus, an R&D subsidy program which increases firms' in-house R&D success probability p leads to a higher growth rate of the economy. A similar analysis, but with varying values of the absorptive capacity (i.e. the imitation success probability q) in case (b) is shown in Figure B.1 (right panels). The figure reveals that an increase in the absorptive capacity q always increases the growth rate ν. Thus, an implication of our model is that policies which positively affect the absorptive capacity q, for example by weakening the intellectual patent protection of incumbent technologies in an industry, can have a positive effect on the growth rate ν of the economy.

A complete numerical analysis of the growth rate ν for general values of q is shown

[3] We will consider the average productivity measured by the geometric mean $\mu = \sqrt[n]{A_1 A_2 \cdots A_n} = \left(\prod_{i=1}^n A_i\right)^{1/n}$, which is related to the arithmetic average of the log-productivity values via $\frac{1}{n}\sum_{i=1}^n a_i = \frac{1}{n}\sum_{i=1}^n \log A_i = \log \mu$. However, our results also hold for the arithmetic average of the productivity values.

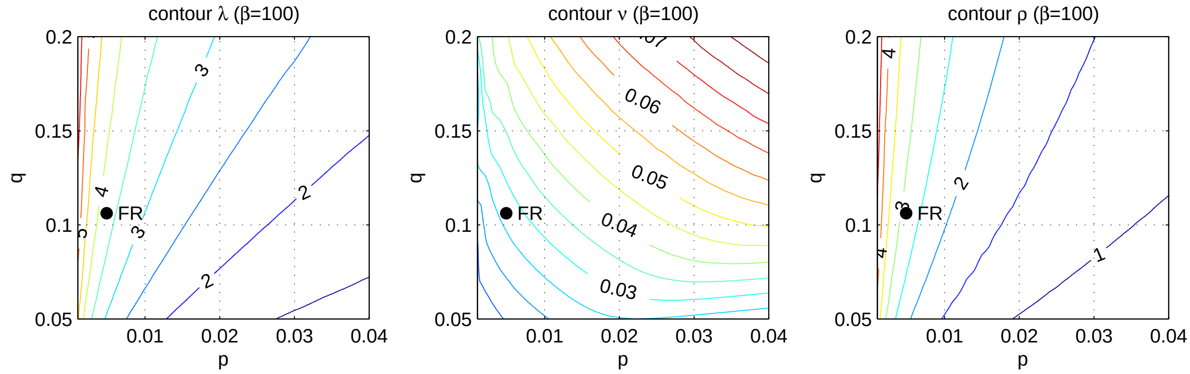

Figure B.2: Exploration of impact of innovation probability p, imitation probability q, on the dependent power-law parameters λ, and ρ, and on the productivity growth rate ν. The contour plots are based on numerical computation of solutions of the system of ODEs in Eq. (16). The black dots mark the calibrated (p, q)-points.

in Figure B.2 (middle panel). We observe that an increase in p or q leads to a higher growth rate ν.

Further, we can investigate the degree of inequality in the economy. As our measure of inequality we take the exponent λ of the right power-law tail of the distribution. A smaller value of λ corresponds to a more dispersed distribution with a higher degree of inequality. For both cases (a) and (b) we provide a numerical analysis in Figure B.1 (left panel). In case (a) we see that the exponent λ is always higher in the limit of strong productivity shocks and the difference increases with increasing innovation success probability p. However, in case (b) the reverse relationship holds: an increase in the absorptive capacity q yields a higher value of λ and thus reduces inequality.

We can draw the following conclusions from our counter factual analysis of the effects of each of the three parameters (p, q). First, we find that both types of policies, those that enhance the in-house innovation success probability p as well as those that facilitate the imitation and diffusion of existing technologies (increasing the value of q) increase the growth rate ν of the economy (cf. Figure B.1). However, while the first leads to an increase in inequality (smaller values of λ), the latter has the opposite effect of decreasing inequality (higher values of λ). It must be noted, however, that an economy in which technologies can easily be imitated (high q) but there is no in-house R&D ($p \to 0$) does not generate growth. Thus, a balanced approach is required, fostering both, the capacities of firms to generate innovations in-house and an environment in which these innovations can diffuse throughout the economy.

References

Corcos, Gregory, Massimo Del Gatto, Giordano Mion, and Gianmarco IP Ottaviano (2007). "Productivity and firm selection: intra- vs international trade." Université Catholique de Louvain, CORE Discussion Papers No. 2007060.

Klette, Tor Jakob, and Samuel Kortum (2004). "Innovating firms and aggregate innovation." *Journal of Political Economy*, 112(5):986–1018.

Levinsohn, James, and Amil Petrin (2003). "Estimating production functions using inputs to control for unobservables." *Review of Economic Studies*, 70(2):317–341.

Melitz, Marc J. (2003). "The impact of trade on intra-industry reallocations and aggregate industry productivity." *Econometrica*, 71(6):1695–1725.

Petrin, Amil, Brian P. Poi, and James Levinsohn (2004). "Production function estimation in Stata using inputs to control for unobservables." *Stata Journal*, 4(2):113–123.