R&D Decisions and Productivity Growth: Evidence from Switzerland and the Netherlands☆

Sabien Dobbelaere[a,b,c], Michael D. König[a,b,d,e], Andrin Spescha[d], Martin Wörter[d,f]

[a] *Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands.*
[b] *Tinbergen Institute (TI), Amsterdam, The Netherlands.*
[c] *Institute of Labor Economics (IZA), Bonn, Germany.*
[d] *ETH Zurich, KOF Swiss Economic Institute, Leonhardstrasse 21, 8092 Zurich, Switzerland.*
[e] *Centre for Economic Policy Research (CEPR), London, United Kingdom.*
[f] *Leibniz Centre for European Economic Research (ZEW), Germany.*

Abstract

We study why the fraction of R&D active firms decreased in Switzerland but increased in the Netherlands between 2000 and 2016, and how this affected productivity growth. Our structural growth model identifies the impact of innovation, imitation and R&D costs on firms' R&D decisions. In Switzerland, in-house R&D became harder and more costly, unlike in the Netherlands. Counterfactual analyses show that policies should prioritize enhancing innovation and imitation success over reducing R&D costs to boost productivity growth. Finally, using an optimal policy framework, we demonstrate that a small revenue tax can optimally finance R&D subsidies.

Key words: R&D, innovation, imitation, R&D costs, policy, productivity growth, traveling wave.
JEL: E61, E65, D22, O31, O47, O52.

☆We would like to thank Gianluca Antonecchia, Pere Arque-Castells, Eric Bartelsman, Doyne Farmer, Hans Gersbach, Mitsuru Igami, Hans Koster, Francois Lafond, José Luis Moraga-González, Mark Roberts, Zheng Michael Song, Steven Poelhekke, Mark Schankerman, Kjetil Storesletten, Jos van Ommeren, Fabrizio Zilibotti, and seminar/conference participants at Oxford University, Vrije Universiteit Amsterdam, the Swiss State Secretariat for Education, Research and Innovation, the KOF Swiss Economic Institute, the Tinbergen Institute Workshop on Innovation, the Mannheim Centre for Competition and Innovation Conference, the Royal Economic Society Conference, the Meeting of the European Economic Association, the Meeting of the European Association for Research in Industrial Economics, and the Comparative Analysis of Enterprise Data Conference for helpful comments. We are grateful to Statistics Netherlands (CBS) for providing access to the Dutch non-public micro-data.

Email addresses: `sabien.dobbelaere@vu.nl` (Sabien Dobbelaere), `m.d.konig@vu.nl` (Michael D. König), `spescha@kof.ethz.ch` (Andrin Spescha), `woerter@kof.ethz.ch` (Martin Wörter)

1. Introduction

The development of R&D activity in European countries is very different. We observe an increase in the share of R&D active companies in the Netherlands and France, in contrast to a declining share in Germany and Switzerland over the period 2000-2016 [Eurostat, 2019].

These diverging trends raise questions about the determinants of firms' R&D decision, the role of R&D decisions in explaining aggregate productivity growth and the effectiveness of innovation policies to revive productivity growth [cf. Goldin et al., 2024]. Can we expect lower growth rates if the share of R&D active companies declines in a country? And should governments consequently take policy action to counteract such decline?

To investigate these questions, our study focuses on two innovation leaders in Europe, Switzerland and the Netherlands, for which we have highly comparable firm-level datasets spanning the years 2000-2016. The selection of these countries is motivated by diverging trends in R&D activity but different approaches to innovation promotion. The Netherlands actively promotes engaging in R&D activities through a favorable corporate tax system (Innovation Box) and specific R&D incentives (R&D tax credit) that support innovation throughout the entire R&D life-cycle [Lokshin and Mohnen, 2012; Mohnen et al., 2017]. Switzerland, in contrast, predominantly uses indirect R&D measures such as the promotion of knowledge and technology transfer between universities and the private sector. Innovation grants are usually not awarded directly to private companies [SERI, 2020].[1]

We first provide descriptive evidence on R&D activity and firm productivity. We document that Swiss companies need increasingly higher productivity levels to stay R&D active whereas this productivity-innovation threshold decreases over time for Dutch companies. As such, fewer (more) but highly (less) productive innovators remain R&D active in Switzerland (the Netherlands). Moreover, we learn from a survival analysis that large, more productive, human-capital intensive and export-oriented firms are less likely to exit R&D in both countries. Innovation policy measures, either indirect ones such as public support for industry-science partnerships in Switzerland or direct ones such as the Innovation Box in the Netherlands, seem effective in keeping firms R&D active.

One could interpret the reduced-form effects of these policy instruments on firms' R&D investment decision through their impact on the difficulty of making innovations (in-house R&D success), R&D costs and technology diffusion (imitation). To disentangle the effects of these three fundamentals on firms' R&D decision and to solve potential endogeneity problems, we construct an endogenous growth model with random interactions (imitation) where firms decide to conduct in-house R&D or to imitate other firms by maximizing the expected profit in every period, following König et al. [2022]. We estimate the model using Simulated Method of Moments by targeting moments of the empirical distribution of R&D and productivity that are salient in the theory. Relying on a structural stability test following Hall [2005], and motivated by the post-2008 rise in R&D active firms in the Netherlands and the 2007 introduction of the Innovation Box, we compare pre- and post-2008 estimates for both

[1] Switzerland has only introduced a Patent Box promotion design in 2020. Innovation promotion is also pursued at the cantonal level. For example, most cantons offer tax relief for R&D expenditures. These measures were introduced after the period covered by this study.

countries to quantify changes in the fundamentals driving firms' R&D decisions.

We find that the estimated in-house R&D success probability decreased and the costs of doing R&D increased significantly in Switzerland in the post 2008-period. This is consistent with evidence from the Swiss Innovation Survey (see Spescha and Wörter [2022]), which reports increasing innovation costs for Swiss firms. The estimated imitation success probability rose for non-R&D firms but declined for R&D firms post-2008. Together, lower R&D success rates and higher R&D costs reduce the expected profit from innovation and, hence, weaker firms' incentives and propensity to conduct R&D, in line with empirical observations.

In the Netherlands, estimated innovation success and imitation success (for both R&D and non-R&D firms) rate rose, while R&D costs fell considerably post-2008, in contrast to Switzerland. Higher success rates and lower costs make innovation more attractive, ceteris paribus, increasing the share of R&D firms over time, consistent with empirical observations. Thus, the importance of R&D success and costs evolved in opposite directions in both countries, thereby generating divergent trends in the prevalence of R&D active firms. The essential question, however, is to what extent this divergence affects productivity growth.

To examine how different developments in the fundamentals driving firms' R&D decisions, and consequently, different shares of R&D active firms in the two countries affect productivity growth, we use the estimated structural growth model to run counterfactual simulations. These simulations allow to quantify how sensitive the growth rate is to changes in the model's fundamental parameters. For both countries, we find that policies targeted towards increasing the success probabilities of innovation and/or imitation are the most effective in boosting the productivity growth rate. Increasing the in-house R&D success probability, however, also amplifies productivity dispersion, leading to larger inequality. Conversely, increasing the likelihood of successful imitation (indicating positive diffusion and spillovers) reduces inequality. Policy instruments such as R&D subsidies or R&D tax credits are likely to reduce the innovation threshold by decreasing R&D costs. This leads to an increase in the number of R&D active, innovative companies, thereby weakly reducing productivity dispersion. However, such policies are found to contribute only moderately to increasing productivity growth.

In our policy experiment, we also simulate for each country a counterfactual in which it experiences the other country's observed changes in fundamental parameters from the pre- to the post-2008 period. We demonstrate that, had Switzerland experienced increasing innovation and imitation success probabilities and declining R&D costs similar to those observed in the Netherlands post-2008 –following the implementation of innovation support measures– its productivity growth rate would have been 40% higher. Conversely, absent this extended post-2008 innovation support in the Netherlands, Dutch productivity growth would have been 20% lower.

Finally, using an optimal policy framework akin to Akcigit et al. [2022], we show that a small revenue tax can be used to optimally finance R&D subsidies. Under the optimal tax and subsidy policy, Switzerland attains 102.74% of the welfare level as in a no-cost benchmark, while the Netherlands reaches roughly the same level as the benchmark. Real-world policies achieve between 84% and 97% of the benchmark welfare in Switzerland and the Netherlands, respectively, but at much higher levels of taxation. The welfare gains of the optimal policy exceeding the benchmark arise from reducing effective R&D costs and discouraging inefficient R&D, thereby reallocating innovation toward more productive firms [König et al., 2022].

Related literature. Our study relates to several lines of literature. First, we contribute to the productivity growth literature, which has intensively investigated declining productivity growth rates and increasing productivity dispersion [e.g. Decker et al., 2020; Goldin et al., 2024; Gordon, 2018]. Such productivity development has led to a growing concentration of economic activity in a smaller number of firms [Autor et al., 2017, 2020]. The existing literature has proposed various underlying causes ranging from "winners-take-it-all" technologies [Autor et al., 2020], a lack of diffusion of newly generated knowledge [Akcigit and Ates, 2021; Andrews et al., 2015], a lack of complementary assets [Andrews et al., 2016], an increasing difficulty to create new knowledge [Bloom et al., 2020; Jones, 2009; Park et al., 2023], higher fixed costs of innovation [Rammer and Schubert, 2016], to a spread of highly efficient firms over multiple markets [Aghion et al., 2023]. We add to this line of research by examining the impact of R&D activities on productivity growth and dispersion. Our model allows us to disentangle whether the difficulty of creating knowledge, innovation costs or knowledge spillovers are responsible for the divergent developments in Switzerland and the Netherlands.

Second, we contribute to the discussion on effective policy instruments for increasing the productivity effects of R&D activities. Bloom et al. [2019] evaluate the effectiveness of several policies aimed at promoting technological innovation. They find that R&D tax credits, skilled immigration, and trade and competition are highly effective. In contrast, direct R&D grants and augmenting local human capital are reported to be less effective, with Patent Boxes being considered the least effective. Except for skilled immigration and augmenting local human capital, all these policies tend to increase inequality across firms. Acemoglu et al. [2018] consider re-allocation effects of scarce R&D resources and find that R&D subsidies to incumbent firms do not necessarily increase welfare because they prevent less productive firms from leaving the R&D market, which would free up scarce innovation inputs for more productive firms. As a result, overall productivity growth falls. Akcigit et al. [2022] examine effective tax policy approaches to encourage firm innovation behavior, taking into account the allocation effects of scarce resources. To support companies that are more productive in the R&D process, R&D subsidies should be non-linear and decreasing in the level of firms' R&D expenditures. Dechezleprêtre et al. [2023] exploit a change in size-based eligibility thresholds for R&D tax relief in the UK and find significant effects of tax relief on R&D, patenting and productivity. Choi and Levchenko [2023] investigate the long-term impact of a large-scale industrial policy in South Korea. The policy subsidized the allocation of foreign credit to Korean heavy and chemical industry firms. The authors find that subsidized firms grew faster than those never subsidized for 30 years after subsidies ended. However, the vast majority of this long-term impact stems from the benefits of learning-by-doing rather than the relaxation of financial constraints. König et al. [2022] find that the level of R&D subsidies is important: subsidies resulting in a moderate increase in R&D investment accelerate productivity growth, whereas excessively high subsidies slow down productivity growth. The latter is due to inducing many firms to innovate, which would benefit more from imitation and the adoption of technologies.[2] We contribute to this strand of literature by estimating the effectiveness of policy measures aimed at fostering innovation implemented in the Netherlands and we apply this specific setting to Switzerland.

Finally, the magnitudes of the effect on the productivity growth rate that we identify when comparing the two policy environments in Switzerland and the Netherlands are comparable with previous studies in the literature. For example, Akcigit [2009] documents that an optimal homogeneous R&D subsidy rate in the US could boost growth up to 2.1%. Kogan et al. [2017] report that a one-standard-deviation increase in the innovation index would result in a 3.4% rise in aggregate productivity growth over a 5-year period in the US. König et al. [2019] find that providing a subsidy to R&D expenditures for firms engaged in R&D collaboration could lead to welfare gains between 1% and 4%. Moreover, König et al. [2022] show that the introduction of an industrial policy that supports more productive firms in China could increase the growth rate by 2 percentage points. Relatedly, Van Reenen and Yueh [2012] find that if China had not implemented its international joint venture policy with the aim of facilitating technology transfer (diffusion), the country's annual GDP growth would have been reduced by between half a percentage point to over one percentage point over the past 30 years. Hence, our findings regarding the growth impact of an innovation-supportive environment do not exceed the estimates obtained in the existing literature.

Organization of the paper. The paper is structured as follows. Section 2 describes the data and presents descriptive evidence on R&D activity and productivity. Section 3 introduces a structural model for the three fundamentals –in-house R&D success (innovation), R&D costs and technology diffusion (imitation)– that affect a firm's decision to conduct R&D. Section 4 reports the estimation results. Section 5 quantifies how the model's fundamental parameters affect productivity growth via counterfactual simulations and examines an optimal policy framework. Section 6 concludes. Additional relevant material can be found in the Supplementary Appendix.

2. Data and Descriptive Evidence

2.1. Data

Production and innovation data. We use highly comparable micro-data sets that are sourced from different surveys. For Switzerland, we use nine waves of the Swiss Innovation Survey (SIS) covering the period 2001-2016. The survey is collected by the KOF Swiss Economic Institute of ETH Zurich. The SIS is the equivalent of the well-known Community Innovation Survey (CIS)

[2]Conceptually, our study is closest to König et al. [2022], who estimate a similar structural model. We differ in four respects. First, we analyze two technologically leading European economies (e.g. in terms of patents per capita), Switzerland and the Netherlands, wheareas König et al. [2022] focus on China. Second, while R&D measurement error may be more severe in the Chinese data, this concern is attenuated in our survey micro-data. Third, unlike König et al. [2022], we exploit multiple survey waves, allowing us to estimate a panel data model without imposing strong stationarity assumptions on the data generating process. Fourth, we extend their productivity dynamics by embedding an optimal policy framework following Akcigit et al. [2022].

of the European Union. Between 2001 and 2010, the survey waves were conducted at three-year intervals. From 2010 onward, the survey waves have been carried out in two-year intervals in order to synchronize with the CIS. The SIS is based on a stratified random sample of firms with more than 5 employees, drawn from the Swiss business census. It is representative of the Swiss economy. The SIS includes all relevant industries in manufacturing, construction and services. Stratification is based on 2-digit industries and within each industry on three firm size classes. For this study, the SIS is restricted to the NACE Rev2 classification headings 10-33, 41-43 and 45-82. In addition to the common harmonized innovation indicators also present in the CIS, the SIS also asks firms about various other firm characteristics such as sales, number of employees, share of high-skilled employees, intermediate input costs and export activity.

For the Netherlands, we use several administrative data sets collected by Statistics Netherlands (CBS). The innovation variables stem from eight waves of the Dutch Community Innovation Survey covering the period 2000-2016. CIS enterprises are merged with data from the Production Surveys (PS). The latter contains data on production value, factor inputs and factor costs. The CIS and PS data are collected at the enterprise level. A combination of census and stratified random sampling is used for each wave of the CIS and PS. The stratification variables are the industry and the number of employees of an enterprise. A census is used for the population of enterprises with at least fifty employees and stratified random sampling is used for enterprises with fewer than fifty employees. This cut-off point of 50 employees is applied to each wave of the CIS and the PS. To define the skill type of each employee in Dutch firms, we use their education type reported in the Education database which comes from the Polis Administration and the Labour Force Survey ("Enquête BeroepsBevolking, EBB"). The Education database provides the highest level of education attained by an individual on October 1st of the year.[3] The education type is based on a 2-digit SOI-code (Dutch education classification, "Standaard Onderwijsindeling") and is converted to the ISCED classification (International Standard Classification of Education). To ensure comparability with the Swiss data, we only consider enterprises active in the manufacturing, construction and services industries. As such, we also use the NACE Rev2 classification headings 10-33, 41-43 and 45-82.

Innovation policy measure. To examine the impact of the main innovation policy instrument on R&D activity and productivity in the Netherlands, we match our PS-CIS data with data on Innovation Box usage collected by the tax office and supplied by Statistics Netherlands. The Innovation Box provides a reduced corporate income tax rate for profits generated from intangible assets. The idea is that lower taxes on future (expected) profits from R&D will make R&D investments more attractive. The policy measure started in 2007 as a Patent Box, lowering the corporate income tax rate from 20-25% to 5% for income derived from patented intangible assets. Throughout the years, several changes were introduced to make the policy more accessible to small and medium-sized firms.

[3] While coverage of this dataset is increasing over time, its sample is skewed towards younger, more educated workers, especially at the beginning of our time period. We use an imputation method to predict the skill level of individuals whose education data is missing. More specifically, we estimate an inverse Mincer equation to predict education levels for workers for whom education information is missing based on individual and firm characteristics in our matched employer-employee micro-data for each year during the period 2000-2016. For the full procedure, we refer to Bartelsman et al. [2015].

The largest change took place in 2010, when the Patent Box was turned into the Innovation Box as the formal patent requirement was removed. Since then, the reduced corporate income tax rate applies to profits arising from both patented and unpatented intangible assets. The most important eligibility criteria are conducting R&D activities that are formally recognized by the tax office and generating in-house R&D. Data on Innovation Box usage is defined at the level of tax units, which are legal entities used by tax authorities for the collection of corporate income tax. To match the Innovation Box usage data at the tax unit-year level to our firm-year panel, we use crosswalks provided by Statistics Netherlands. During the 2008-2016 period, about 14% of the firms that are in our matched PS-CIS sample are Innovation Box users. Among the Innovation Box users, 55% are manufacturing companies, 4% are construction companies and 41% are service companies.

The other innovation policy instruments in the Netherlands are the WBSO and the RDA,[4] for which we do not have data. These are two similar policy instruments. The WBSO was introduced in 1994 and reduces the wage costs of R&D workers through tax credits. The RDA was introduced in 2012 and is more general than the WSBO, as it encompasses tax credits for non-labor related R&D expenses. Thus, the Dutch innovation support environment that we measure in Section 4 extends beyond the scope of the Innovation Box instrument only. The WBSO improved the financial situation of firms already in the pre-2008 period. In contrast, the RDA, along with the Innovation Box, enhanced the financial situation of firms in the post-2008 period.

Estimation samples. For estimation purposes, we use information from the aforementioned SIS waves (Switzerland) and matched PS-CIS sample (the Netherlands). After some basic cleaning and deleting missing firm-year observations on real value added per worker (our main productivity measure) and engagement in R&D activity,[5] we end up with an unbalanced panel of 16,553 observations corresponding to 6,201 enterprises (48.2% in manufacturing, 9.5% in construction, and 42.1% in services) over the period 2000-2016 in Switzerland. For the Netherlands, we have an unbalanced panel of 43,341 observations corresponding to 21,624 enterprises (29.4% in manufacturing, 9.2% in construction and 61.4% in services) over the period 2000-2016.

2.2. Descriptive Evidence on R&D Activity and Productivity

In the following, we present descriptive evidence on the fraction of R&D active companies, the development of the productivity distribution and the innovation (R&D) decision of firms in Switzerland and the Netherlands.

[4] WBSO is an acronym for "The Wage Tax and Social Insurance Act" ("Wet Bevordering Speur- en Ontwikkelingswerk" and RDA stands for R&D tax Allowance.

[5] We use firms' reported indicators for in-house or outsourced R&D to define their R&D decision, in contrast to König et al. [2022], who rely on reported R&D expenditures (the intensive margin). We focus on the extensive margin for two reasons: (i) it aligns with the discrete choice model we estimate and is required for our counterfactual experiments, and (ii) it is less subject to measurement error, tax evasion, and manipulation than the intensive margin. Moreover, only a small share of firms undertake any R&D and the distribution of R&D expenditures is highly skewed, which might lead to difficulties in the estimation of the intensive margin. Further, note that in our data the R&D status refers to R&D reported at the domestic enterprise level, regardless of ownership.

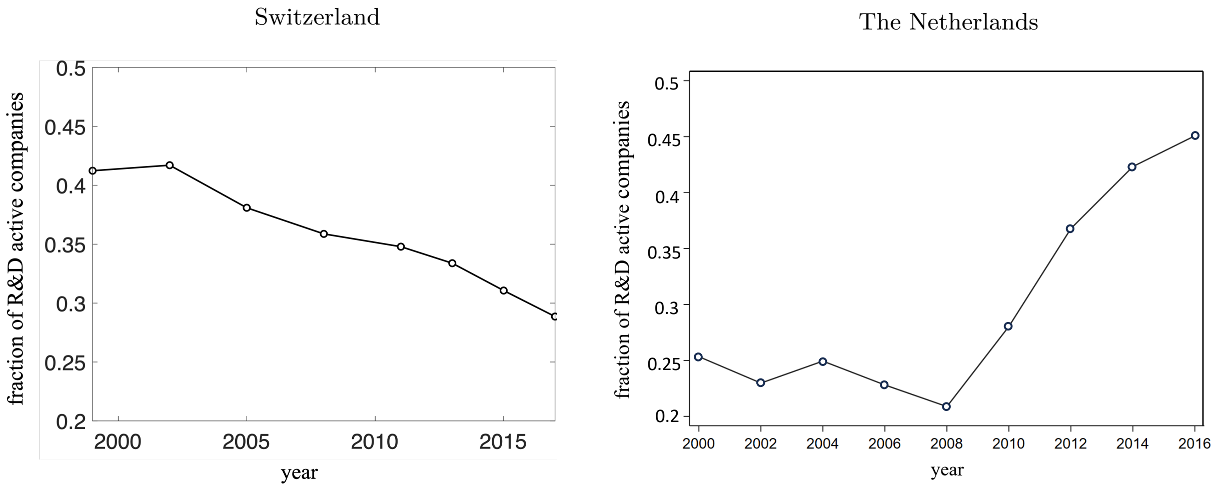

Figure 1: Fraction of R&D active firms over the years 2000-2016 in Switzerland (left panel) and the Netherlands (right panel).

Figure 1 shows the share of R&D active firms in Switzerland (left panel) and the Netherlands (right panel), where R&D active companies are defined as those engaging in either in-house or outsourced R&D activities.[6] In Switzerland, the number of R&D active firms sharply decreases from more than 41% at the beginning of the 2000s to less than 30% in 2016. In the Netherlands, we observe a completely different development: the share of R&D active firms ranges between 20-25% during the 2000-2008 period but sharply increases afterwards and reaches 45% in 2016.[7,8]

Contrary to the development of R&D activity, productivity (measured as real value added per employee) evolves similarly in both countries during our considered time span. Figure 2 shows the log-productivity distribution in Switzerland (left panel) and the Netherlands (right panel), respectively.[9] The distribution is computed for different years using a Kernel smoothing procedure. In

[6]The composition of R&D activities has remained relatively stable over time in both countries. The relationship between domestic R&D and R&D conducted abroad has remained steady for the past 20 years in Switzerland [Spescha and Wörter, 2022].

[7]The observed differences in the share of R&D performing firms in the two countries could be related to different industrial structures. For example, countries with a substantial high-tech sector may experience a lower likelihood of reduced rates of R&D diffusion than countries with a large service sector. This is indeed crucial for making cross-sectional descriptive comparisons between countries. However, it is less important for the analysis presented in this paper, where our primary focus is on studying variations over time within a country. Furthermore, the industrial structure in Switzerland and the Netherlands did not change significantly during our observation period.

[8]Moreover, the share of R&D expenditures in sales shows an upward trend in both countries. This implies an increasing concentration of more R&D expenditures in an ever smaller number of R&D active firms in Switzerland while firms in the Netherlands are not only more frequently pursuing R&D but are also increasing their respective R&D expenditures.

[9]For a more accurate comparison, log productivity in the Netherlands has been adjusted by a constant to align with average log productivity in Switzerland at the beginning of our observation period.

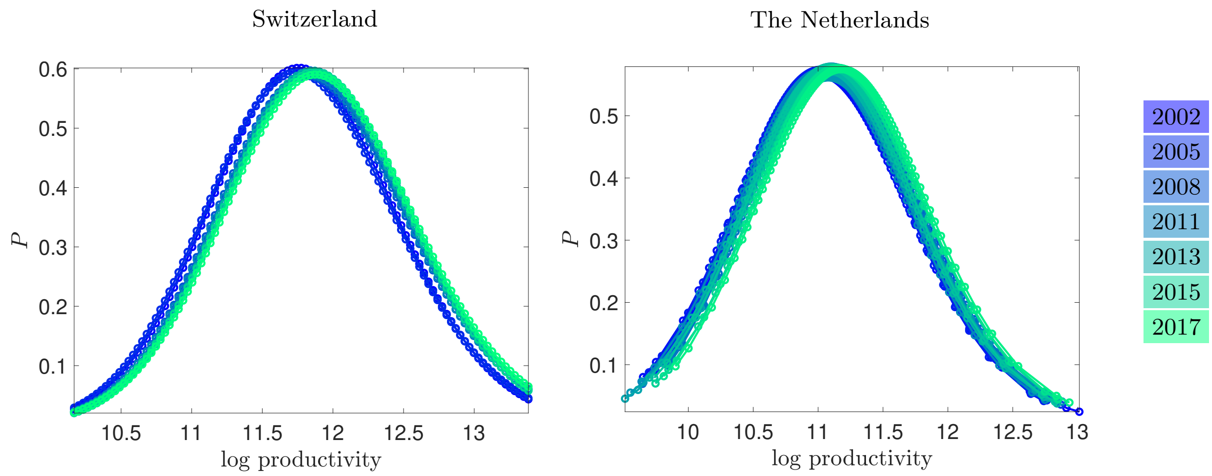

Figure 2: Log-productivity distribution (P) in Switzerland (left panel) and the Netherlands (right panel). Productivity is measured as real value added per employee (in full-time equivalents).

both countries, the distribution has the shape of a "traveling wave" [cf. König et al., 2022, 2016], with the peak of the distribution moving to the right, indicating an increase in the average productivity level over time. This is consistent with Figure 3 which shows the yearly variation of the average and median log productivity in Switzerland (left panel) and the Netherlands (right panel), respectively. In both countries, average and median productivity levels slightly increase over time. In the Netherlands, we observe a slightly steeper increase in the post-2008 period, while in Switzerland we see a decline in productivity growth after 2008.[10] When considering the productivity growth rates in both countries in the bottom panels in Figure 3, we observe that the growth rate is declining on average in Switzerland and increasing on average in the Netherlands. We investigate potential causes of these diverging trends in Section 3 and beyond.

Let us now look at the correlation between a firm's innovation decision (i.e. whether a firm conducts R&D or not) and its productivity level. The left panels of Figures 4 and 5 show firms' innovation decision plotted over their productivity levels in Switzerland and the Netherlands, respectively. In both countries, we see that firms with higher productivity tend to innovate more. The correlation between firms' innovation decision and productivity pooled across all years equals 0.75 in Switzerland and 0.90 in the Netherlands. The dashed line indicates the innovation-productivity 50%-threshold, that is, the lowest productivity level at which the likelihood of conducting R&D is more than 50%. The right panels of Figures 4 and 5 show the evolution of this innovation-productivity

[10]Note that there are minor differences compared to the official statistics of the respective countries, see for example "OECD (2023), GDP per hour worked (indicator). doi: 10.1787/1439e590-en (Accessed on 11 October 2023)". This may be due to different samples. For example, for Switzerland, the KOF data refer to companies with more than 5 employees (in full-time equivalents) and the value added statistics on which the official data are based refer to companies with more than 3 employees (see Federal Statistical Office, Production and Value-Added Statistics, Factsheet). For the Netherlands, official figures based on the OECD database encompass very small businesses (see e.g. Grabska et al. [2017]). In contrast, our estimation sample is restricted to firms having at least 5 employees.

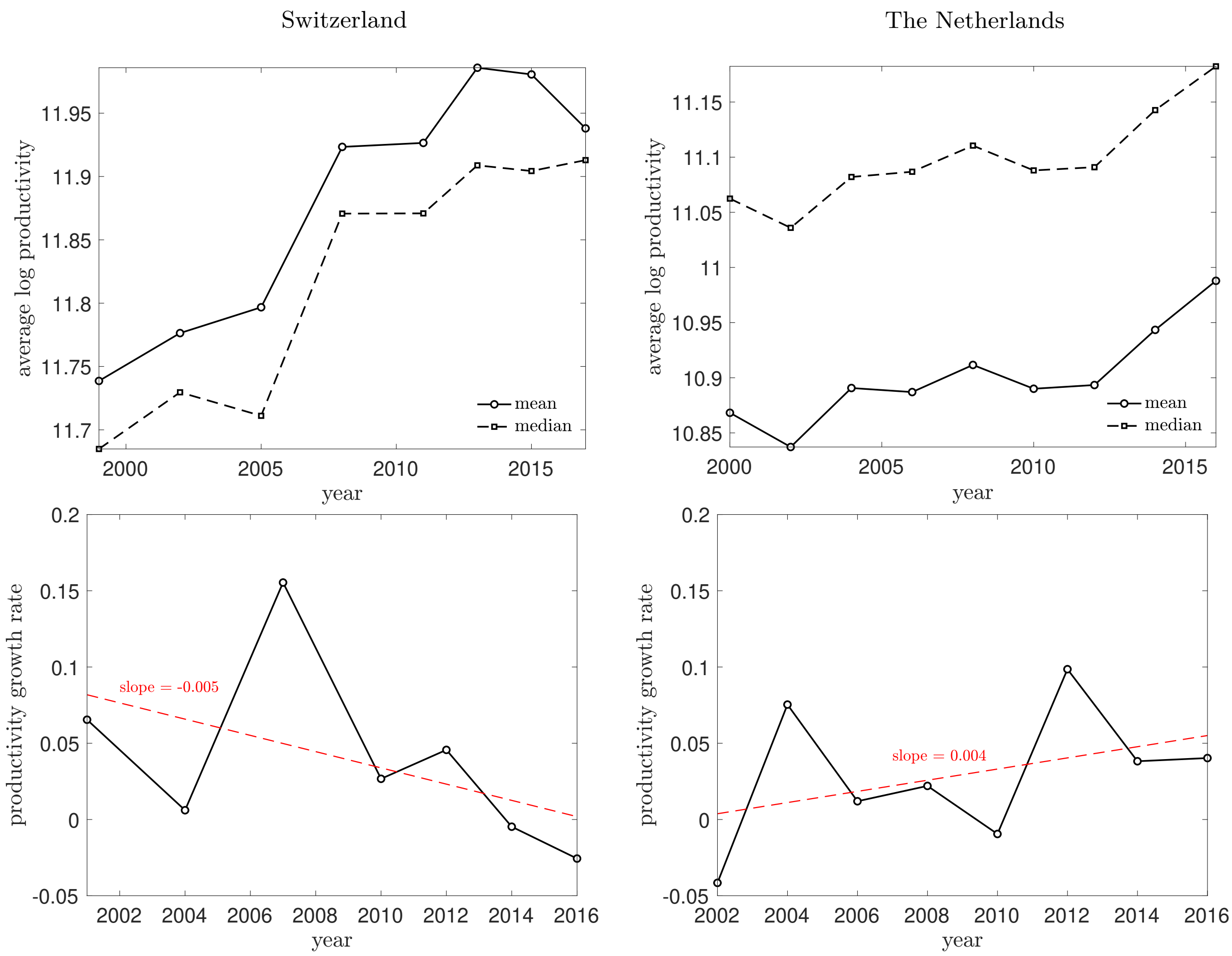

Figure 3: Evolution of average log productivity (top panels) and the productivity growth rate (bottom panels) in Switzerland (left panels) and the Netherlands (right panels). In the top panels the arithmetic mean and the median are shown. Productivity is measured as real value added per employee (in full-time equivalents). The dashed line in the bottom panels represent a linear regression fit.

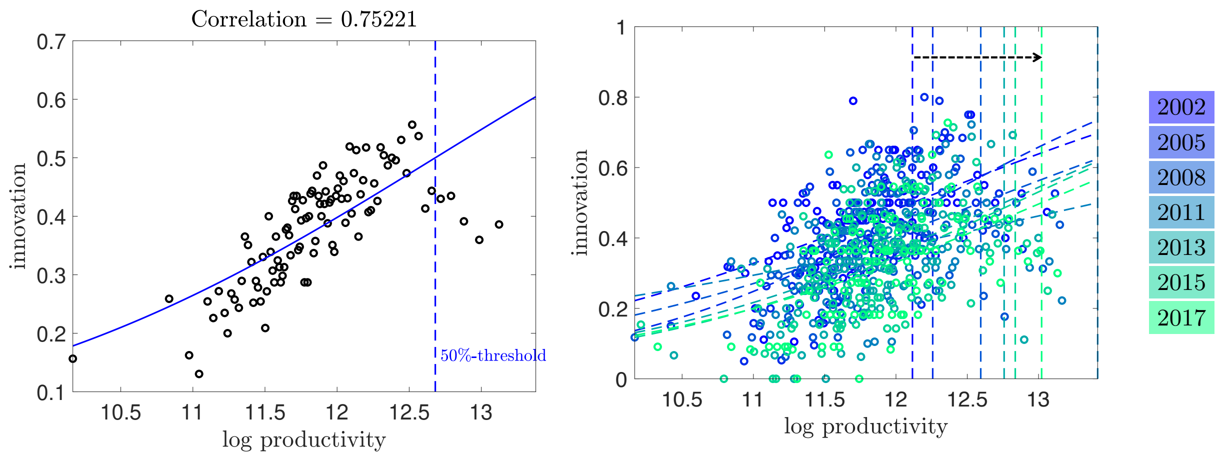

Figure 4: (Left panel) The innovation decision over productivity pooled across all years of observation in Switzerland. Productivity is measured as real value added per employee. The solid line shows the fit of a logistic function, $f(x) = 1/(1+\exp(-\beta_1(x-\beta_2)))$. The threshold is given by $x^* = \beta_2$ such that $f(\beta_2) = 1/2$. The 50%-threshold (dashed line) indicates the lowest productivity level at which the likelihood of doing R&D is more than 50%. (Right panel) The innovation decision over productivity across different years. The arrow indicates an upward shift of the 50%-threshold (see also the left panel of Figure 6).

50%-threshold over time in Switzerland and the Netherlands, respectively. The same evolution is depicted in Figure 6. The innovation-productivity threshold is increasing over time in Switzerland but decreasing in the Netherlands. This implies that Swiss (Dutch) companies need increasingly higher (lower) productivity levels to stay R&D active. In a complementary analysis, we examine in Supplementary Appendix A the factors (either firm characteristics or innovation policy instruments) that correlate with a firm's decision to continue doing R&D by running a survival analysis [Cleves et al., 2008]. We show that firm size (measured by employment in full-time equivalents), firm productivity (measured by real value added per worker), absorptive capacity (measured by the employment share of academics and the share of employees with higher education in Switzerland, and the share of employees with tertiary education in the Netherlands) and access to international markets (measured by export status) correlate negatively with the hazard rate of exiting from R&D. These factors can impact firms' R&D decision by increasing the likelihood of innovation success and easing the imitation of other firms' technologies. We also document that domestic and international innovation support decrease the hazard rate of exiting R&D. are correlated with firms' R&D decision. While these measures may also affect innovation success, their primary impact is on firms' R&D costs. These Cox proportional hazards estimates are simple, partial correlations and could potentially suffer from endogeneity problems arising from omitted variable bias and reverse causality, a concern we address in the next sections.

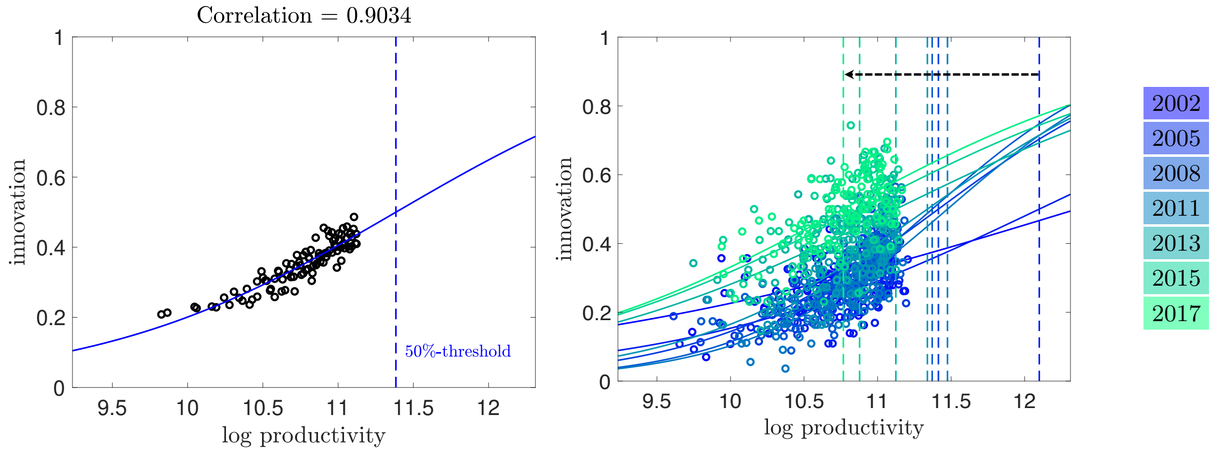

Figure 5: (Left panel) The innovation decision over productivity pooled across all years of observation in the Netherlands. Productivity is measured as real value added per employee. The solid line shows the fit of a logistic function, $f(x) = 1/(1 + \exp(-\beta_1(x - \beta_2)))$. The threshold is given by $x^* = \beta_2$ such that $f(\beta_2) = 1/2$. The 50%-threshold (dashed line) indicates the lowest productivity level at which the likelihood of doing R&D is more than 50%. (Right panel) The innovation decision over productivity across different years. The arrow indicates a downward shift of the 50%-threshold (see also the right panel of Figure 6).

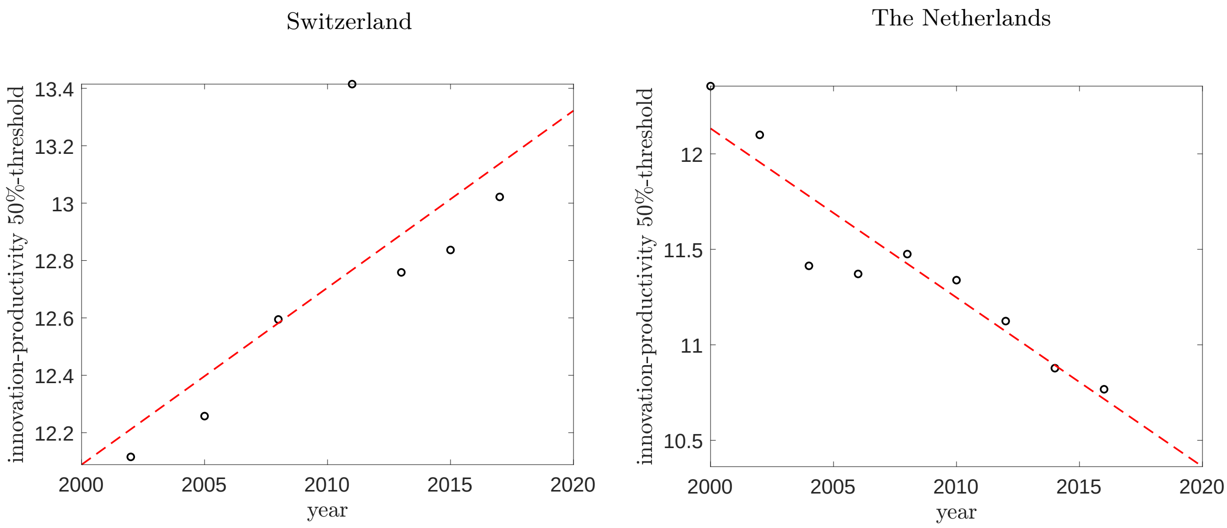

Figure 6: The innovation-productivity 50%-threshold for Switzerland (left panel) and the Netherlands (right panel). Productivity is measured as real value added per employee. Dashed lines indicate a linear regression fit.

3. Determinants of Firms' R&D Decisions

In the following sections, we jointly model the impact of three fundamentals –the likelihood of innovation success (in-house R&D success), the innovation or R&D costs and the likelihood of imitation success (resulting in technology diffusion)– on the R&D decision of a firm. Such a model allows us to take into account potential endogeneity concerns. The goal is to disentangle the effects of these three fundamentals, which underlie every firm's R&D decision problem. To this end, we introduce a model of endogenous technological change, productivity growth and technology spillovers where firms' choice between in-house R&D and imitation is endogenous and based on firms' profit maximization motive. The model follows König et al. [2022]. The structural model also allow us to simulate counterfactuals to examine the productivity growth effects of R&D policies in Section 5.

3.1. Firms' Profits and R&D Costs

For simplicity, we assume a uniform revenue tax τ and no R&D subsidies ($s = 0$) and relax these assumptions in Section 5.2. Firm i's profits at time $t \geq 0$ are given by

$$\pi_i(t) = \widetilde{\Psi} A_i(t)^{\eta-1} - c_i(t), \tag{1}$$

with a proportionality factor $\widetilde{\Psi} = \Psi(1-\tau)^{\eta}$, where $\Psi = \frac{1}{\eta}(1-\tau)^{1-\eta}\left(\sum_{i=1}^{n} A_i^{\eta-1}\right)^{\frac{1}{\eta-1}-1}$, demand elasticity $\eta > 1$ and $A_i(t)$ the productivity of firm i at time t. For the derivation of the profit function, we refer to Supplementary Appendix B.

Further, following König et al. [2022], we assume that the cost of innovation is given by:[11]

$$c_i(t) = \begin{cases} \Psi \cdot \left(\kappa + 1 - \exp\left(\xi_i - \frac{\sigma_\kappa}{2}\right)\right) \cdot \left(\overline{A}(t)^{\theta} A_i(t)^{1-\theta}\right)^{\eta-1} & \text{if } \quad i \text{ innovates,} \\ 0 & \text{if } \quad i \text{ imitates,} \end{cases} \tag{2}$$

where $\xi_i \sim \mathcal{N}(0, \sigma_\kappa^2)$ are innovation wedges (i.i.d. across firms), $\overline{A}(t)$ is average productivity at time t, and $\kappa > 0$, $\sigma_\kappa > 0$ and $\theta \in [0,1]$ are cost parameters. Note that $\mathbb{E}\left[\kappa + 1 - \exp\left(\xi_i - \sigma_\kappa/2\right)\right] = \kappa$, so σ_κ is a mean-preserving spread. We assume that R&D costs are a function of the geometric combination of A_i and $\overline{A}$. If $\theta = 1$, R&D costs are independent of productivity. If $\theta < 1$, R&D costs vary across firms, being higher for high-productivity firms that are closer to the technological frontier.[12] When taking the model to the data, we estimate the value of θ that best fits the empirical

[11] For a similar cost specification, we refer also to Benhabib et al. [2014] and Chen et al. [2021]. In particular, Chen et al. [2021] note that any homothetic production function with Hicks neutral productivity admits this representation. For empirical support, see e.g. Cohen [2010], who shows that internal R&D funds increase with firm size and thus also with productivity.

[12] Arora et al. [2018], Bloom et al. [2020] and Griliches [1998] demonstrate that it has become more expensive to innovate. This is especially true for companies at the technological frontier. Arora et al. [2018] find that "more research effort may be necessary to produce one unit of scientific output if producing scientific breakthroughs (or 'climbing over

application at hand.

3.2. Innovation vs. Imitation

Productivity is measured along a quality ladder, $A_i \in \{\widetilde{A}, \widetilde{A}^2, \widetilde{A}^3, \ldots\}$. Firms can increase their productivity by a factor $\widetilde{A}$ along the ladder via two alternative channels: through imitating other firms' technologies (diffusion) or through conducting costly in-house R&D (innovation). Let P_a denote the probability mass function of the fraction of firms with log-productivity $a = \log(A)$ and let the cumulative distribution function be $F_a = \sum_{b=1}^{a} P_b$. Further, to simplify the notation, we set the parameter σ_κ to zero in what follows.

Imitation. A firm pursuing the imitation strategy is randomly matched with another firm in the empirical distribution. If the firm is matched with a more productive firm, its productivity increases by one notch with probability $q \in [0, 1]$ and remains constant with probability $1 - q$. If the firm is matched with a less productive firm, it retains its initial productivity. Because of random matching, the probability that an imitating firm with log-productivity a moves up the productivity ladder equals $q \sum_{b=1}^{\infty} P_{a+b} = q(1 - F_a)$.

Innovation. A firm can discover something genuinely new that is unrelated to the knowledge set of other firms with probability $p_i \in [0, 1]$ (in-house R&D success). The realization of p_i is observed at the beginning of each period t, before firms choose whether to innovate or imitate.

3.3. Innovation Decision and Threshold

We assume that firms choose whether to innovate through in-house R&D or to imitate other firms based on a standard value-maximization objective. In our environment, this is equivalent to maximizing the expected profit in every period t. In turn, Equation (1) shows that the profit is linearly increasing in the productivity level.

Let $\mathbb{E}_i^{\text{in}}[\pi_i(t + \Delta t)| \cdot]$ and $\mathbb{E}_i^{\text{im}}[\pi_i(t + \Delta t)| \cdot]$ denote the expected profit for firm i from choosing in-house R&D or imitation, respectively. The probability of success in innovating through in-house R&D is given by $p_i(t) \in [\underline{p}, \overline{p}]$, with $0 \leq \underline{p} \leq \overline{p} \leq 1$. The probabilities $p_i(t)$ are i.i.d. and realized at the beginning of each period t.

Firm i chooses innovation whenever, conditional on its current productivity $A_i(t)$ and the state

the shoulders of giants') is getting harder". Bloom et al. [2020] document that it has become difficult for firms with higher productivity to generate new ideas. There might be several reasons for this. For example, more productive firms could be forced to devote management and labor resources to R&D, which has higher opportunity costs [cf. König et al., 2022].

of $p_i(t)$, we have that

$$\mathbb{E}_i^{\text{in}}[\pi_i(t+\Delta t)|\, A_i(t), p_i(t), P(t)] = \mathbb{E}_i^{\text{in}}\left[\widetilde{\Psi}A_i(t+\Delta t)^{\eta-1} - \Psi\kappa\left(\overline{A}(t)^{\theta}A_i(t)^{1-\theta}\right)^{\eta-1}\middle|\, A_i(t), p_i(t), P(t)\right] >$$
$$\mathbb{E}_i^{\text{im}}\left[\pi_i(t+\Delta t)|\, A_i\left(t\right), P(t)\right] = \mathbb{E}_i^{\text{im}}\left[\widetilde{\Psi}A_i(t+\Delta t)^{\eta-1}\middle|\, A_i\left(t\right), P(t)\right]. \quad (3)$$

The expected profit in Equation (3) from imitation is given by

$$\mathbb{E}_i^{\text{im}}\left[\pi_i(t+\Delta t)|\, A_i\left(t\right), P(t)\right] = q\left(1-F_{a_i(t)}(t)\right)\widetilde{\Psi}A_i(t)^{\eta-1}\widetilde{A}^{\eta-1} + \left(1-q\left(1-F_{a_i(t)}(t)\right)\right)\widetilde{\Psi}A_i(t)^{\eta-1},$$

while the expected profit from innovation is given by

$$\mathbb{E}_i^{\text{in}}[\pi_i(t+\Delta t)|\, A_i(t), p_i(t), P(t)] = p_i(t)\widetilde{\Psi}A_i(t)^{\eta-1}\widetilde{A}^{\eta-1} - \Psi\kappa\left(\overline{A}(t)^{\theta}A_i(t)^{1-\theta}\right)^{\eta-1} + (1-p_i(t))\widetilde{\Psi}A_i(t)^{\eta-1}. \quad (4)$$

In terms of log-productivities $a_i(t) = \log A_i(t)$, $\bar{a}(t) = \log \overline{A}(t)$ and $\log \widetilde{A} = \widetilde{a}$,[13] we can write

$$\begin{aligned}\mathbb{E}_i^{\text{in}}[\pi_i(t+\Delta t)|\, a_i(t), p_i(t), P(t)] &= p_i(t)\widetilde{\Psi}e^{(\eta-1)(a_i(t)+\widetilde{a})} - \Psi\kappa e^{(\eta-1)\theta\bar{a}(t)}e^{(\eta-1)(1-\theta)a_i(t)}\\ &\quad + (1-p_i(t))\widetilde{\Psi}e^{(\eta-1)a_i(t)},\end{aligned}$$

and

$$\mathbb{E}_i^{\text{im}}\left[\pi_i(t+\Delta t)|\, a_i(t), P(t)\right] = \widetilde{\Psi}e^{(\eta-1)a_i(t)}\left(1+q\left(1-F_{a_i(t)}(t)\right)\left(e^{(\eta-1)\widetilde{a}}-1\right)\right).$$

In the following, we denote by $\overline{\pi}_i^{\text{im}}(a_i(t), P(t)) \equiv \mathbb{E}_i^{\text{im}}[\pi_i(t+\Delta t)|\, a_i(t), P(t)]$ and $\overline{\pi}_i^{\text{in}}(a_i(t), p_i(t), P(t)) \equiv \mathbb{E}_i^{\text{in}}[\pi_i(t+\Delta t)|\, a_i(t), p_i(t), P(t)]$. The indicator function for whether firm i pursues imitation can then be written as $\chi^{\text{im}}(a,p,P) = \mathbb{1}_{\{\overline{\pi}_i^{\text{im}}(a,P) > \overline{\pi}_i^{\text{in}}(a,p,P)\}}$. We also define the indicator function for innovation as $\chi^{\text{in}}(a,p,P) \equiv 1-\chi^{\text{im}}(a,p,P)$. Further, we define $\widetilde{\kappa} = \frac{(1-\tau)^{-\eta}\kappa}{e^{(\eta-1)\widetilde{a}}-1}$. We can then write

$$\chi^{\text{im}}(a,p,P) = 1-\chi^{\text{in}}(a,p,P) = \begin{cases}1 & \text{if } p < q(1-F_a) + \widetilde{\kappa}e^{\theta(\eta-1)(\bar{a}-a)},\\ 0 & \text{otherwise.}\end{cases} \quad (5)$$

3.4. Innovation Decision and Comparative Statics

Based on Equation (5), if we denote by

$$D_i = \underbrace{p_i}_{\substack{\text{innovation}\\ \text{potential}}} \underbrace{-\tilde{\kappa}e^{\theta(\eta-1)(\bar{a}-a_i)}}_{\substack{\text{R\&D cost}\\ \text{effect}}} \underbrace{-q(1-F_{a_i})}_{\substack{\text{imitation}\\ \text{potential}}}, \quad (6)$$

[13]Observe that $A_i \in \{\widetilde{A}, \widetilde{A}^2, \widetilde{A}^3, \ldots\} = \{e^{\widetilde{a}}, e^{2\widetilde{a}}, e^{3\widetilde{a}}, \ldots\}$.

then firm i conducts innovation ($\chi^{\text{in}}_{a_i}(p, P) = 1$) if $D_i > 0$ and chooses imitation ($\chi^{\text{im}}_{a_i}(p, P) = 1$) if $D_i \leq 0$. From a comparative statics analysis, we obtain that

$$\frac{\partial D_i}{\partial a_i} = \theta\widetilde{\kappa}e^{\theta(\eta-1)(\overline{a}-a_i)} + qf_{a_i} > 0, \qquad \text{(size)},$$

$$\frac{\partial D_i}{\partial p_i} = 1 > 0, \qquad \text{(in-house R\&D success probability)},$$

$$\frac{\partial D_i}{\partial \widetilde{\kappa}} = -e^{\theta(\eta-1)(\overline{a}-a_i)} < 0, \qquad \text{(R\&D cost)},$$

$$\frac{\partial D_i}{\partial q(1-F_{a_i})} < 0, \qquad \text{(imitation success probability / laggards)}.$$

In particular, firms with a higher log-productivity a_i tend to be larger and, according to Equation (6), have a higher probability of conducting R&D ($\frac{\partial D_i}{\partial a_i} > 0$). A higher value of the in-house R&D success probability p_i, due to e.g. a higher technological potential of the firm from industry-university collaborations, leads to a higher probability of engaging in R&D ($\frac{\partial D_i}{\partial p_i} > 0$). More costly R&D (increase in $\widetilde{\kappa}$) leads to a lower probability of pursuing R&D ($\frac{\partial D_i}{\partial \widetilde{\kappa}} < 0$), consistent with survey data.[14] Finally, the term $q(1-F_{a_i})$ reduces the likelihood of the firm conducting R&D, which is increasing in the imitation success probability q and is higher for firms lagging further behind in their log-productivity due to higher values of $1 - F_{a_i}$ [cf. Aghion et al., 2005; Hashmi, 2013]. .

3.5. Failed In-house R&D and Passive Imitation

So far, we have assumed that a firm will not be able to enhance its productivity if it fails to innovate. This is a very strict assumption and excludes the possibility that a firm alters its strategy and shifts its focus to imitation instead of innovation when the latter proves unsuccessful. In an extended version of our model, we assume that if innovation fails, the firm gets a second chance to improve its technology via (passive) imitation. However, in such case, the probability of success is different from that of a firm actively pursuing imitation, being equal to $\delta q(1-F_a)$, with δ the passive imitation success probability. Thus, the total probability of success of a firm pursuing innovation is $p_i + (1-p_i)\delta q(1-F_a)$. The expected profit from innovation can then be computed by expanding Equation (4) as follows

$$\begin{aligned}&\mathbb{E}^{\text{in}}_i[\pi_i(t+\Delta t)|\, A_i(t), p_i(t), P(t)] = p_i(t)\Psi A_i(t)^{\eta-1}\widetilde{A}^{\eta-1} - \Psi\kappa\left(\overline{A}(t)^{\theta}A_i(t)^{1-\theta}\right)^{\eta-1} + (1-p_i(t))\\ &\times\left\{\delta\left[q\left(1-F_{a_i(t)}(t)\right)\Psi A_i(t)^{\eta-1}\widetilde{A}^{\eta-1} + \left(1-q\left(1-F_{a_i(t)}(t)\right)\right)\Psi A_i(t)^{\eta-1}\right] + (1-\delta)\Psi A_i(t)^{\eta-1}\right\}.\end{aligned} \tag{7}$$

[14]Note also that the term $e^{\theta(\eta-1)(\overline{a}-a_i)}$ becomes smaller the further firm i's log-productivity a_i lies above average log-productivity $\overline{a}$, thereby weakening the cost effect via $\widetilde{\kappa}$ for firms at the technological frontier.

In terms of log-productivities (where we have defined $a_i(t) = \log A_i(t)$, $\overline{a}(t) = \log \overline{A}(t)$ and $\log \widetilde{A} = \widetilde{a}$), we can write this as

$$\mathbb{E}_i^{\text{in}}[\pi_i(t+\Delta t)|\, a_i(t), p_i(t), P(t)] = p_i(t)\Psi e^{(\eta-1)(a_i(t)+\widetilde{a})} - \Psi\kappa e^{(\eta-1)\theta\overline{a}(t)} e^{(\eta-1)(1-\theta)a_i(t)} + (1-p_i(t)) \times \left\{\delta \left[q\left(1-F_{a_i(t)}(t)\right)\Psi e^{(\eta-1)(a_i(t)+\widetilde{a})} + \left(1-q\left(1-F_{a_i(t)}(t)\right)\right)\Psi e^{(\eta-1)a_i(t)}\right] + (1-\delta)\Psi e^{(\eta-1)a_i(t)}\right\}.$$

With passive imitation, the indicator function for whether firm i pursues imitation in Equation (5) then needs to be modified as follows

$$\chi^{\text{im}}(a,p,P) = 1-\chi^{\text{in}}(a,p,P) = \begin{cases} 1 & \text{if } p < \frac{(1-\delta)q(1-F_a)+\widetilde{\kappa}e^{\theta(\eta-1)(\overline{a}-a)}}{1-\delta q(1-F_a)}, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

3.6. Law of Motion of the Productivity Distribution

We consider an environment in which firms decide to conduct in-house R&D or to imitate other firms by maximizing the expected profit in every period t (see Sections 3.3 to 3.5). These decisions determine how the distribution of productivity evolves over time t. The following proposition provides a complete characterization of the evolution of the productivity distribution with heterogeneous firms in terms of their in-house R&D success probabilities, $p_i(t) \in [\underline{p}, \overline{p}]$.

Proposition 1. *The evolution of the* log*-productivity distribution $P_a(t)$, $a \in \mathcal{A}$, is given by the following system of integro-differential equations*

$$\begin{aligned}\frac{\partial P_a(t)}{\partial t} = \int_{[\underline{p},\overline{p}]} g(dp)\Big[&\left(\chi^{im}(a-1,p,P) + \delta(1-p)\chi^{in}(a-1,p,P)\right) q(1-F_{a-1}(t))P_{a-1}(t) \\ &- \left(\chi^{im}(a,p,P) + \delta(1-p)\chi^{in}(a,p,P)\right) q(1-F_a(t))P_a(t) \\ &+ \chi^{in}(a-1,p,P)pP_{a-1}(t) - \chi^{in}(a,p,P)pP_a(t)\Big],\end{aligned} \tag{9}$$

where $g : [\underline{p},\overline{p}] \to [0,1]$ is the density function of a random variable over the interval $[\underline{p},\overline{p}]$ and

$$\chi^{im}(a,p,P) = 1-\chi^{in}(a,p,P) = \begin{cases} 1 & \textit{if } p < \frac{(1-\delta)q(1-F_a)+\widetilde{\kappa}e^{\theta(\eta-1)(\overline{a}-a)}}{1-\delta q(1-F_a)}, \\ 0 & \textit{otherwise,} \end{cases} \tag{10}$$

with $F_a = \sum_{b=1}^{a} P_b$ and average log*-productivity given by $\overline{a} = \sum_{a=1}^{\infty} aP_a$.*

The proof of Proposition 1 is provided in Supplementary Appendix D. We obtain the evolution of the productivity distribution $P_a(t)$ by numerically solving the system of ordinary differential equations provided in Equation (9) for a given initial condition $P_a(0)$. This will be important for estimating the model in Section 4.2. In Supplementary Appendix C, we simplify Equation (9) by assuming, as in König et al. [2022], that the in-house R&D success probability is uniformly distributed, consistent with the assumption used when estimating the structural model.

4. Estimation

In the following, we estimate the structural model introduced in Section 3, which takes into account the endogenous evolution of firms' productivity from their innovation and imitation decisions. Relying on a structural stability test following Hall [2005], and motivated by the post-2008 rise in R&D active firms in the Netherlands and the 2007 introduction of the Innovation Box, we compare pre- and post-2008 estimates for both countries to quantify changes in the fundamentals driving firms' R&D decisions.

4.1. External Calibration and TFP Calculation

Following König et al. [2022] and Song et al. [2011], we set the demand elasticity parameter η to 5. Labor productivity is simply computed as real value added per employee. Further, from Equation (2), we obtain an estimate of the cost elasticity parameter θ by estimating a linear regression of de-meaned log R&D expenditures on de-meaned log productivities in each period.[15] Using data from Switzerland we obtain a value of $\theta = 0.68$, which means that around two thirds of the R&D cost is driven by average productivity and one third by the idiosyncratic productivity level of the firm. This is similar to the estimate for θ of 0.75 obtained in König et al. [2022]. For a better comparison, we use the same value for the Netherlands. Finally, the step size $\widetilde{a}$ is set to 0.05, balancing computational cost (a smaller step size increases the computation time due to larger arrays necessary for storing data) with granularity.

4.2. Structural Estimation: Productivity Distribution and R&D Decision

To estimate the parameters of the structural model introduced in Section 3, we use a Simulated Method of Moments (SMM) procedure where we target a set of moments derived from the data and our model [Hall, 2005; McFadden, 1989].

The first moment is the average log-productivity, $\overline{a} = \sum_a aP(a)$, the second the log-productivity growth rate, ν, and the third the log-productivity variance, $\sigma^2 = \sum_a (a - \overline{a})^2 P(a)$.[16] The next set of moments is given by the R&D profile, $H(a)$, its shape parameters, β_1 (steepness) and β_2 (50%-threshold) (cf. Figures 4 and 5 in Section 2.2) and the unweighted bin-sum of R&D shares, $\sum_a H(a)$. The final set of moments is given by the fraction of R&D firms, $\sum_a P(a)H(a)$, the average log-productivity of R&D firms, $\sum_a aP(a)H(a) \ / \ \sum_a P(a)H(a)$, and the log-productivity variance of R&D firms, $\sum_a a^2 \ H(a)P(a) \ / \ \sum_a H(a)P(a) - (\sum_a a \ H(a)P(a) \ / \ \sum_a H(a)P(a))^2$.

With the innovation decision in Equation (10) and assuming uniform draws of p in the interval

[15] Note that Equation (2) implies $\log c_i(t) - \overline{\log c}(t) = (\eta - 1)(1 - \theta)(a_i(t) - \overline{a}(t))$, where $c_i(t)$ is measured by the R&D expenditure level of firm i at time t, $a_i(t)$ by its log-productivity level and over-lined variables correspond to averages across firms.

[16] The sensitivity of the growth rate ν and variance σ^2 with respect to the parameters $\bar{p}$, $\widetilde{\kappa}$, q and δ is shown in Figures 9 and 10, respectively. Increases in $\bar{p}$, q and δ increase the growth rate while an increase in $\widetilde{\kappa}$ decreases the growth rate. Increases in $\bar{p}$ and $\widetilde{\kappa}$ increase the variance, while increases in q and δ decrease it.

$[0, \overline{p}]$, the (model-implied) R&D profile can be computed as follows

$$H(a, P) \equiv \int_{[0,\overline{p}]} \chi^{\text{in}}(a, p, P)dp = \int_{[0,\overline{p}]} \mathbb{1}_{\{p>C(a,P)\}}dp, \tag{11}$$

where we have denoted by (cf. Equation (C.2) in Supplementary Appendix C)

$$C(a, P) \equiv \frac{(1-\delta)q(1-F_a) + \widetilde{\kappa}e^{\theta(\eta-1)(\overline{a}-a)}}{1-\delta q\,(1-F_a)}.$$

The moments for our SMM approach are then given by

$$\widehat{\mathbf{m}}(P, H|\boldsymbol{\theta}) = \begin{bmatrix} \vdots \\ \sum_a aP(a,t) \\ \sum_a (a-\overline{a}(P))^2 P(a,t) \\ \nu(P) \\ \vdots \\ H(a,t) \\ \vdots \\ \beta_1(H) \\ \beta_2(H) \\ \sum_a P(a,t)H(a,t) \\ \sum_a H(a,t) \\ \sum_a aP(a,t)H(a,t)/\sum_a P(a,t)H(a,t) \\ \frac{\sum_a a^2 H(a,t)P(a,t)}{\sum_a H(a,t)P(a,t)} - \left(\frac{\sum_a aH(a,t)P(a,t)}{\sum_a H(a,t)P(a,t)}\right)^2 \\ \vdots \end{bmatrix}, \quad \mathbf{m}(P_{\text{obs}}, H_{\text{obs}}) = \begin{bmatrix} \vdots \\ \sum_a aP_{\text{obs}}(a,t) \\ \sum_a (a-\overline{a}(P_{\text{obs}}))^2 P_{\text{obs}}(a,t) \\ \nu(P_{\text{obs}}) \\ \vdots \\ H_{\text{obs}}(a,t) \\ \vdots \\ \beta_1(H_{\text{obs}}) \\ \beta_2(H_{\text{obs}}) \\ \sum_a P_{\text{obs}}(a,t)H_{\text{obs}}(a,t) \\ \sum_a H_{\text{obs}}(a,t) \\ \sum_a aP_{\text{obs}}(a,t)H_{\text{obs}}(a,t)/\sum_a P_{\text{obs}}(a,t)H_{\text{obs}}(a,t) \\ \frac{\sum_a a^2 H_{\text{obs}}(a)P_{\text{obs}}(a,t)}{\sum_a H_{\text{obs}}(a,t)P_{\text{obs}}(a,t)} - \left(\frac{\sum_a aH_{\text{obs}}(a,t)P_{\text{obs}}(a,t)}{\sum_a H_{\text{obs}}(a,t)P_{\text{obs}}(a,t)}\right)^2 \\ \vdots \end{bmatrix}$$

with the error function given by $\mathbf{e}(P, H, P_{\text{obs}}, H_{\text{obs}}|\boldsymbol{\theta}) = \widehat{\mathbf{m}}(P, H|\boldsymbol{\theta}) - \mathbf{m}(P_{\text{obs}}, H_{\text{obs}})$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is the vector of parameters, $(P_{\text{obs}}(a,t))_{a\geq 0}$ the observed empirical productivity distribution at time t and $(H_{\text{obs}}(a,t))_{a\geq 0}$ the observed empirical R&D profile at time t for $t = 0, \ldots, T$. The distributions $P(a,t)$ and $H(a,t)$ are the corresponding theoretical predictions from the solution to Equations (9) and (11) with the previous time period $(t-1)$ as initial condition. The SMM estimator is then given by

$$\widehat{\boldsymbol{\theta}}_{SMM} = \text{argmin}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\ \mathbf{e}(P, H, P_{\text{obs}}, H_{\text{obs}}|\boldsymbol{\theta})^\top\ \mathbf{W}\ \mathbf{e}(P, H, P_{\text{obs}}, H_{\text{obs}}|\boldsymbol{\theta}). \tag{12}$$

The SMM estimation algorithm uses an Ordinary Least Squares (OLS) regression of Equation (6) to obtain starting values for a numerical optimization algorithm to solve Equation (12) and to obtain the point estimates.[17] Standard errors are computed using the heteroscedasticity and autocorrelation consistent (HAC) weighting matrix [Newey and West, 1987]. The productivity data are trimmed at the bottom and top 1% percentile for robust inference [Čížek, 2008; Hill and Renault, 2010]. In all SMM model specifications, we assume that firms get a random draw from a uniform distribution

[17] While informative about basic correlations, these reduced-form OLS estimates may suffer from endogeneity bias (reverse causality) due to productivity (and other variables) being affected by the innovation decision (and vice versa). Moreover, applying such regression models to a binary dependent variable can yield inefficient, inconsistent and biased coefficient estimates [cf. e.g. Long, 1997].

Table 1: Identification of structural parameters.

Parameter	Main identifying moments	
$\bar{p}$	Log-Productivity growth rate	(ν)
q	Log-productivity variance	$(\sum_a (a-\overline{a})^2 P(a))$
δ	Log-productivity variance of R&D firms	$(\frac{\sum_a a^2 H(a)P(a)}{\sum_a H(a)P(a)} - \left(\frac{\sum_a aH(a)P(a)}{\sum_a H(a)P(a)}\right)^2)$
$\widetilde{\kappa}, \sigma_\kappa$	R&D profile and fraction of R&D firms	$(H(a), \sum_a P(a)H(a))$
θ	Productivity–R&D cost elasticity	(cf. Footnote 15)

with support $[0, \bar{p}]$ for their in-house R&D success probabilities in each period (cf. Supplementary Appendix C).

4.3. Identification

Because the estimation minimizes the weighted distance between theoretical and empirical moments, all parameters are jointly identified. Nevertheless, given the model's dynamics, a heuristic discussion of the identification strategy is informative.

To a first order, the in-house R&D success probability $\bar{p}$ is identified by the aggregate log-productivity growth rate ν, while the imitation success probability q is primarily identified by the dispersion of the productivity distribution, as measured by the log-productivity variance σ^2. The passive imitation success probability δ is identified by productivity dispersion among R&D-performing firms, which captures the extent to which unsuccessful innovators are able to catch up through diffusion. The innovation cost parameters $\widetilde{\kappa}$ and σ_κ are identified by their effects on the level and shape of the R&D participation profile $H(a)$ as well as by the aggregate fraction of R&D-active firms. Table 1 summarizes the identification of the structural parameters.

Additional moments are used to reduce the variance of the estimator and to match multiple features of the data simultaneously (e.g. growth rate, dispersion, threshold, and R&D profile), reducing the risk that parameters are driven by a single statistic, and enabling a distribution-based identification strategy [Lewbel, 2019].

4.4. Estimation Results

Tables 2 and 4 present the estimates for Switzerland and the Netherlands, respectively. The R-squared goodness-of-fit measure, $R^2_{\mathrm{KL}}(\cdot|\cdot)$, reported in these tables follows Cameron and Windmeijer [1997] and is defined as $R^2_{\mathrm{KL}}(\cdot|\cdot) = 1 - D_{\mathrm{KL}}(\cdot|\cdot)$, where $D_{\mathrm{KL}}(\cdot|\cdot)$ is the Kullback-Leibler divergence. The Kullback-Leibler divergence, $D_{\mathrm{KL}}(H|H_{\mathrm{obs}}) = \sum_a H(a) \log\left(H(a)/H_{\mathrm{obs}}(a)\right)$, measures the difference between the predicted (H) and the observed (empirical) distribution (H_{obs}) [Song, 2002]. The reported R-squared measures are averages over time periods. Across all samples, the J_T-statistic is rejected at the 1% level indicating that the over-identifying restrictions are satisfied. To assess structural stability of our model, we conduct the test proposed by Hall [2005]. Using a set of statistics (Wald, Lagrange Multiplier, and the D- and O-statistics), we find evidence of a structural break in 2008, consistent with the post-2008 increase in R&D-active firms in the Netherlands and the

2007 introduction of the Innovation Box. The Wald (W_T), Lagrange Multiplier (LM_T), D- (D_T), and O-statistics (O_T, O_{1T}, and O_{2T}) are computed following Equations (5.75), (5.77), (5.78), and (5.80)–(5.82) in Hall [2005], respectively (see also Ghysels and Guay [2003]). Based on the inferred breakpoint in 2008, we compare pre- and post-2008 estimates for both countries to quantify changes in the fundamentals driving firms' R&D decisions. These estimates are reported in columns (3) and (4) of Tables 2 and 4 for Switzerland and the Netherlands, respectively. Tables 3 and 5 report the first set of targeted moments for Switzerland and the Netherlands, respectively. Figure 7 for Switzerland and Figure 8 for the Netherlands compare the estimated and observed log-productivity distributions (non-targeted) and R&D profiles (targeted) for different years.

Estimation results and goodness-of-fit for Switzerland. The SMM estimation results for Switzerland are presented in Table 2. Column (1) reports the OLS estimates for Equation (6), while columns (2)–(4) display the SMM estimates described in Section 4.2. Columns (1) and (2) use the full sample from 2001 to 2016, whereas columns (3) and (4) incorporate a structural break in 2008, providing separate estimates for the periods 2001–2008 and 2010–2016, respectively. All test statistics, except for LM_T, indicate a structural break in 2008.

Comparing the estimates before and after the structural break, we find a 23.98% decline in the in-house R&D success probability ($\bar{p}$).[18] The estimated R&D cost parameter ($\widetilde{\kappa}$) rose by 11.86%, indicating that R&D costs gained importance for Swiss firms' R&D decisions in the post-2008 period. This is consistent with evidence from the Swiss Innovation Survey (see Spescha and Wörter [2022]), which reports increasing innovation costs for Swiss firms. An increase in R&D costs is also documented in the empirical literature: for example, Bloom et al. [2020] and Pammolli et al. [2011] report substantial cost increases and difficulties of innovating in the pharmaceutical sector, which is particularly important in Switzerland. The estimate of the dispersion in innovation costs (σ_κ) decreased by 20.04% post-2008. We observe a 37.48% increase in the imitation success probability among non-R&D firms (q) post-2008. This finding underscores the rising importance of imitation and technology diffusion for non-R&D firms in order to fit the theoretical model to the empirical data. By contrast, the passive imitation success probability of R&D firms (δ) decreased by 10.77% in the post-2008 period. Taken together, these results suggest that the decline in the in-house R&D success probability and the increase in R&D costs appear to be the driving forces behind the post-2008 decline in the share of R&Dactive firms (cf. the left panel of Figure 3 in Section 2.2).

The goodness-of-fit of the estimated SMM model is shown in Table 3 and the right panel of Figure 7. Non-targeted moments are shown in the left panel of Figure 7.[19] Except for the productivity growth rate in the post-2008 period, in which Switzerland experienced a decline in productivity, across the different years considered, the SMM model fits both, the targeted and the non-targeted moments well.

[18] The corresponding estimated average in-house R&D success probability is given by $\bar{p}/2$.

[19] Note that the SMM procedure targets only the first two moments of the log-productivity distribution—its mean and variance—rather than the entire distribution.

Table 2: Estimation results for Switzerland.

		OLS	SMM Pooled	SMM Breakpoint in 2008	
		w/o passive imitation	with passive imitation	with passive imitation	with passive imitation
		$\delta = 0$	$\delta \neq 0$	$\delta \neq 0$	$\delta \neq 0$
		(1)	(2)	(3)	(4)
Periods		2001–2016	2001–2016	2001–2008	2010–2016
Innovation	$(\overline{p})$	0.4564***	0.3516***	0.4099***	0.3116***
		(0.0021)	(0.0067)	(0.0042)	(0.0059)
Imitation	(q)	0.1537***	0.0009	0.0108*	0.0149
		(0.0052)	(0.0033)	(0.0065)	(0.0479)
Passive Imitation	(δ)		0.0050	0.9058***	0.8082
			(0.0000)	(0.0001)	(2.6519)
Cost Mean	$(\widetilde{\kappa})$	0.0068***	0.0009	0.0030***	0.0033***
		(0.0019)	(0.0037)	(0.0006)	(0.0009)
Cost Spread	(σ_κ)		0.0143	0.0022***	0.0018***
			(0.0105)	(0.0012)	(0.0023)
Observations		14,172	14,172	6,920	7,252
Firms		5,676	5,676	4,223	3,579
$R^2_{\text{KL}}(P\|P_{\text{obs}})$			0.92608	0.9414	0.9453
$R^2_{\text{KL}}(H\|H_{\text{obs}})$			0.93361	0.9441	0.9301
J_T-stat.			5.2381	47.8588	5.7570
Change Before/After 2008 Breakpoint					
Innovation	$(\Delta\overline{p}/\overline{p})$				–23.98%
Imitation	$(\Delta q/q)$				+37.48%
Passive Imitation	$(\Delta\delta/\delta)$				–10.77%
Cost Mean	$(\Delta\widetilde{\kappa}/\widetilde{\kappa})$				+11.86%
Cost Spread	$(\Delta\sigma_\kappa/\sigma_\kappa)$				–20.04%
W_T-stat.				843.4179	
LM_T-stat.				0.0364	
D_T-stat.				3990.2081	
O_T-stat.				21.5916	
O_{1T}-stat.				19.2607	
O_{2T}-stat.				2.3309	

Notes: Column (1) presents the OLS estimates for Equation (6). Columns (2)–(4) report the estimation results from the SMM algorithm. Standard errors are computed using the heteroscedasticity and autocorrelation consistent (HAC) optimal weighting matrix [Newey and West, 1987] and reported in parentheses. The asterisks *** /**/* indicate significance at the 1%/5%/10% level. The R-squared goodness-of-fit measure is defined as $R^2_{\text{KL}}(\cdot \mid \cdot) = 1 - D_{\text{KL}}(\cdot \mid \cdot)$, where $D_{\text{KL}}(H \mid H_{\text{obs}}) = \sum_a H(a) \log\left(H(a)/H_{\text{obs}}(a)\right)$ is the Kullback–Leibler divergence, and is reported as an average across time periods. The J_T-statistic tests the overidentifying restrictions. The Wald (W_T), Lagrange Multiplier (LM_T), D-(D_T), and O-statistics (O_T, O_{1T}, and O_{2T}) are structural stability tests. Additional goodness-of-fit measures are reported in Table 3 and illustrated in Figure 7.

Table 3: Targeted moments for Switzerland: Model vs. data.

Periods		2001–2008		2010–2016	
		Model	Data	Model	Data
Fraction of R&D firms	$(\sum_a P(a)H(a))$	0.3822	0.3908	0.2860	0.3119
Unweighted bin-sum of R&D shares	$(\sum_a H(a))$	21.0927	20.8590	16.2663	17.4528
Average log-productivity of R&D firms	$(\sum_a aP(a)H(a)/\sum_a P(a)H(a))$	11.9981	12.1048	12.1926	12.2895
Log-productivity variance of R&D firms	$(\frac{\sum_a a^2 H(a)P(a)}{\sum_a H(a)P(a)} - \left(\frac{\sum_a aH(a)P(a)}{\sum_a H(a)P(a)}\right)^2)$	0.6157	0.5386	0.6598	0.5639
Logistic fit R&D profile: steepness	(β_1)	0.0881	0.1209	0.1267	0.1447
Logistic fit R&D profile: 50%-threshold	(β_2)	17.2427	16.1589	19.3264	18.2022
Average log-productivity	$(\overline{a} = \sum_a aP(a))$	11.7229	11.7514	11.9311	11.9011
Log-productivity variance	$(\sum_a (a-\overline{a})^2 P(a))$	0.2128	0.2599	0.2734	0.2550
Productivity growth rate	(ν)	0.0316	0.0450	0.0273	0.0043

Notes: Reported numbers correspond to averages across time periods.

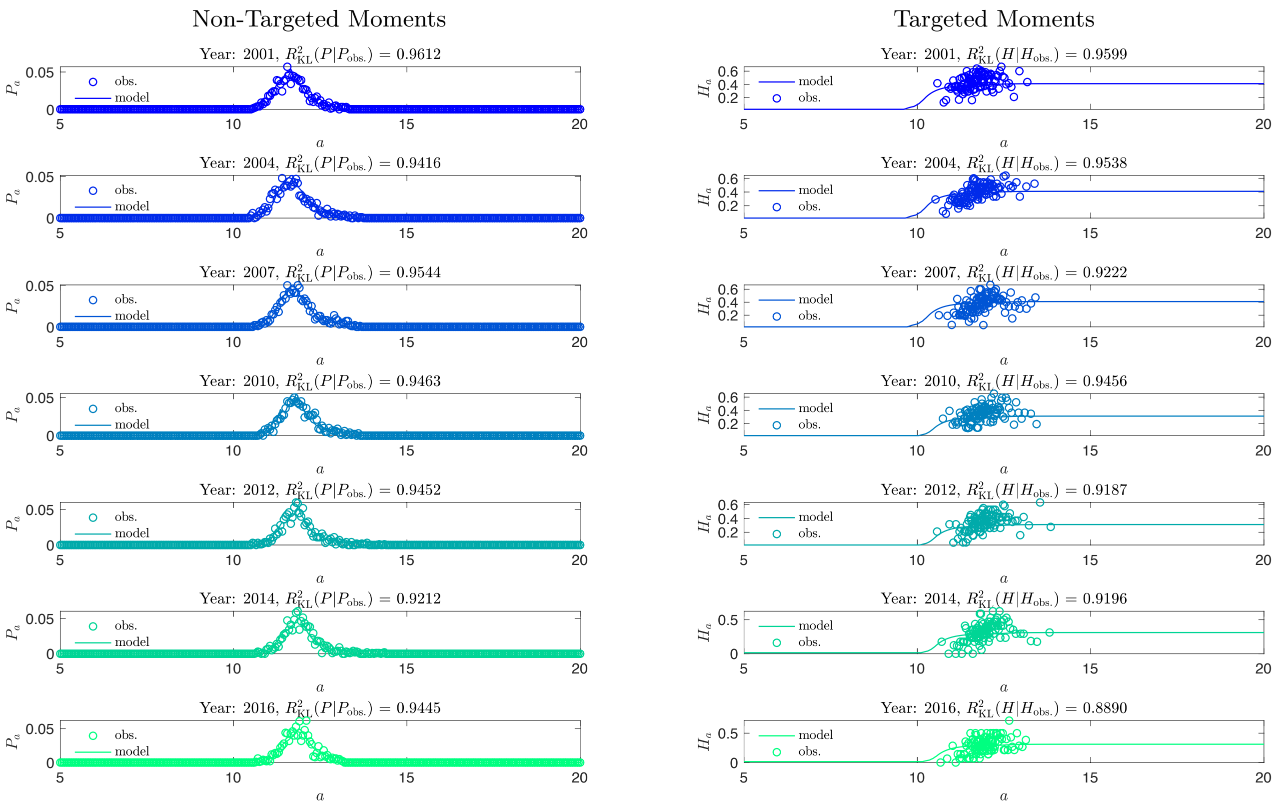

Figure 7: Goodness-of-fit of the estimated model for Switzerland with estimates, based on the estimates in Table 2 and a structural break in 2008. The panels on the left correspond to non-targeted moments, the panels on the right to targeted moments.

Estimation results and goodness-of-fit for the Netherlands. The SMM estimates for the Netherlands are reported in Table 4. In line with the Swiss results, all test statistics, except for LM_T, point to a structural break in 2008. Comparing the pre- and post-2008 periods, we find a 39.01% increase in the in-house R&D success probability ($\bar{p}$) and a 40.10% decline in R&D costs ($\widetilde{\kappa}$). This substantial reduction in costs may reflect post-2008 public support measures aimed at lowering innovation expenses. The dispersion in innovation costs (σ_κ) also decreases by 3.16% post-2008. Technology diffusion among non-R&D firms gained importance in the post-2008 period, as indicated by a 41.85% increase in the imitation success probability (q). Likewise, the passive imitation success probability of R&D firms (δ) rose by 45.95%. Overall, favorable innovation policies appear to have reduced R&D costs and supported firms in remaining R&D-active, e.g. by enabling them to hire STEM workers, which in turn raises innovation success for high-productivity R&D firms and imitation success for lower-productivity R&D firms. Taken together, higher R&D success rates and lower costs make innovation more attractive, ceteris paribus, helping to explain the post-2008 increase in the share of R&D firms (cf. the right panel of Figure 3 in Section 2.2).

The goodness-of-fit of the estimated SMM model is shown in Table 5 and in the right panel of Figure 8, while the non-targeted moments are displayed in the left panel. Across all years, the model matches both targeted and non-targeted moments well, indicating a satisfactory overall fit.

Robustness Analysis. As a robustness check, we re-estimated the structural model using a balanced panel that includes only firms with at least one observation in both the pre- and post-2008 periods. The corresponding results are reported in Tables E.1 and E.2 in Supplementary Appendix E for Switzerland and the Netherlands, respectively.

For Switzerland, the findings are qualitatively consistent with those in Table 2, except that, in the balanced panel, R&D costs and the imitation success probability of R&D firms appear to have decreased and increased, respectively, post-008. This pattern is likely driven by the balanced panel being dominated by larger firms, for which R&D costs are relatively less important than the in-house R&D success rate in affecting R&D decisions. For the Netherlands, the balanced panel estimates are qualitatively similar to those in Table 4, with the exception that the imitation success probability of R&D firms shows a slight decline in the post-2008 period.

Table 4: Estimation results for the Netherlands.

		OLS	SMM		
			Pooled	Breakpoint in 2008	
		w/o passive imitation	with passive imitation	with passive imitation	with passive imitation
		$\delta = 0$	$\delta \neq 0$	$\delta \neq 0$	$\delta \neq 0$
		(1)	(2)	(3)	(4)
Periods		2000–2016	2000–2016	2002–2008	2010–2016
Innovation	$(\overline{p})$	0.4481***	0.4883***	0.4004***	0.5566***
		(0.0004)	(0.0000)	(0.0000)	(0.0000)
Imitation	(q)	0.2042***	0.1749	0.1383	0.1962
		(0.0004)	(0.1656)	(0.1173)	(0.1342)
Passive Imitation	(δ)		0.6552*	0.6486*	0.9467***
			(0.3738)	(0.3436)	(0.1421)
Cost Mean	$(\widetilde{\kappa})$	0.0000	0.0068	0.0113**	0.0068***
		(0.0003)	(0.0045)	(0.0055)	(0.0026)
Cost Spread	(σ_κ)		0.0250*	0.0346**	0.0335***
			(0.0140)	(0.0156)	(0.0116)
Observations		43,719	43,719	26,178	17,541
Firms		25,892	25,892	15,659	10,233
$R^2_{\text{KL}}(P\|P_{\text{obs}})$			0.9145	0.9452	0.9580
$R^2_{\text{KL}}(H\|H_{\text{obs}})$			0.96499	0.9483	0.9659
J_T-stat.			1.4245	1.7482	3.5340
Change Before/After 2008 Breakpoint					
Innovation	$(\Delta\overline{p}/\overline{p})$				+39.01%
Imitation	$(\Delta q/q)$				+41.85%
Passive Imitation	$(\Delta\delta/\delta)$				+45.95%
Cost Mean	$(\Delta\widetilde{\kappa}/\widetilde{\kappa})$				–40.10%
Cost Spread	$(\Delta\sigma_\kappa/\sigma_\kappa)$				–3.16%
W_T-stat.				1.6570×10^{11}	
LM_T-stat.				2.2509	
D_T-stat.				93685.1520	
O_T-stat.				36.6260	
O_{1T}-stat.				6.8802	
O_{2T}-stat.				29.7459	

Notes: Column (1) presents the OLS estimates for Equation (6). Columns (2)–(4) report the estimation results from the SMM algorithm. Standard errors are computed using the heteroscedasticity and autocorrelation consistent (HAC) optimal weighting matrix [Newey and West, 1987] and reported in parentheses. The asterisks *** /**/* indicate significance at the 1%/5%/10% level. The R-squared goodness-of-fit measure is defined as $R^2_{\text{KL}}(\cdot \mid \cdot) = 1 - D_{\text{KL}}(\cdot \mid \cdot)$, where $D_{\text{KL}}(H \mid H_{\text{obs}}) = \sum_a H(a) \log\big(H(a)/H_{\text{obs}}(a)\big)$ is the Kullback–Leibler divergence, and is reported as an average across time periods. The J_T-statistic tests the overidentifying restrictions. The Wald (W_T), Lagrange Multiplier (LM_T), D- (D_T), and O-statistics (O_T, O_{1T}, and O_{2T}) are structural stability tests. Additional goodness-of-fit measures are reported in Table 5 and illustrated in Figure 8.

Table 5: Targeted moments for the Netherlands: Model vs. data.

Periods		2002–2008		2010–2016	
		Model	Data	Model	Data
Fraction of R&D firms	$(\sum_a P(a)H(a))$	0.3011	0.2728	0.4652	0.4189
Unweighted bin-sum of R&D shares	$(\sum_a H(a))$	15.5735	15.6408	24.7992	24.8488
Average log-productivity of R&D firms	$(\sum_a aP(a)H(a)/\sum_a P(a)H(a))$	12.1266	12.0263	12.1968	12.0353
Log-productivity variance of R&D firms	$\left(\frac{\sum_a a^2 H(a)P(a)}{\sum_a H(a)P(a)} - \left(\frac{\sum_a aH(a)P(a)}{\sum_a H(a)P(a)}\right)^2\right)$	0.5267	0.6140	0.6086	0.7565
Logistic fit R&D profile: steepness	(β_1)	0.1831	0.1961	0.2367	0.2985
Logistic fit R&D profile: 50%-threshold	(β_2)	17.0787	17.0165	13.5866	13.3440
Average log-productivity	$(\overline{a} = \sum_a aP(a))$	11.8882	11.8182	12.0084	11.9432
Log-productivity variance	$(\sum_a (a-\overline{a})^2 P(a))$	0.2615	0.2418	0.2436	0.2655
Productivity growth rate	(ν)	0.0295	0.0184	0.0444	0.0349

Notes: Reported numbers correspond to averages across time periods.

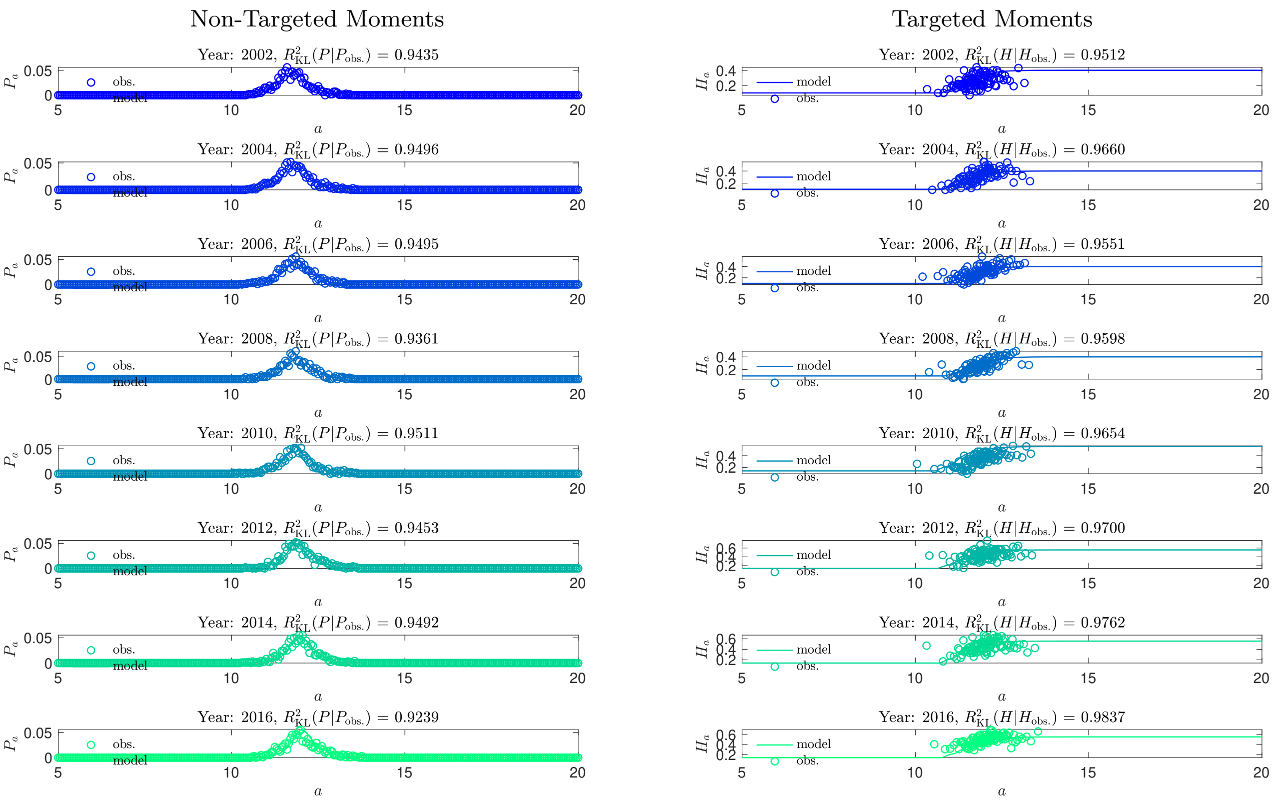

Figure 8: Goodness-of-fit of the estimated model for Switzerland with estimates, based on the estimates in Table 4 and a structural break in 2008. The panels on the left correspond to non-targeted moments, the panels on the right to targeted moments.

5. Counterfactuals

5.1. Parameter Sensitivity Analysis

Our estimated structural growth model allows us to analyze how the productivity growth rate (ν) and dispersion (σ^2) depend on the in-house R&D success probability ($\bar{p}$), R&D costs ($\widetilde{\kappa}$) and the imitation success probabilities of non-R&D and R&D firms (q and δ, respectively).

We first provide a comparative statics analysis in which we compute the change in the productivity growth rate ν between (i) a simulation of the model using the estimated value $\hat{\theta}$ of a parameter $\theta \in \Theta = \{\bar{p}, \widetilde{\kappa}, q, \delta\}$ from Tables 2 and 4 for Switzerland and the Netherlands, respectively and (ii) a counterfactual simulation in which this parameter is set to zero ($\theta = 0$), while all other parameters are held at their estimated values. We report the relative difference, $(\nu(\theta = \hat{\theta}) - \nu(\theta = 0))/\nu(\theta = \hat{\theta})$, in Table 6 for the periods 2000–2008 and 2010–2016. Column (1) presents the results for Switzerland, and column (2) those for the Netherlands. For both countries, the in-house R&D success probability ($\bar{p}$) accounts by far for the largest contribution to the relative change in the productivity growth rate. If firms never succeeded in in-house R&D (i.e. $\bar{p} = 0$) in the 2000–2008 period, productivity growth would be 86.21% lower in Switzerland and 67.02% lower in the Netherlands. By contrast, the effects of the other fundamental parameters on productivity growth are much smaller.

These magnitudes and their relative contributions are relatively stable across the pre- and post-2008 periods. For Switzerland, the reduction in the productivity growth rate associated with setting the innovation success probability to zero ($\bar{p} = 0$) is 4.53 percentage points (p.p.) smaller in the post-2008 period, while the corresponding reductions attributable to imitation (q) and passive imitation (δ) success probabilities are 1.37 p.p. and 1.13 p.p. larger, respectively. The signs of these pre- versus post-2008 changes are the same for the Netherlands.

For both countries, Table 6 illustrates that a lack of success with in-house R&D would be most detrimental to productivity growth, followed by a lack of imitation success (technology diffusion). This is consistent with the trivial observation that a deficiency of innovation would bring growth to a halt in the long run in our model. The smallest impact on growth results from reducing R&D costs. Consequently, policies that are effective in supporting firms' capabilities to successfully innovate or imitate are likely to yield the largest gains in productivity growth.

To get a better understanding of the sensitivity of the productivity growth rate to changes in the fundamental parameters over a broader range of admissible values, Figure 9 shows the relative changes in the productivity growth rate (ν) when varying the in-house R&D success probability ($\bar{p}$), the imitation success probability (q), the passive imitation success probability (δ) and the R&D cost parameter ($\widetilde{\kappa}$), respectively, over a whole range of parameter values for the post-2008 period for Switzerland (top panels) and the Netherlands (bottom panels). The changes are computed relative to a benchmark scenario in which the parameters are set to their estimated values post-2008 (see column (4) in Table 2 for Switzerland and column (4) in Table 4 for the Netherlands). We find that the productivity growth rate ν monotonically increases in $\bar{p}$, q and δ, and decreases in $\widetilde{\kappa}$. Figure 9 corroborates the results of Table 6, showing that by far the largest increase in productivity growth can be achieved by increasing the success probabilities of innovation ($\bar{p}$) and imitation of non-R&D firms (q).

Table 6: The percentage change of the productivity growth rate from the parameter estimate to the parameter being set to zero.

		Switzerland (1)	the Netherlands (2)
Sample period: 2000-2008			
Innovation	($\bar{p} = 0$)	–86.21	–67.02
Imitation	($q = 0$)	–1.92	–13.71
Passive Imitation	($\delta = 0$)	–1.67	–10.82
Cost	($\widetilde{\kappa} = 0$)	+0.05	+0.28
Sample period: 2010-2016			
Innovation	($\bar{p} = 0$)	–81.68	–66.87
Imitation	($q = 0$)	–3.29	–17.40
Passive Imitation	($\delta = 0$)	–2.79	–14.68
Cost	($\widetilde{\kappa} = 0$)	+0.18	+0.45
Change: 2000-2008 to 2010-2016			
Innovation	(p.p.)	+4.53	+0.15
Imitation	(p.p.)	–1.37	–3.69
Passive Imitation	(p.p.)	–1.13	–3.86
Cost	(p.p.)	+0.13	+0.17

Notes: The relative change in the productivity growth rate ν when moving from the estimated parameter value $\hat{\theta}$ to $\theta = 0$ is computed as $(\nu(\theta = \hat{\theta}) - \nu(\theta = 0))/\nu(\theta = \hat{\theta})$, where $\theta \in \bar{p}, \widetilde{\kappa}, q, \delta$. Column (1) reports the results for Switzerland for the pre- and post-2008 periods, based on the parameter estimates in Table 2, while column (2) presents the corresponding results for the Netherlands using the estimates in Table 4.

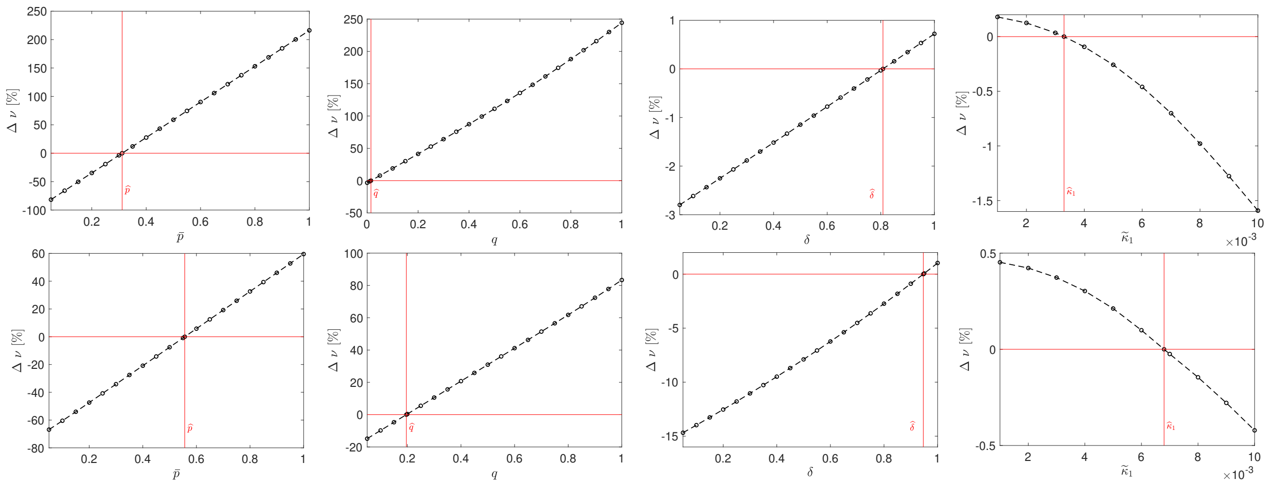

Figure 9: Changes in the productivity growth rate (ν) in Switzerland (top panels) and the Netherlands (bottom panels) when changing the in-house R&D success probability $\bar{p}$ (first column), the imitation success probability q (second column), the passive imitation success probability δ (third column) and the R&D cost parameter $\widetilde{\kappa}$ (last column), respectively, relative to their estimated values post-2008, with the remaining parameters set to their post-2008 estimates. These post-2008 estimates for Switzerland are shown in column (4) of Table 2, and those for the Netherlands in column (4) of Table 4. Vertical lines correspond to the corresponding 2008 estimated values.

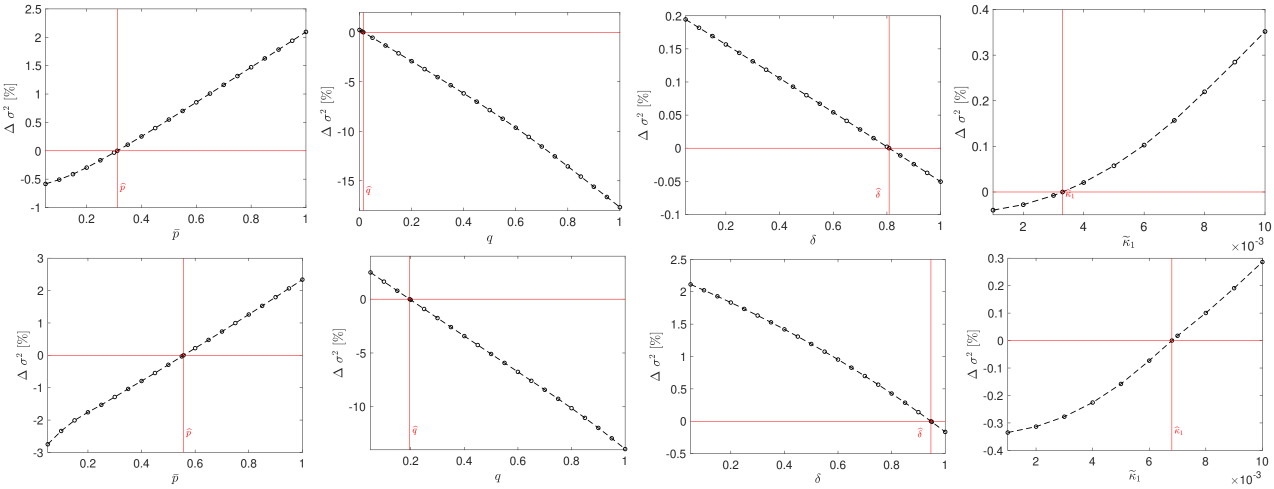

Figure 10: Changes in the productivity variance (σ^2) in Switzerland (top panels) and the Netherlands (bottom panels) when changing the in-house R&D success probability $\bar{p}$ (first column), the imitation success probability q (second column), the passive imitation success probability δ (third column) and the R&D cost parameter $\widetilde{\kappa}$ (last column), respectively, relative to their estimated values post-2008, with the remaining parameters set to their post-2008 estimates. These post-2008 estimates for Switzerland are shown in column (4) of Table 2, and those for the Netherlands in column (4) of Table 4. Vertical lines correspond to the corresponding 2008 estimated values.

How do the fundamental parameters affect productivity dispersion? Figure 10 shows the relative changes in the productivity variance (σ^2) when varying the in-house R&D success probability ($\bar{p}$), the imitation success probability (q), the passive imitation success probability (δ) and the R&D cost parameter ($\widetilde{\kappa}$) for Switzerland and the Netherlands, respectively. For both countries, the variance (σ^2) increases $\bar{p}$) and $\widetilde{\kappa}$, and decreases in q and δ, respectively. This illustrates that policies that increase the innovation success probability ($\bar{p}$) and/or decrease R&D costs ($\widetilde{\kappa}$) not only boost productivity growth but also increase inequality in the economy.

Finally, we simulate a counterfactual for Switzerland in which we apply to the pre-2008 Swiss estimates (reported in column (3) of Table 2) the parameter changes observed in the Netherlands between the pre- and post-2008 periods (reported in Table 4). Imposing the same relative changes on Switzerland implies a 39.92% increase in the productivity growth rate in the counterfactual relative to the benchmark. We interpret the difference between the counterfactual and the benchmark as the potential gain in annual productivity growth had Switzerland experienced the same favorable changes in the fundamental parameters as the Netherlands.

Similarly, we simulate a counterfactual for the Netherlands in which we impose the parameter changes observed for Switzerland between the pre- and post-2008 periods. This simulation exercise mimics a situation in which the Netherlands did not experience extended innovation support. In this counterfactual, the productivity growth rate is 19.17% lower than in the benchmark scenario. We interpret this as evidence that heightened innovation support contributed to an increase of about 20% in the productivity growth rate in the Netherlands.

5.2. Optimal Policy

The comparative statics in Section 5.1 motivate the analysis of an optimal R&D subsidy policy program. In this section, we solve a planner's problem whose objective is to maximize social welfare (consumer plus producer surplus) through revenue taxes, τ_{it}, and R&D subsidies, s_{it}, for $i = 1, \ldots, n$ and time periods $t = 1 \ldots, T$ [cf. Akcigit et al., 2022]. Details of the planner's problem are provided in Supplementary Appendix F.

Period t consumer surplus without transfers is given by (see Equation (B.1) in Supplementary Appendix B):

$$Q_t = \left(\sum_{i=1}^{n} q_{it}^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}}$$

Given firms' profits π_{it}, as defined in Equation (B.11) in Supplementary Appendix B, producer surplus in period t is given by:

$$\Pi_t = \sum_{i=1}^{n} \pi_{it} = \sum_{i=1}^{n} \left((1 - \tau_{it}) p_{it} q_{it} - w L_{it} - (1 - s_{it}) c_{it} \right),$$

where $\tau_{it} \in [0, 1]$ is a sales tax, $s_{it} \in [0, 1]$ is an R&D subsidy and c_{it} denotes the cost of R&D in Equation (2). The firms' participation or incentive constraint is given by

$$\pi_{it} = (1 - \tau_{it}) p_{it} q_{it} - w L_{it} - (1 - s_{it}) c_{it} \geq 0, \ \forall i = 1, \ldots, n,$$

that is, firms' profits after transfers must be non-negative in every period. The government is not allowed to accumulate debt, so the following budget constraint must hold in every period

$$\sum_{i=1}^{n} p_{it} q_{it} \tau_{it} \geq \sum_{i=1}^{n} c_{it} s_{it}. \tag{13}$$

where the transfer payments to the firms in period t are given by $T_t = \sum_{i=1}^{n} c_{it} s_{it} - \sum_{i=1}^{n} p_{it} q_{it} \tau_{it}$ with $T_t \leq 0$. The period t welfare function is given by

$$\overline{W}_t = Q_t + \Pi_t = \left(\sum_{i=1}^{n} q_{it}^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}} + \sum_{i=1}^{n} \left((1 - \tau_{it}) p_{it} q_{it} - w L_{it} - (1 - s_{it}) c_{it} \right),$$

and the planner's problem is[20]

$$\max_{\{\tau_{it}, s_{it}\}} \mathbb{E}\left[\sum_{t=1}^{T} \overline{W}_t\right], \qquad \text{(objective function)}$$

$$\text{s.t. } \overline{W}_t = \left(\sum_{i=1}^{n} q_{it}^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}} - w\sum_{i=1}^{n} L_{it} + \sum_{i=1}^{n}(1-\tau_{it})p_{it}q_{it} - \sum_{i=1}^{n}(1-s_{it})c_{it}, \qquad \text{(social welfare)}$$

$$\forall i: \; \pi_{it} = (1-\tau_{it})p_{it}q_{it} - wL_{it} - (1-s_{it})c_{it} \geq 0, \qquad \text{(incentive constraint)}$$

$$\sum_{i=1}^{n} p_{it}q_{it}\tau_{it} \geq \sum_{i=1}^{n} c_{it}s_{it}, \qquad \text{(budget constraint)}$$

where the expectation in the objective function is over the evolution of the productivity distribution given by Equation (9), the indicator function for whether firm i pursues imitation is

$$\chi^{\text{im}}(a,p,P) = 1 - \chi^{\text{in}}(a,p,P) = \begin{cases} 1 & \text{if } p < \frac{(1-\delta)q(1-F_a)+(1-\tau(a))^{-\eta}(1-s(a))\widehat{\kappa}e^{\theta(\eta-1)(\overline{a}-a)}}{1-\delta q(1-F_a)}, \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

with $\widehat{\kappa} = \kappa/(e^{(\eta-1)\widetilde{a}} - 1)$, $F_a = \sum_{b=1}^{a} P_b$ and the average log-productivity $\overline{a} = \sum_{a=1}^{\infty}(1 - F_{a-1})$.

The planner's problem can be solved numerically to obtain the optimal policies $\{\tau_{it}^*, s_{it}^*\}$ for $i = 1, \ldots, n$ and $t = 1 \ldots, T$ with the law of motion for the productivity distribution from Proposition 1. We initiate the search for a solution using a simulated annealing algorithm with a penalty term for the inequality constraints ($T_t \leq 0$ and $\pi_{it} \geq 0$ for all $i = 1, \ldots, n$ and $t = 1 \ldots, T$) to find a solution. The constraint can be implemented as an additive term $-\mu \times (\max\{0, T_t\} + \sum_{i=1}^{n} \max\{0, -\pi_{it}\})$ to the objective function with penalty parameter $\mu > 0$ [Delahaye et al., 2018; Van Laarhoven et al., 1987]. We then use this candidate solution as an initial condition for a constrained optimization algorithm (such as `fmincon` in `Matlab`).

To better understand the impact of the policy variables, we first investigate how taxes and subsidies affect R&D costs. Figure 11 shows the threshold $a^* = \max\{a \in \mathcal{A} : \chi^{\text{im}}(a,p,P) \geq 1\}$ from Equation (14) (top panels) and total R&D costs $\sum_{t=1}^{T}\sum_{i=1}^{n}(1-s)c_{it}$ with c_{it} defined in Equation (2) in Section 3.1) as a function of the revenue tax rate τ and the R&D subsidy rate s for Switzerland (left panel) and the Netherlands (right panel). The threshold a^* is increasing with the tax τ and decreasing with the subsidy rate s as Equation (2) indicates. Total R&D costs decline with the subsidy rate s and show a non-monotonic relationship with the tax rate τ in the case of Switzerland. To understand the latter, note that the scaling factor Ψ in the R&D cost increases with τ.[21] However,

[20] The planner's problem can be written as a function of the productivity distribution only. We refer to Supplementary Appendix F for details.

[21] In Equation (B.20) in Section B.2, we show that under a uniform revenue tax, the scaling factor can be written as $\Psi = \frac{1}{\eta}(1-\tau)^{1-\eta}\left(\sum_{i=1}^{n} A_{it}^{\eta-1}\right)^{\frac{2-\eta}{\eta-1}}$, which is increasing in τ for $\eta > 1$. Moreover, one can show that $\Psi = \frac{1}{\eta(1-\tau)}\left(\frac{\eta-1}{\eta}\right)^{\eta-2} w^{2-\eta}$, where w is the equilibrium wage rate. As the tax rate τ rises, the wage w falls (for $\eta > 2$), and the scaling factor Ψ increases.

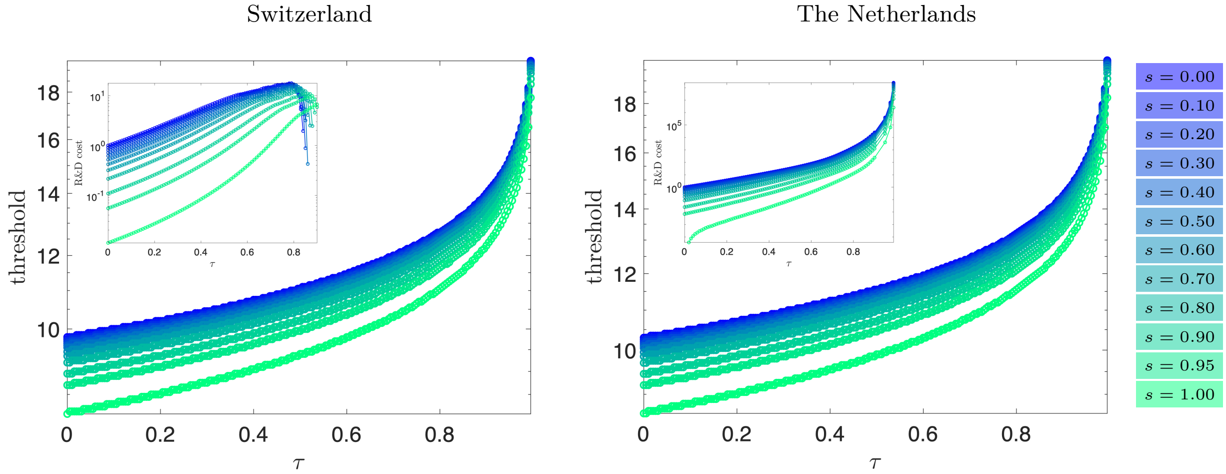

Figure 11: Threshold $a^* = \max\{a \in \mathcal{A} : \chi^{\text{im}}(a, p, P) \geq 1\}$ from Equation (14) with the productivity distribution P in the year 2010 and total R&D costs $\sum_{t=1}^{T}\sum_{i=1}^{n}(1-s)c_{it}$ relative to R&D costs with zero taxes and subsidies (insets) as a function of the revenue tax rate τ and R&D subsidy rate s for Switzerland (left panel) and the Netherlands (right panel).

as τ rises, the threshold in Equation (14) also increases, making firms more likely to imitate rather than to conduct in-house R&D, which eventually reduces effective R&D costs.

Table 7 reports the optimal policy parameters (τ^* and s^*), together with the real-world policies for Switzerland ($\tau = 0.08$, $s = 0$), the Netherlands ($\tau = 0.25$, $s = 0.12$), and the United States ($\tau = 0.23$, $s = 0.19$). The optimal policy for Switzerland is $\tau^* = 2.37 \times 10^{-5}$ and $s^* = 0.99$, while for the Netherlands it is $\tau^* = 9.34 \times 10^{-6}$ and $s^* = 0.98$. As illustrated in Figure 11, R&D costs are mimimized when subsidies are high and taxes are low. However, note that while the optimal tax is indeed relatively small it must be different from zero as the latter would violate the government's budget constraint in Equation (13).

To facilitate comparison across policy regimes, welfare is expressed relative to a benchmark scenario with zero R&D costs ($c_{it} = 0$). Under the optimal tax and subsidy policy, Switzerland attains 102.74 % of the welfare in the no-cost benchmark, while the Netherlands reaches 100.11% of the benchmark level (see row (D) in Table 7). The observed (real-world) policies for Switzerland, the Netherlands, and the United States perform similarly across countries, yielding 97.38%, 86.09%, and 87.43% of the benchmark welfare (without R&D costs) for Switzerland and slightly lower values of 94.80%, 83.69%, and 85.03% for the Netherlands, respectively (see rows (A), (B), and (C) in Table 7, respectively).[22] The gains of the optimal policy exceeding the benchmark in Switzerland is driven

[22]As a robustness check, we have also solved for the optimal R&D policy while explicitly accounting for marginal costs of public funds in the government budget constraint, $(1-\chi)\sum_{i=1}^{n} p_{it}q_{it}\tau_{it} \geq \sum_{i=1}^{n} c_{it}s_{it}$, where $\chi \in [0,1]$ denotes the marginal cost of public funds [Dahlby, 2008]. For a relatively high value of $\chi = 0.5$, the optimal policy for Switzerland implies $\tau = 4.69 \times 10^{-5}$ and $s = 0.99$, while for the Netherlands it implies $\tau = 5.42 \times 10^{-5}$ and $s = 0.96$. Despite these adjustments, the qualitative effectiveness of the optimal policy remains similar to the results reported in Table 7. Overall, introducing marginal costs of public funds leads to higher optimal revenue taxes and lower optimal R&D

by both lower effective R&D costs and a taxation that discourages inefficient R&D while reallocating innovation toward more productive firms [König et al., 2022].

Table 7: Welfare comparison across policies and countries.

			Welfare relative to no R&D cost scenario ($c_{it} = 0$)	
Policy type	τ_{it}	s_{it}	Switzerland	The Netherlands
A. CH policy	0.08	0.00	97.38%	94.80%
B. NL policy	0.25	0.12	86.09%	83.69%
C. US policy	0.23	0.19	87.43%	85.03%
D. Optimal policy	τ^*	s^*	102.74%	100.11%

Notes: The optimal policy for Switzerland is given by $\tau^* = 2.37 \times 10^{-5}$ and $s^* = 0.99$, while for the Netherlands it is $\tau^* = 9.34 \times 10^{-6}$ and $s^* = 0.98$.

6. Conclusion

We analyze the productivity distribution and the R&D decisions of firms in two innovation leaders in Europe, the Switzerland and the Netherlands, using almost 20 years of panel data. These countries are characterized by different trends in R&D activity (a decrease in Switzerland versus an increase in the Netherlands) and pursue contrasting innovation policies (Switzerland relying primarily on indirect R&D promotion, the Netherlands on active support).

To understand the origins and consequences of these divergent patterns, with fewer but highly productive innovators being R&D active in Switzerland versus more but less productive innovators conducting R&D in the Netherlands, we estimate a structural growth model and examine the role of three fundamental drivers of R&D decisions: innovation success, R&D costs, and technology spillovers (imitation success). Relying on a structural stability test, and motivated by the post-2008 rise in R&D active firms in the Netherlands and the 2007 introduction of the Innovation Box, we compare pre- and post-2008 Simulated Method of Moments estimates for both countries. For Switzerland, the estimates point to a decline in the in-house innovation success probability and an increase in R&D costs, in contrast to the opposite pattern observed in the Netherlands. These shifts in fundamentals appear to underlie the diverging trends of R&D activity in the two countries.

Using the estimated model, we run counterfactual simulations to assess how differences in these fundamentals affect productivity growth. Policies that increase innovation or imitation success probabilities are most effective at stimulating productivity growth. However, raising innovation success also increases productivity dispersion, whereas boosting imitation success reduces inequality through

subsidy rates.

spillover effects. We also show that if Switzerland had experienced the same favorable changes in fundamentals as the Netherlands post-2008, its productivity growth would have been 40% higher. Conversely, had the Netherlands undergone the changes observed in Switzerland instead of benefiting from an expanded innovation-supportive environment, its productivity growth would have been 20% lower. Finally, we show that a small revenue tax can optimally finance R&D subsidies. For Switzerland, the resulting policy delivers welfare gains exceeding those in a no–R&D-cost benchmark. These gains arise from both, reducing effective R&D costs and discouraging inefficient R&D, thereby reallocating innovation toward more productive firms.

References

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N., and Kerr, W. (2018). Innovation, reallocation, and growth. *American Economic Review*, 108(11):3450–91.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., and Li, H. (2023). A theory of falling growth and rising rents. *The Review of Economic Studies*, 90(6):2675–2702.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., and Howitt, P. (2005). Competition and innovation: An inverted-U relationship. *The Quarterly Journal of Economics*, 120(2):701–728.

Akcigit, U. (2009). Firm size, innovation dynamics and growth. In *2009 Meeting Papers*, number 1267. Society for Economic Dynamics.

Akcigit, U. and Ates, S. T. (2021). Ten facts on declining business dynamism and lessons from endogenous growth theory. *American Economic Journal: Macroeconomics*, 13(1):257–98.

Akcigit, U., Hanley, D., and Stantcheva, S. (2022). Optimal taxation and R&D policies. *Econometrica*, 90(2):645–684.

Andrews, D., Criscuolo, C., and Gal, P. N. (2015). Frontier firms, technology diffusion and public policy: Micro evidence from OECD countries. Technical report, OECD Working Papers.

Andrews, D., Criscuolo, C., and Gal, P. N. (2016). The best versus the rest: the global productivity slowdown, divergence across firms and the role of public policy. Technical report, OECD Working Papers.

Arora, A., Belenzon, S., and Patacconi, A. (2018). The decline of science in corporate R&D. *Strategic Management Journal*, 39(1):3–32.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Reenen, J. V. (2017). Concentrating on the fall of the labor share. *American Economic Review*, 107(5):180–185.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, 135(2):645–709.

Bartelsman, E., Dobbelaere, S., and Peters, B. (2015). Allocation of human capital and innovation at the frontier: firm-level evidence on Germany and the Netherlands. *Industrial and Corporate Change*, 24(5):875–949.

Benhabib, J., Perla, J., and Tonetti, C. (2014). Catch-up and fall-back through innovation and imitation. *Journal of Economic Growth*, 19:1–35.

Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. (2020). Are ideas getting harder to find? *American Economic Review*, 110(4):1104–1144.

Bloom, N., Van Reenen, J., and Williams, H. (2019). A toolkit of policies to promote innovation. *Journal of Economic Perspectives*, 33(3):163–84.

Buera, F. J. and Oberfield, E. (2020). The global diffusion of ideas. *Econometrica*, 88(1):83–114.

Cameron, A. C. and Windmeijer, F. A. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of econometrics*, 77(2):329–342.

Chen, Z., Liu, Z., Suárez Serrato, J. C., and Xu, D. Y. (2021). Notching R&D investment with corporate income tax cuts in China. *American Economic Review*, 111(7):2065–2100.

Choi, J. and Levchenko, A. A. (2023). The long-term effects of industrial policy. Technical report, National Bureau of Economic Research, Working Paper No. w29263.

Čížek, P. (2008). General trimmed estimation: Robust approach to nonlinear and limited dependent variable models. *Econometric Theory*, 24(6):1500–1529.

Cleves, M., Gould, W., Gutierrez, R., and Marchenko, Y. (2008). *An introduction to survival analysis using Stata*. Stata Press.

Cohen, W. M. (2010). Fifty years of empirical studies of innovative activity and performance. *Handbook of the Economics of Innovation*, 1:129–213.

Dahlby, B. (2008). *The marginal cost of public funds: Theory and applications*. MIT press.

Dechezleprêtre, A., Einiö, E., Martin, R., Nguyen, K.-T., and Van Reenen, J. (2023). Do tax incentives increase firm innovation? An RD design for R&D, patents, and spillovers. *American Economic Journal: Economic Policy*, 15(4):486–521.

Decker, R. A., Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2020). Changing business dynamism and productivity: Shocks versus responsiveness. *American Economic Review*, 110(12):3952–3990.

Delahaye, D., Chaimatanan, S., and Mongeau, M. (2018). Simulated annealing: From basics to applications. In *Handbook of Metaheuristics*, pages 1–35. Springer.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *The American economic review*, 67(3):297–308.

Eurostat (2019). Community innovation survey CIS, database. *Online: https://ec.europa.eu/eurostat/data/database*.

Ghysels, E. and Guay, A. (2003). Structural change tests for simulated method of moments. *Journal of econometrics*, 115(1):91–123.

Goldin, I., Koutroumpis, P., Lafond, F., and Winkler, J. (2024). Why is productivity slowing down? *Journal of Economic Literature*, 62(1):196–268.

Gordon, R. J. (2018). Why has economic growth slowed when innovation appears to be accelerating? Technical report, National Bureau of Economic Research.

Grabska, K., Bettendorf, L., Luginbuhl, R., Meijerink, G., and Elbourne, A. (2017). Productivity slowdown: Evidence for the Netherlands. *CPB Communication*, 9.

Griliches, Z. (1998). *R&D and Productivity: The Econometric Evidence*. University of Chicago Press.

Hall, A. R. (2005). *Generalized method of moments*. Oxford University Press.

Hashmi, A. R. (2013). Competition and innovation: The inverted-U relationship revisited. *The Review of Economics and Statistics*, 95(5):1653–1668.

Hill, J. B. and Renault, E. (2010). Generalized method of moments with tail trimming. *Dept. of Economics, University of North Carolina-Chapel Hill*.

Jones, B. F. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *The Review of Economic Studies*, 76(1):283–317.

Kogan, L., Papanikolaou, D., Seru, A., and Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, 132(2):665–712.

König, M., Storesletten, K., Song, Z., and Zilibotti, F. (2022). From imitation to innovation: Where is all that chinese R&D going? *Econometrica*, 90(4):1615–1654.

König, M. D., Liu, X., and Zenou, Y. (2019). R&D networks: Theory, empirics, and policy implica-

tions. *Review of Economics and Statistics*, 101(3):476–491.
König, M. D., Lorenz, J., and Zilibotti, F. (2016). Innovation vs. imitation and the evolution of productivity distributions. *Theoretical Economics*, 11(3):1053–1102.
Lewbel, A. (2019). The identification zoo: Meanings of identification in econometrics. *Journal of Economic Literature*, 57(4):835–903.
Lokshin, B. and Mohnen, P. (2012). How effective are level-based R&D tax credits? Evidence from the Netherlands. *Applied Economics*, 44(12):1527–1538.
Long, S. J. (1997). *Regression models for categorical and limited dependent variables.* Sage Publications.
McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica:*, pages 995–1026.
Mohnen, P., Vankan, A., and Verspagen, B. (2017). Evaluating the innovation box tax policy instrument in the Netherlands, 2007–13. *Oxford Review of Economic Policy*, 33(1):141–156.
Newey, W. and West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.
Pammolli, F., Magazzini, L., and Riccaboni, M. (2011). The productivity crisis in pharmaceutical R&D. *Nature reviews Drug discovery*, 10(6):428–438.
Park, M., Leahey, E., and Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature*, 613(7942):138–144.
Rammer, C. and Schubert, T. (2016). Concentration on the few? R&D and innovation in German firms between 2001 and 2013. Technical report, Fraunhofer ISI Discussion Papers-Innovation Systems and Policy Analysis.
SERI (2020). Forschung und Innovation in der Schweiz 2020. Technical report, State Secretariat for Education, Research, and Innovation.
Song, K.-S. (2002). Goodness-of-fit tests based on Kullback-Leibler discrimination information. *IEEE Transactions on Information Theory*, 48(5):1103–1117.
Song, Z., Storesletten, K., and Zilibotti, F. (2011). Growing like China. *The American Economic Review*, 101(1):196–233.
Spescha, A. and Wörter, M. (2022). Innovation und Digitalisierung in der Schweizer Privatwirtschaft-Ergebnisse der Innovationserhebung 2020. Technical report, State Secretariat for Education, Research, and Innovation, ETH Zurich, KOF Swiss Economic Institute.
Van Laarhoven, P. J., Aarts, E. H., van Laarhoven, P. J., and Aarts, E. H. (1987). *Simulated annealing.* Springer.
Van Reenen, J. and Yueh, L. (2012). Why has China grown so fast? The role of international technology transfers. *Centre for Economic Performance (CEP) Discussion Paper 1121.*

Supplementary Appendix for "R&D Decisions and Productivity Growth: Evidence from Switzerland and the Netherlands"

Sabien Dobbelaere[a,b,c], Michael D. König[a,b,d,e,f], Andrin Spescha[d] and Martin Wörter[d,f]

[a] *Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands.*
[b] *Tinbergen Institute (TI), Amsterdam, The Netherlands.*
[c] *Institute of Labor Economics (IZA), Bonn, Germany.*
[d] *ETH Zurich, KOF Swiss Economic Institute, Zurich, Switzerland.*
[e] *Centre for Economic Policy Research (CEPR), London, United Kingdom.*
[f] *Leibniz Centre for European Economic Research (ZEW), Germany.*

Contents

A. Descriptive Evidence on the Continuation of R&D Activities

In Section 2.2 in the main text, we document that the decrease (increase) in the fraction of R&D active firms in Switzerland (the Netherlands) comes with an increasing (decreasing) productivity threshold above which firms engage in R&D. In this section, we examine the factors (either firm characteristics or innovation policy instruments) that correlate with a firm's decision to continue doing R&D by running a survival analysis [Cleves et al., 2008]. These factors can be interpreted through the lens of three fundamentals that affect a firm's R&D decision: how likely it is for an innovation to succeed (innovation success probability), how costly it is to pursue an innovation strategy (R&D costs) and how easy it is to imitate others (imitation success probability). The Cox proportional hazards estimates are partial correlations, i.e. reduced-form associations between the firm/policy variables and these fundamentals. Section 3 in the main text develops a structural growth model from these fundamentals and quantifies their impact on aggregate productivity growth.

Survival analysis estimation samples. In the survival analysis, we include all R&D active firms that can either continue or exit R&D. Given these R&D active firms, we investigate whether they exit R&D and if so how long they lasted until they exit. Non-R&D active firms and firms entering R&D are not part of the analysis. The Swiss estimation sample consists of 1,669 firms that have been at risk for exiting from R&D activities. In total, we observe 809 exits between 2000 and 2016. The Dutch estimation sample consists of 3,720 firms, which result in 1,579 exits from R&D between 2000 and 2016. From the Kaplan-Meier estimates, we find that the probability that a Swiss (Dutch) company which conducted R&D in 2000, still conducts R&D in 2010 is 45.9% (40.1%), whereas it drops to 24.9% (27.3%) in 2016.[23]

Baseline Cox model. In the baseline Cox proportional hazard model, we examine the role of firm size (measured by employment in full-time equivalents), firm productivity (measured by real value added per worker), absorptive capacity (measured by the employment share of academics and the share of employees with higher education in Switzerland, and the share of employees with tertiary education in the Netherlands) and access to international markets (measured by export status) in affecting the risk of exit from R&D. In all specifications, we include 2-digit industry dummies to control for unobserved time-invariant heterogeneity at the industry level. Tables A.1 and A.2 show the coefficients for the explanatory variables of the estimated baseline Cox model for Switzerland and the Netherlands, respectively. These coefficients are exponentiated, indicating the extent to which the explanatory variables shift the hazard of exiting from R&D. For example, the hazard ratio of employment of 0.775 in the first column of Table A.1 means that a one percent increase in employment reduces the hazard rate of exiting from R&D by 0.225%. For both countries, all firm covariates of our baseline model are statistically significant and negatively correlated with the hazard rate. Ceteris paribus, large, productive, human-capital intensive and export-oriented companies experience a lower hazard rate of exiting from R&D in both countries.

[23] Results not reported but available upon request.

Table A.1: Baseline Cox proportional hazard model: Exit from R&D in Switzerland.

	(1)	(2)	(3)	(4)
Ln(Employment)	0.775***	0.781***	0.773***	0.786***
	(0.018)	(0.018)	(0.019)	(0.018)
Ln(Value added/employee)		0.769***	0.800***	0.825**
		(0.053)	(0.053)	(0.053)
Share of academics			0.985***	0.988***
			(0.003)	(0.003)
Share of employees with higher education			0.991***	0.991***
			(0.003)	(0.002)
Export (0/1)				0.697***
				(0.049)
Observations	3,307	3,250	3,250	3,250
Industry fixed effects	Yes	Yes	Yes	Yes

Notes: ***/**/* denotes statistical significance at the 1%/5%/10% level.

Table A.2: Baseline Cox proportional hazard model: Exit from R&D in the Netherlands.

	(1)	(2)	(3)	(4)
Ln(Employment)	0.832***	0.839***	0.840***	0.848***
	(0.018)	(0.018)	(0.018)	(0.019)
Ln(Value added/employee)		0.767***	0.779***	0.751***
		(0.043)	(0.044)	(0.043)
Share of high-skilled employees			0.611***	0.618***
			(0.108)	(0.113)
Export (0/1)				0.886**
				(0.052)
Observations	6,299	6,299	6,273	6,109
Industry fixed effects	Yes	Yes	Yes	Yes

Notes: ***/**/* denotes statistical significance at the 1%/5%/10% level.

Competitive vs. less competitive firms. To examine how the survival of competitive firms differs from less competitive ones, we distinguish three types of firms based on the explanatory variables in Tables A.1 and A.2. We consider firms with low values for the respective firm characteristics (that is, values in the 10% quantile of the distribution of the variables), firms with values lying at the median and firms with high values (that is, values in the 90% quantile of the distribution). Firms with high values (that is, large, productive, export-oriented firms with a high-skilled workforce) are generally considered to be more competitive than firms with low values. Figures A.1 and A.2 plot the survivor functions of the three types of firms in Switzerland and the Netherlands, respectively. In both countries, we observe that more competitive firms are significantly less likely to exit R&D

compared to less competitive firms. For example, by the end of our analysis period, Swiss companies scoring high on the aforementioned firm characteristics have a 62% probability of remaining R&D active, compared to 43% for Dutch companies. In contrast, firms with low scores on these characteristics have less than a 5% probability of continuing R&D activities. Swiss (Dutch) firms at the median of the firm characteristics have a probability of about 30% (20%) to continue conducting R&D. Comparing the development of the survivor functions in the two countries highlights two differences, though. First, we observe a flattening of the survivor function from 2010 onward in the Netherlands for all firm types, which is not the case in Switzerland. Second, Dutch competitive firms show a more pronounced decline in the probability to remain R&D active compared to Swiss competitive firms. This suggests that Swiss competitive companies exhibit greater resistance to discontinuing R&D than their Dutch counterparts. One possible reason for this divergence is that there is more formal support for innovation activities in the Netherlands, to which we turn now.

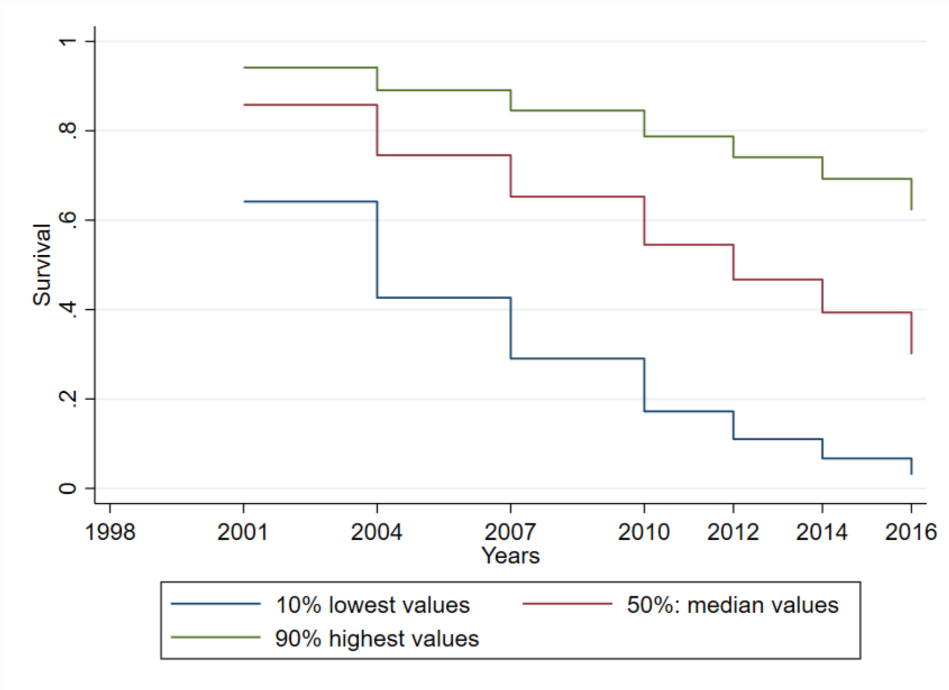

Figure A.1: Survivor functions: Competitive vs. less competitive firms in Switzerland.

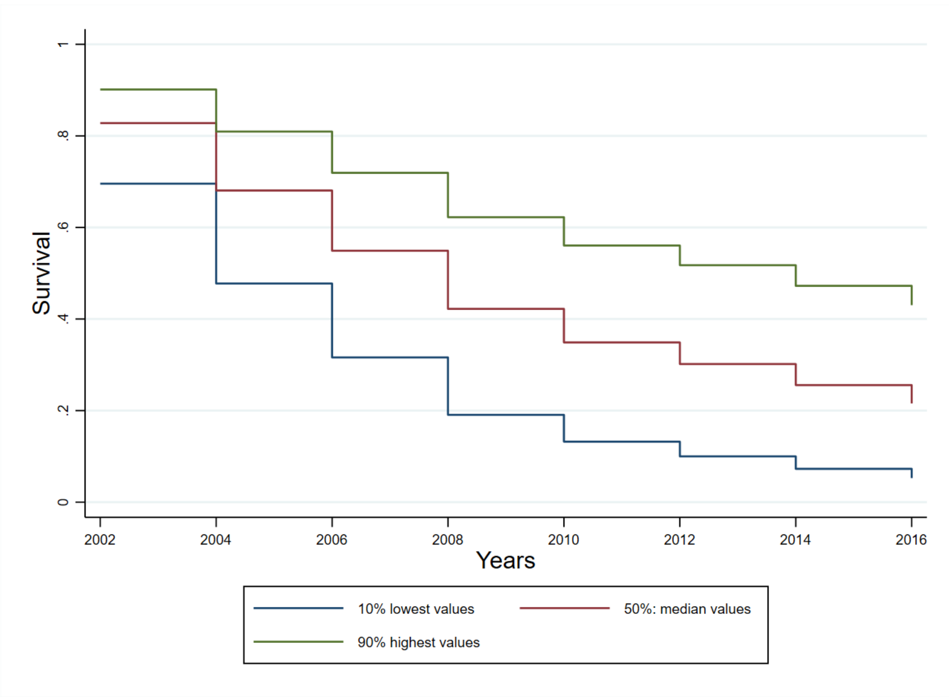

Figure A.2: Survivor functions: Competitive vs. less competitive firms in the Netherlands.

Extended Cox model with innovation input/support. To investigate partial correlations between explanatory variables that measure the innovation input as well as the innovation support of firms and the hazard rate of exiting from R&D, we enrich our baseline model specification (specification (4) in Tables A.1 and A.2) and retain the firm covariates from the baseline estimations as control variables. As reported in Table A.3, we include R&D expenditures, R&D cooperation with universities, R&D cooperation with other research institutes, and domestic and international innovation support when estimating this extended model on the Swiss micro-data. These variables are all negatively associated with the hazard rate of exiting R&D. Except for the coefficients of R&D cooperation with other research institutes and international innovation support, the exponentiated coefficients are statistically significant and below the value of one. For example, the coefficient of R&D expenditures is 0.859, indicating that a one percent increase in R&D expenditures reduces the hazard of exiting R&D by 0.141%. Domestic innovation support shows a coefficient of 0.425, which implies that the hazard of exiting R&D is only half as large for firms enjoying domestic innovation support than for firms without such support. Innovation support in Switzerland does not include any financial contribution to the R&D activities of the company. Instead, the funding is directed exclusively to the university partner of the project. The company must bear its own share of the costs associated with the cooperation project. This suggests that even “soft” measures of public innovation promotion, focusing primarily on supporting knowledge and technology transfer between the public research sector and industry rather than contributing financially to R&D activities of companies, are effective in keeping Swiss firms R&D active.

To estimate the extended model on the Dutch micro-data, we include R&D expenditures, R&D cooperation, government funding (excluding Patent/Innovation Box), funding from the EU (framework program) and being a Patent/Innovation Box user, in addition to the explanatory variables from the baseline estimation. Table A.4 shows that Dutch firms engaging intensively in R&D activities (either through innovation input or cooperation) and Dutch firms receiving financial support from the government display a much lower hazard of exiting R&D. Dutch firms spending the same amount on R&D are much less likely to exit from R&D than their Swiss counterparts (see exponentiated coefficient of 0.761 as compared to 0.859 in Table A.3). Table A.4 suggests that the Innovation Box policy instrument might be quite effective in keeping firms R&D active: being a Patent/Innovation Box user decreases the hazard of exiting R&D by 64%.

Receiving indirect innovation policy support, such as promoting knowledge and technology transfer between universities and the private sector as in Switzerland, may influence the firm's R&D decision by enhancing the ease of generating innovations (innovation success probability). Receiving direct innovation policy support, such as lowering taxes on expected profits from R&D through the Innovation Box in the Netherlands, may influence the firm's R&D investment decision by reducing the cost associated with conducting R&D. Note, however, that the estimated coefficients from Cox proportional hazard models are simple, partial correlations. As such, we are not in a position to establish causal relationships. For example, firms that benefit from the Innovation Box might differ in observable and unobservable characteristics from those that cannot, extending beyond the measurable observables included in our estimations. These hidden characteristics might in fact cause the observed difference.

Table A.3: Cox proportional hazard model with innovation input/support: Exit from R&D in Switzerland.

	(1)	(2)	(3)
Ln(R&D expenditures)	0.859***		
	(0.019)		
R&D Cooperation with universities		0.709***	
		(0.077)	
R&D Cooperation with other research institutes		0.918	
		(0.104)	
Innovation support domestic			0.425***
			(0.102)
Innovation support international			0.611
			(0.213)
Observations	3,175	3,224	1,807
Controls	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes

Notes: ***/**/* denotes statistical significance at the 1%/5%/10% level.

Table A.4: Cox proportional hazard model innovation input/support: Exit from R&D in the Netherlands.

	(1)	(2)	(3)	(4)
Ln(R&D expenditures)	0.761***			
	(0.015)			
R&D Cooperation		0.725***		
		(0.037)		
Funding from government (excl. Patent/Innovation Box)			0.441***	
			(0.028)	
Funding from EU (excl. framework program)			0.975	
			(0.122)	
Funding from EU framework program			0.728*	
			(0.130)	
Patent/Innovation Box user				0.363***
				(0.061)
Observations	5,255	6,109	5,209	6,109
Control	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes

Notes: ***/**/* denotes statistical significance at the 1%/5%/10% level.

B. Economic Environment

This section provides a complete description of the economic environment. To simplify the notation we drop time indices in the following.

B.1. Final Goods Sector

Consider a monopolistic competition model à la Dixit and Stiglitz [1977] with heterogeneous firms.[24] The representative consumer's utility function from consumption of the final good is given by

$$U = \left(\sum_{i=1}^{n} q_i^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}} \tag{B.1}$$

where q_i is the quantity of consumption of good i and $\eta > 1$ is the elasticity of substitution among varieties. A higher η makes utility less sensitive to the number of varieties consumed.

Consumers want to maximize utility U subject to the budget constraint, I. The corresponding

[24]For brevity, we omit time indices in this and the next section.

Lagrangian function is given by

$$\mathscr{L} = \left(\sum_{i=1}^{n} q_i^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}} - \lambda\left(\sum_{i=1}^{n} p_i q_i - I\right)$$

The first-order condition is given by

$$\frac{\partial \mathscr{L}}{\partial q_i} = \frac{\eta}{\eta-1}\left(\sum_{i=1}^{n} q_i^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}-1} \frac{\eta-1}{\eta} q_i^{-\frac{1}{\eta}} - \lambda p_i = 0.$$

This gives

$$\left(\sum_{i=1}^{n} q_i^{\frac{\eta-1}{\eta}}\right)^{\frac{1}{\eta-1}} q_i^{-\frac{1}{\eta}} = \lambda p_i$$

We then can write

$$\frac{q_i}{q_j} = \left(\frac{p_i}{p_j}\right)^{-\eta}$$

That is

$$q_i = q_j\left(\frac{p_i}{p_j}\right)^{-\eta} \tag{B.2}$$

and

$$p_i q_i = p_i q_j\left(\frac{p_i}{p_j}\right)^{-\eta} = p_i^{1-\eta} p_j^{\eta} q_j$$

Hence, we have that

$$I = \sum_{i=1}^{n} p_i q_i = q_j p_j^{\eta} \sum_{i=1}^{n} p_i^{1-\eta}$$

Solving for q_j gives

$$q_j p_j^{\eta} = \frac{I}{\sum_{i=1}^{n} p_i^{1-\eta}}$$

since $I = \sum_{i=1}^{n} p_i q_i$. We next define the price index as

$$P = \left(\sum_{i=1}^{n} p_i^{1-\eta}\right)^{\frac{1}{1-\eta}}$$

so that the demand function can be written as

$$q_j = \frac{p_j^{-\eta} I}{P^{1-\eta}} \tag{B.3}$$

The elasticity of demand is given by

$$\begin{aligned}\frac{dq_j}{dp_j}\frac{p_j}{q_j} &= -\frac{\eta p_j^{-\eta-1} I}{P^{1-\eta}}\frac{p_j}{q_j} \\ &= \frac{-\eta p_j^{-\eta} I P^{1-\eta}}{P^{1-\eta} p_j^{-\eta} I} \\ &= -\eta \qquad \text{(B.4)}\end{aligned}$$

where firms' pricing decisions are assumed not to influence the price index P. This is a valid approximation when there are many firms, or in a small open economy where world prices are assumed to be fixed.

Note that the price index P equals the cost of obtaining one unit of utility. To see this, use the first-order in Equation (B.2) to obtain

$$\begin{aligned}U &= \left(\sum_{i=1}^{n} q_i^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}} \\ &= \left(\sum_{i=1}^{n}\left(q_j\left(\frac{p_i}{p_j}\right)^{-\eta}\right)^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}} \\ &= \left(\sum_{i=1}^{n} q_j^{\frac{\eta-1}{\eta}} p_j^{\eta-1} p_i^{1-\eta}\right)^{\frac{\eta}{\eta-1}} \\ &= q_j p_j^{\eta}\left(\sum_{i=1}^{n} p_i^{1-\eta}\right)^{\frac{\eta}{\eta-1}} \\ &= q_j p_j^{\eta} P^{-\eta} \qquad \text{(B.5)}\end{aligned}$$

Inserting Equation (B.3) into Equation (B.5) gives

$$q_j = \frac{p_j^{-\eta} I}{P^{1-\eta}} \qquad \text{(B.6)}$$

Inserting Equation (B.6) into Equation (B.5) gives

$$U = \frac{I}{P^{1-\eta}} P^{-\eta} = \frac{I}{P}$$

and hence

$$P = \frac{I}{U}$$

In the following, we normalize the price index to one, $P = \left(\sum_{i=1}^{n} p_i^{1-\eta}\right)^{\frac{1}{1-\eta}} = 1$. Then, firm i's demand function is given by

$$q_i = p_i^{-\eta} I \qquad \text{(B.7)}$$

where aggregate income (or revenue) is given by $I = \sum_{i=1}^{n} p_i q_i$. Since setting $P = 1$ implies $U = I$, it follows that

$$U = I = \left(\sum_{i=1}^{n} q_i^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}} = Q$$

which is the constant elasticity of substitution (CES) aggregate of firms' outputs.

B.2. Intermediate Goods Sector

The production function of the intermediate good q_i produced by firm i is given by a labor-only, linear production technology

$$q_i = A_i L_i, \tag{B.8}$$

where A_i denotes the firm-specific productivity of firm i [cf. e.g. Buera and Oberfield, 2020]. From firm i's demand function (Equation (B.7)), we have that

$$p_i = q_i^{-\frac{1}{\eta}} I^{\frac{1}{\eta}}. \tag{B.9}$$

Inserting Equation (B.9) into Equation (B.8) gives

$$q_i p_i = I^{\frac{1}{\eta}} q_i^{\frac{\eta-1}{\eta}} = I^{\frac{1}{\eta}} \left(A_i L_i \right)^{\frac{\eta-1}{\eta}} \tag{B.10}$$

The profit function of firm i is given by

$$\pi_i = (1 - \tau_i) p_i q_i - w L_i - (1 - s_i) c_i \tag{B.11}$$

where c_i is firm i's R&D cost, s_i the R&D subsidy rate, w the wage, τ_i the revenue tax, and L_i the labor force employed by firm i. The total labor force is normalized to one, $\sum_{i=1}^{n} L_i = 1$.

Firms maximize profits by choosing labor optimally. The corresponding first-order condition is given by

$$\frac{\partial \pi_i}{\partial L_i} = (1 - \tau_i) \frac{\eta - 1}{\eta} I^{\frac{1}{\eta}} A_i^{\frac{\eta-1}{\eta}} L_i^{\frac{\eta-1}{\eta} - 1} - w = 0 \tag{B.12}$$

Equation (B.12) can simplified to

$$(1 - \tau_i) \frac{\eta - 1}{\eta} \frac{p_i q_i}{L_i} = w \tag{B.13}$$

From Equation (B.12), we get

$$w = (1 - \tau_i) \frac{\eta - 1}{\eta} I^{\frac{1}{\eta}} A_i^{\frac{\eta-1}{\eta}} L_i^{-\frac{1}{\eta}}$$

and therefore

$$L_i = \left(\frac{\eta-1}{\eta w}\right)^{\eta} I(1-\tau_i)^{\eta} A_i^{\eta-1} \propto A_i^{\eta-1}. \tag{B.14}$$

Next, inserting Equation (B.14) into Equation (B.10) yields for firm i's revenues

$$\begin{aligned}
p_i q_i &= I^{\frac{1}{\eta}} \left(A_i L_i\right)^{\frac{\eta-1}{\eta}} \\
&= I^{\frac{1}{\eta}} A_i^{\frac{\eta-1}{\eta}} L_i^{\frac{\eta-1}{\eta}} \\
&= I^{\frac{1}{\eta}} A_i^{\frac{\eta-1}{\eta}} \left(\left(\frac{\eta-1}{\eta w}\right)^{\eta} I(1-\tau_i)^{\eta} A_i^{\eta-1}\right)^{\frac{\eta-1}{\eta}} \\
&= I \left(\frac{\eta-1}{\eta w}\right)^{\eta-1} (1-\tau_i)^{\eta-1} A_i^{\eta-1}. \qquad \text{(B.15)}
\end{aligned}$$

Inserting Equation (B.13) into firm i's profit function (Equation (B.11)) gives

$$\begin{aligned}
\pi_i &= (1-\tau_i)p_i q_i - (1-\tau)\frac{\eta-1}{\eta} p_i q_i - (1-s_i)c_i \\
&= \frac{1}{\eta}(1-\tau_i)p_i q_i - (1-s_i)c_i.
\end{aligned} \tag{B.16}$$

Inserting revenues (Equation (B.15)) into Equation (B.16) gives

$$\pi_i = \frac{I}{\eta}(1-\tau_i)^{\eta} \left(\frac{\eta-1}{\eta w}\right)^{\eta-1} A_i^{\eta-1} - (1-s_i)c_i.$$

Further, summation over i in Equation (B.15) gives

$$I = \sum_{i=1}^{n} p_i q_i = I \left(\frac{\eta-1}{\eta w}\right)^{\eta-1} \sum_{i=1}^{n} (1-\tau_i)^{\eta-1} A_i^{\eta-1}.$$

Canceling I on both sides and solving for the equilibrium wage gives

$$w = \frac{\eta-1}{\eta} \left(\sum_{i=1}^{n} (1-\tau_i)^{\eta-1} A_i^{\eta-1}\right)^{\frac{1}{\eta-1}}. \tag{B.17}$$

It then follows that

$$\left(\frac{\eta-1}{\eta w}\right)^{\eta-1} = \frac{1}{\sum_{i=1}^{n} (1-\tau_i)^{\eta-1} A_i^{\eta-1}}$$

Profits are given by

$$\pi_i = \frac{I}{\eta}(1-\tau_i)^{\eta} \left(\frac{\eta-1}{\eta w}\right)^{\eta-1} A_i^{\eta-1} - (1-s_i)c_i.$$

Hence,

$$\pi_i \;=\; \frac{I}{\eta}\,\frac{(1-\tau_i)^{\eta}A_i^{\eta-1}}{\sum_{j=1}^{n}(1-\tau_j)^{\eta-1}A_j^{\eta-1}} \;-\; (1-s_i)c_i.$$

Inserting the equilibrium wage (Equation (B.17)) back into Equation (B.15) gives

$$p_i q_i = I\frac{A_i^{\eta-1}}{\sum_{j=1}^{n}A_j^{\eta-1}}$$

Since total labor supply is normalized to 1, Equation (B.14) implies

$$1 = \sum_{i=1}^{n} L_i = I\left(\frac{\eta-1}{\eta w}\right)^{\eta}\sum_{i=1}^{n}(1-\tau_i)^{\eta}A_i^{\eta-1} \tag{B.18}$$

Inserting the equilibrium wage (Equation (B.17)) into Equation (B.18) gives aggregate expenditures

$$I = \frac{\left(\sum_{i=1}^{n}(1-\tau_i)^{\eta-1}A_i^{\eta-1}\right)^{\frac{\eta}{\eta-1}}}{\sum_{i=1}^{n}(1-\tau_i)^{\eta}A_i^{\eta-1}}.$$

We can then write profits as follows

$$\pi_i = \frac{1}{\eta}\,\frac{\left(\sum_{j=1}^{n}(1-\tau_j)^{\eta-1}A_j^{\eta-1}\right)^{\frac{1}{\eta-1}}}{\sum_{j=1}^{n}(1-\tau_j)^{\eta}A_j^{\eta-1}}\,(1-\tau_i)^{\eta}A_i^{\eta-1} \;-\; (1-s_i)c_i.$$

In the following, we denote by

$$\Psi = \frac{1}{\eta}\,\frac{\left(\sum_{j=1}^{n}(1-\tau_j)^{\eta-1}A_j^{\eta-1}\right)^{\frac{1}{\eta-1}}}{\sum_{j=1}^{n}(1-\tau_j)^{\eta}A_j^{\eta-1}}, \tag{B.19}$$

so profits can be written compactly as

$$\pi_i = \Psi\,(1-\tau_i)^{\eta}A_i^{\eta-1} \;-\; (1-s_i)c_i.$$

In the special case of a uniform tax ($\tau_i = \tau$), we have

$$\Psi = \frac{1}{\eta}(1-\tau)^{1-\eta}\left(\sum_{i=1}^{n}A_i^{\eta-1}\right)^{\frac{1}{\eta-1}-1}. \tag{B.20}$$

C. Uniformly Distributed R&D Success Probability

In this section, we simplify Equation (9) in Section 3.6 of the main text by assuming, as in König et al. [2022], that the in-house R&D success probability is uniformly distributed, consistent with the assumption used when estimating the structural model.

Lemma 1. *Assuming a uniform distribution of the in-house R&D success probability with support* $[0, \overline{p}]$ *allows us to write Equation (9) as follows*

$$\begin{aligned}
&\frac{\partial P_a(t)}{\partial t} = \frac{1}{\overline{p}} \Big[q(1-F_{a-1}(t))P_{a-1}(t) \left(\min\{C(a-1,P),\overline{p}\}\right. \\
&\left. +\delta\frac{1}{2}\left(\overline{p}(2-\overline{p}) - C(a-1,P)(2-C(a-1,P))\right) \mathbb{1}_{\{C(a-1,P)<\overline{p}\}}\right) \\
&-q(1-F_a(t))P_a(t)\left(\min\{C(a,P),\overline{p}\} + \delta\frac{1}{2}\left(\overline{p}(2-\overline{p}) - C(a,P)(2-C(a,P))\right)\mathbb{1}_{\{C(a,P)<\overline{p}\}}\right) \\
&-\frac{1}{2}P_a(t)\left(\overline{p}^2 - C(a,P)^2\right)\mathbb{1}_{\{C(a,P)<\overline{p}\}} + \frac{1}{2}P_{a-1}(t)\left(\overline{p}^2 - C(a-1,P)^2\right)\mathbb{1}_{\{C(a-1,P)<\overline{p}\}}\Big], \qquad \text{(C.1)}
\end{aligned}$$

where we have denoted by

$$C(a,P) \equiv \frac{(1-\delta)q(1-F_a) + \widetilde{\kappa}e^{\theta(\eta-1)(\overline{a}-a)}}{1-\delta q\,(1-F_a)}. \qquad \text{(C.2)}$$

Note that $C(a,P)$ is non-negative and decreasing in a. Let the threshold productivity be defined as $a^* = \{\min a \in \mathcal{A} : C(a,P) < \overline{p}\} = \{\max a \in \mathcal{A} : C(a,P) \geq \overline{p}\}$. Then for all $a \leq a^*$, we have that $\min\{C(a-1,P),\overline{p}\} = \min\{C(a,P),\overline{p}\} = \overline{p}$ and $\mathbb{1}_{\{C(a,P)<\overline{p}\}} = \mathbb{1}_{\{C(a-1,P)<\overline{p}\}} = 0$ so that we can write Equation (C.1) for all $a \leq a^*$ as follows

$$\frac{\partial P_a(t)}{\partial t} = q(1-F_{a-1}(t))P_{a-1}(t) - q(1-F_a(t))P_a(t).$$

Then for $a = a^*+1$, we have that $\min\{C(a-1,P),\overline{p}\} = \overline{p}$ but $\min\{C(a,P),\overline{p}\} = C(a,P)$ and $\mathbb{1}_{\{C(a,P)<\overline{p}\}} = 1$ but $\mathbb{1}_{\{C(a-1,P)<\overline{p}\}} = 0$ so that for $a = a^*+1$, we can write Equation (C.1) as follows

$$\begin{aligned}
\frac{\partial P_a(t)}{\partial t} = \frac{1}{\overline{p}} \Big[& q(1-F_{a-1}(t))P_{a-1}(t)\overline{p} - q(1-F_a(t))P_a(t)\Big(C(a,P) + \delta\frac{1}{2}\big(\overline{p}(2-\overline{p}) \\
& - C(a,P)(2-C(a,P))\big)\Big) - \frac{1}{2}P_a(t)\left(\overline{p}^2 - C(a,P)^2\right)\Big],
\end{aligned}$$

For all $a > a^*+1$, we have that $\min\{C(a-1,P),\overline{p}\} = C(a-1,P)$ and $\min\{C(a,P),\overline{p}\} = C(a,P)$

and $\mathbb{1}_{\{C(a,P)<\overline{p}\}} = \mathbb{1}_{\{C(a-1,P)<\overline{p}\}} = 1$ so that Equation (C.1) for all $a > a^* + 1$ can be written as

$$\begin{aligned}\frac{\partial P_a(t)}{\partial t} = \frac{1}{\overline{p}} \Bigg[& q(1-F_{a-1}(t))P_{a-1}(t) \left(C(a-1,P) + \delta \frac{1}{2} \left(\overline{p}(2-\overline{p}) - C(a-1,P)(2-C(a-1,P))\right) \right) \\ & -q(1-F_a(t))P_a(t) \left(C(a,P) + \delta \frac{1}{2} \left(\overline{p}(2-\overline{p}) - C(a,P)(2-C(a,P))\right) \right) \\ & -\frac{1}{2} P_a(t) \left(\overline{p}^2 - C(a,P)^2\right) + \frac{1}{2} P_{a-1}(t) \left(\overline{p}^2 - C(a-1,P)^2\right) \Bigg]. \end{aligned} \tag{C.3}$$

We obtain the evolution of the productivity distribution $P_a(t)$ by numerically solving the system of ordinary differential equations provided in Equation (C.3) for a given initial condition $P_a(0)$.

D. Proofs

This section presents the proof of Proposition 1 in Section 3.6 of the main text. For simplicity, we consider the case in which all firms have the same in-house R&D success probability p. The generalization to heterogeneous probabilities is straightforward.

Proof of Proposition 1. The evolution of the log-productivity distribution $P_a(t)$ can be written as

$$\begin{aligned} P_a(t+\Delta t) - P_a(t) = & \left(\chi^{\text{im}}(a-1,p,P) + \delta\chi^{\text{in}}(a-1,p,P)(1-p)\right) q(1-F_{a-1}(t))P_{a-1}(t) \\ & - \left(\chi^{\text{im}}(a,p,P) + \delta\chi^{\text{in}}(a,p,P)(1-p)\right) q(1-F_a(t))P_a(t) \\ & + \chi^{\text{in}}(a-1,P)pP_{a-1}(t) - \chi^{\text{in}}(a,P)pP_a(t), \end{aligned} \tag{D.1}$$

where δ is the passive imitation probability, and

$$\chi^{\text{im}}(a,p,P) = 1 - \chi^{\text{in}}(a,p,P) = \begin{cases} 1 & \text{if } p < \frac{(1-\delta)q(1-F_a)+\widetilde{\kappa}e^{\theta(\eta-1)(\overline{a}-a)}}{1-\delta q(1-F_a)}, \\ 0 & \text{otherwise}, \end{cases} \tag{D.2}$$

with the average log-productivity given by $\bar{a} = \sum_{a=1}^{\infty} F_a$.

The first term in Equation (D.1) corresponds to the case in which a firm with log-productivity $a-1$ is selected, times the indicator that it wants to imitate, $\chi^{\text{im}}(a-1,p,P) = 1$, or that it wants to innovate, failed to do so and then engages in passive imitation, $\chi^{\text{in}}(a-1,p,P)(1-p)\delta$, times the probability that it draws a firm with log-productivity larger than $a-1$ and successfully improves its log-productivity by one unit with probability q. The second term corresponds to the event that a firm with log-productivity a is selected and successfully imitates. The third term corresponds to the case in which a firm with log-productivity $a-1$ is selected, wants to innovate, $\chi^{\text{in}}(a-1,p,P) = 1-\chi^{\text{im}}(a-1,p,P) = 1$, and succeeds to improve its log-productivity by one unit with probability p. The fourth term corresponds to the case in which a firm with log-productivity a is selected, wants to innovate, $\chi^{\text{in}}(a,p,P) = 1-\chi^{\text{im}}(a,p,P) = 1$, and succeeds with probability p. Finally, it follows from Equation (D.1) that for all $t \geq 0$: $\sum_{a=1}^{\infty}(P_a(t+\Delta t) - P_a(t)) = 0$. □

Proof of Lemma 1. Assuming a uniform distribution of the in-house R&D success probability allows us to write Equation (9) as follows

$$\frac{\partial P_a(t)}{\partial t} = \frac{1}{\overline{p}-\underline{p}} \int_{[\underline{p},\overline{p}]} \Big[\left(\chi^{\text{im}}(a-1,p,P) + \delta(1-p)\chi^{\text{in}}(a-1,p,P)\right) q(1-F_{a-1}(t))P_{a-1}(t)$$
$$- \left(\chi^{\text{im}}(a,p,P) + \delta(1-p)\chi^{\text{in}}(a,p,P)\right) q(1-F_a(t))P_a(t) + \chi^{\text{in}}(a-1,p,P)pP_{a-1}(t) - \chi^{\text{in}}(a,p,P)pP_a(t)\Big]\, dp.$$

Using the fact that

$$\int_{[\underline{p},\overline{p}]} \chi^{\text{im}}(a,p,P)dp = \int_{[\underline{p},\overline{p}]} \mathbb{1}_{\{p<C(a,P)\}}dp = (\min\{C(a,P),\overline{p}\} - \underline{p})\mathbb{1}_{\{C(a,P)>\underline{p}\}},$$

and

$$\begin{aligned}\int_{[\underline{p},\overline{p}]} p\chi^{\text{in}}(a,p,P)dp &= \int_{[\underline{p},\overline{p}]} p\mathbb{1}_{\{p>C(a,P)\}}dp \\ &= \int_{[\max\{\underline{p},C(a,P)\},\overline{p}]} pdp\mathbb{1}_{\{C(a,P)<\overline{p}\}} \\ &= \frac{1}{2}\left(\overline{p}^2 - (\max\{\underline{p},C(a,P)\})^2\right)\mathbb{1}_{\{C(a,P)<\overline{p}\}},\end{aligned}$$

and

$$\begin{aligned}\int_{[\underline{p},\overline{p}]} (1-p)\chi^{\text{in}}(a,p,P)dp &= \int_{[\underline{p},\overline{p}]} (1-p)\mathbb{1}_{\{p>C(a,P)\}}dp \\ &= \int_{[\max\{\underline{p},C(a,P)\},\overline{p}]} (1-p)dp\mathbb{1}_{\{C(a,P)<\overline{p}\}} \\ &= \frac{1}{2}\, p(2-p)\Big|_{\max\{\underline{p},C(a,P)\}}^{\overline{p}} \mathbb{1}_{\{C(a,P)<\overline{p}\}} \\ &= \frac{1}{2}\left(\overline{p}(2-\overline{p}) - \max\{\underline{p},C(a,P)\}(2-\max\{\underline{p},C(a,P)\})\right)\mathbb{1}_{\{C(a,P)<\overline{p}\}},\end{aligned}$$

where we denote by

$$C(a,P) \equiv \frac{(1-\delta)q(1-F_a) + \widetilde{\kappa}e^{\theta(\eta-1)(\overline{a}-a)}}{1-\delta q\,(1-F_a)}, \tag{D.3}$$

we can write

$$\frac{\partial P_a(t)}{\partial t} = \frac{1}{\overline{p}-\underline{p}}\Big[q(1-F_{a-1}(t))P_{a-1}(t)\Big((\min\{C(a-1,P),\overline{p}\}-\underline{p})\mathbb{1}_{\{C(a-1,P)>\underline{p}\}}$$
$$+\ \delta\frac{1}{2}\left(\overline{p}(2-\overline{p})-\max\{\underline{p},C(a-1,P)\}(2-\max\{\underline{p},C(a-1,P)\})\right)\mathbb{1}_{\{C(a-1,P)<\overline{p}\}}\Big)$$
$$-q(1-F_a(t))P_a(t)\Big((\min\{C(a,P),\overline{p}\}-\underline{p})\mathbb{1}_{\{C(a,P)>\underline{p}\}}$$
$$+\ \delta\frac{1}{2}\left(\overline{p}(2-\overline{p})-\max\{\underline{p},C(a,P)\}(2-\max\{\underline{p},C(a,P)\})\right)\mathbb{1}_{\{C(a,P)<\overline{p}\}}\Big)$$
$$-\frac{1}{2}P_a(t)\left(\overline{p}^2-(\max\{\underline{p},C(a,P)\})^2\right)\mathbb{1}_{\{C(a,P)<\overline{p}\}}$$
$$+\frac{1}{2}P_{a-1}(t)\left(\overline{p}^2-(\max\{\underline{p},C(a-1,P)\})^2\right)\mathbb{1}_{\{C(a-1,P)<\overline{p}\}}\Big]. \tag{D.4}$$

From Equation (D.3), we observe that $C(a,P)\geq 0$. Further, assuming $\underline{p}=0$ allows us to write

$$\frac{\partial P_a(t)}{\partial t} = \frac{1}{\overline{p}}\left[q(1-F_{a-1}(t))P_{a-1}(t)\left(\min\{C(a-1,P),\overline{p}\}\right.\right.$$
$$\left.+\ \delta\frac{1}{2}\left(\overline{p}(2-\overline{p})-C(a-1,P)(2-C(a-1,P))\right)\mathbb{1}_{\{C(a-1,P)<\overline{p}\}}\right)$$
$$-q(1-F_a(t))P_a(t)\left(\min\{C(a,P),\overline{p}\}\right.$$
$$\left.+\ \delta\frac{1}{2}\left(\overline{p}(2-\overline{p})-C(a,P)(2-C(a,P))\right)\mathbb{1}_{\{C(a,P)<\overline{p}\}}\right)$$
$$-\frac{1}{2}P_a(t)\left(\overline{p}^2-C(a,P)^2\right)\mathbb{1}_{\{C(a,P)<\overline{p}\}}$$
$$\left.+\frac{1}{2}P_{a-1}(t)\left(\overline{p}^2-C(a-1,P)^2\right)\mathbb{1}_{\{C(a-1,P)<\overline{p}\}}\right]. \tag{D.5}$$

□

Figure D.1 shows the log-productivity distribution $P_a(t)$ from a numerical solution of Equation (D.5).

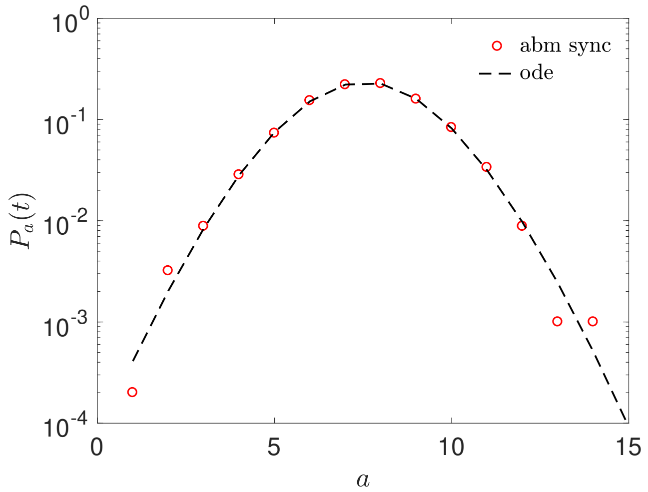

Figure D.1: The log-productivity distribution $P_a(t)$ from a Monte Carlo simulation of the stochastic process indicated with circles. The dashed line indicates the solution of the ordinary differential equation (ode) of Equation (D.5).

E. Robustness Analysis: Balanced panel

This section reports a robustness check in which we re-estimated the structural model presented in Section 3 in the main text using a balanced panel that includes only firms with at least one observation in both the pre- and post-2008 periods. The estimates are reported in Tables E.1 and E.2 for Switzerland and the Netherlands, respectively.

For Switzerland, the findings are qualitatively consistent with those in Table 2 in the main text, except that, in the balanced panel, R&D costs and the imitation success probability of R&D firms appear to have decreased and increased, respectively, post-2008. This pattern is likely driven by the balanced panel being dominated by larger firms, for which R&D costs are relatively less important than the in-house R&D success rate in affecting R&D decisions. For the Netherlands, the balanced panel estimates are qualitatively similar to those in Table 4 in the main text, with the exception that the imitation success probability of R&D firms shows a slight decline in the post-2008 period.

Table E.1: Estimation results for Switzerland with a balanced panel.

		OLS	SMM		
			Pooled	Breakpoint in 2008	
		w/o passive imitation	with passive imitation	with passive imitation	with passive imitation
		$\delta = 0$	$\delta \neq 0$	$\delta \neq 0$	$\delta \neq 0$
		(1)	(2)	(3)	(4)
Periods		2004–2016	2004–2016	2004–2008	2010–2016
Innovation	($\overline{p}$)	0.4573***	0.3470***	0.3726***	0.3084***
		(0.0050)	(0.0201)	(0.0147)	(0.0159)
Imitation	(q)	0.1859***	0.0596	0.0055	0.0059
		(0.0138)	(0.9168)	(0.0304)	(0.0455)
Passive Imitation	(δ)		0.9847***	0.4802***	0.9975***
			(0.0701)	(0.0003)	(0.0003)
Cost Mean	($\widetilde{\kappa}$)	0.0068**	0.0003	0.0007	0.0005
		(0.0030)	(0.0045)	(0.0043)	(0.0047)
Cost Spread	(σ_κ)		0.0091	0.0076	0.0062
			(0.0139)	(0.0087)	(0.0114)
Observations		8,618	8,618	2,697	5,799
Firms		3,540	3,540	2,039	2,849
$R^2_{\mathrm{KL}}(P\|P_{\mathrm{obs}})$			0.8967	0.8635	0.9103
$R^2_{\mathrm{KL}}(H\|H_{\mathrm{obs}})$			0.9152	0.9076	0.9140
J_T-stat.			1.1002	11.1246	0.7813
Change Before/After 2008 Breakpoint					
Innovation	($\Delta\overline{p}/\overline{p}$)				-17.23%
Imitation	($\Delta q/q$)				7.11%
Passive Imitation	($\Delta\delta/\delta$)				107.73%
Cost Mean	($\Delta\widetilde{\kappa}/\widetilde{\kappa}$)				-24.37%
Cost Spread	($\Delta\sigma_\kappa/\sigma_\kappa$)				-18.70%
W_T-stat.				3.3032×10^{12}	
LM_T-stat.				0.0606	
D_T-stat.				151.5784	
O_T-stat.				3.1904	
O_{1T}-stat.				1.0080	
O_{2T}-stat.				2.1824	

Notes: Column (1) presents the OLS estimates for Equation (6). Columns (2)–(4) report the estimation results from the SMM algorithm. Standard errors are computed using the heteroscedasticity and autocorrelation consistent (HAC) optimal weighting matrix [Newey and West, 1987] and reported in parentheses. The asterisks *** /**/* indicate significance at the 1%/5%/10% level. The R-squared goodness-of-fit measure is defined as $R^2_{\mathrm{KL}}(\cdot \mid \cdot) = 1 - D_{\mathrm{KL}}(\cdot \mid \cdot)$, where $D_{\mathrm{KL}}(H \mid H_{\mathrm{obs}}) = \sum_a H(a) \log\left(H(a)/H_{\mathrm{obs}}(a)\right)$ is the Kullback–Leibler divergence, and is reported as an average across time periods. The J_T-statistic tests the overidentifying restrictions. The Wald (W_T), Lagrange Multiplier (LM_T), D-(D_T), and O-statistics (O_T, O_{1T}, and O_{2T}) are structural stability tests.

Table E.2: Estimation results for the Netherlands with a balanced panel.

		OLS	SMM		
			Pooled	Breakpoint in 2008	
		w/o passive imitation	with passive imitation	with passive imitation	with passive imitation
		$\delta = 0$	$\delta \neq 0$	$\delta \neq 0$	$\delta \neq 0$
		(1)	(2)	(3)	(4)
Periods		2000–2016	2000–2016	2000–2008	2010–2016
Innovation	$(\overline{p})$	0.5527***	0.4311***	0.3902***	0.4521***
		(0.0014)	(0.0004)	(0.0057)	(0.0074)
Imitation	(q)	0.2750***	0.0602	0.1216	0.2015
		(0.0022)	(0.6517)	(0.3415)	(0.3040)
Passive Imitation	(δ)		0.99997	0.9960***	0.9717***
			(0.6687)	(0.1604)	(0.0683)
Cost Mean	$(\widetilde{\kappa})$	0.0002***	0.0009	0.0030*	0.0007
		(0.0001)	(0.0014)	(0.0016)	(0.0005)
Cost Spread	(σ_κ)		0.0024	0.0004*	0.0006
			(0.0033)	(0.0010)	(0.0005)
Observations		17,721	17,721	9,378	8,343
Firms		4,384	4,384	4,384	4,268
$R^2_{\text{KL}}(P\|P_{\text{obs}})$			0.9195	0.9267	0.9290
$R^2_{\text{KL}}(H\|H_{\text{obs}})$			0.93922	0.9322	0.9388
J_T-stat.			1.0133	7.1062	1.4311
Change Before/After 2008 Breakpoint					
Innovation	$(\Delta\overline{p}/\overline{p})$				+15.87%
Imitation	$(\Delta q/q)$				+65.73%
Passive Imitation	$(\Delta\delta/\delta)$				–2.44%
Cost Mean	$(\Delta\widetilde{\kappa}/\widetilde{\kappa})$				–75.94%
Cost Spread	$(\Delta\sigma_\kappa/\sigma_\kappa)$				+38.70%
W_T-stat.				400.8318	
LM_T-stat.				0.7497	
D_T-stat.				4132.6426	
O_T-stat.				5.2365	
O_{1T}-stat.				2.1794	
O_{2T}-stat.				3.0570	

Notes: Column (1) presents the OLS estimates for Equation (6). Columns (2)–(4) report the estimation results from the SMM algorithm. Standard errors are computed using the heteroscedasticity and autocorrelation consistent (HAC) optimal weighting matrix [Newey and West, 1987] and reported in parentheses. The asterisks *** /**/* indicate significance at the 1%/5%/10% level. The R-squared goodness-of-fit measure is defined as $R^2_{\text{KL}}(\cdot \mid \cdot) = 1 - D_{\text{KL}}(\cdot \mid \cdot)$, where $D_{\text{KL}}(H \mid H_{\text{obs}}) = \sum_a H(a) \log \big(H(a)/H_{\text{obs}}(a)\big)$ is the Kullback–Leibler divergence, and is reported as an average across time periods. The J_T-statistic tests the overidentifying restrictions. The Wald (W_T), Lagrange Multiplier (LM_T), D- (D_T), and O-statistics (O_T, O_{1T}, and O_{2T}) are structural stability tests.

F. Planner's Problem

In this section, we present the planner's problem that underlies the optimal policy framework discussed in Section 5.2 of the main text. The planner's problem is given by:

$$\max_{\{\tau_{it},s_{it}\}} \mathbb{E}\left[\sum_{t=1}^{T}\overline{W}_t\right], \qquad \text{(objective function)}$$

$$\text{s.t. } \overline{W}_t = \left(\sum_{i=1}^{n} q_{it}^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}} - w\sum_{i=1}^{n} L_{it} + \sum_{i=1}^{n}(1-\tau_{it})p_{it}q_{it} - \sum_{i=1}^{n}(1-s_{it})c_{it}, \qquad \text{(social welfare)}$$

$$\forall i:\ \pi_{it} = (1-\tau_{it})p_{it}q_{it} - wL_{it} - (1-s_{it})c_{it} \geq 0, \qquad \text{(incentive constraint)}$$

$$\sum_{i=1}^{n} p_{it}q_{it}\tau_{it} \geq \sum_{i=1}^{n} c_{it}s_{it}. \qquad \text{(budget constraint)}$$

Recall that, using the firms' first-order conditions, we showed in Supplementary Appendix B that

$$wL_{it} = (1-\tau_{it})\frac{\eta-1}{\eta}p_{it}q_{it}$$

and

$$\pi_{it} = \frac{1}{\eta}(1-\tau_{it})p_{it}q_{it} - (1-s_{it})c_{it}.$$

and

$$p_{it}q_{it} = I\left(\frac{\eta-1}{\eta w}\right)^{\eta-1}(1-\tau_{it})^{\eta-1}A_{it}^{\eta-1}$$

so that the planner's problem can be written as

$$\max_{\{\tau_{it},s_{it}\}} \mathbb{E}\left[\sum_{t=1}^{T}\overline{W}_t\right], \qquad \text{(objective function)}$$

$$\text{s.t. } \overline{W}_t = \left(\sum_{i=1}^{n} q_{it}^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}} + \sum_{i=1}^{n}\frac{2\eta-1}{\eta}(1-\tau_{it})p_{it}q_{it} - \sum_{i=1}^{n}(1-s_{it})c_{it}, \qquad \text{(social welfare)}$$

$$\forall i:\ \pi_{it} = \frac{1}{\eta}(1-\tau_{it})p_{it}q_{it} - (1-s_{it})c_{it} \geq 0, \qquad \text{(incentive constraint)}$$

$$\sum_{i=1}^{n} p_{it}q_{it}\tau_{it} \geq \sum_{i=1}^{n} c_{it}s_{it}, \qquad \text{(budget constraint)}$$

$$p_{it}q_{it} = I\left(\frac{\eta-1}{\eta w}\right)^{\eta-1}(1-\tau_{it})^{\eta-1}A_{it}^{\eta-1}. \qquad \text{(firm revenues)}$$

Next, using that $q_{it}^{\frac{\eta-1}{\eta}} = I^{-\frac{1}{\eta}} p_{it} q_{it}$ (under the assumption $P = 1$), we obtain

$$\left(\sum_{i=1}^{n} q_{it}^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}} = \left(\sum_{i=1}^{n} I^{-\frac{1}{\eta}} p_{it} q_{it}\right)^{\frac{\eta}{\eta-1}} = I^{-\frac{1}{\eta-1}} \left(\sum_{i=1}^{n} p_{it} q_{it}\right)^{\frac{\eta}{\eta-1}} = I^{-\frac{1}{\eta-1}} I^{\frac{\eta}{\eta-1}} = I,$$

with $\sum_{i=1}^{n} p_{it} q_{it} = I$, and denoting by

$$\Omega = \left(\frac{\eta-1}{\eta w}\right)^{\eta-1},$$

this can be written more compactly as follows

$$\max_{\{\tau_{it}, s_{it}\}} \mathbb{E}\left[\sum_{t=1}^{T} \overline{W}_t\right], \qquad \text{(objective function)}$$

$$\text{s.t. } \overline{W}_t = I + I\Omega \frac{2\eta-1}{\eta} \sum_{i=1}^{n} (1-\tau_{it})^{\eta} A_{it}^{\eta-1} - \sum_{i=1}^{n} (1-s_{it}) c_{it}, \qquad \text{(social welfare)}$$

$$\forall i: \frac{I\Omega}{\eta}(1-\tau_{it})^{\eta} A_{it}^{\eta-1} \geq (1-s_{it}) c_{it}, \qquad \text{(incentive constraint)}$$

$$I\Omega \sum_{i=1}^{n} (1-\tau_{it})^{\eta-1} A_{it}^{\eta-1} \tau_{it} \geq \sum_{i=1}^{n} c_{it} s_{it}, \qquad \text{(budget constraint)}$$

$$c_{it} = \frac{I\Omega}{\eta} \kappa \left(\overline{A}(t)^{\theta} A_i(t)^{1-\theta}\right)^{\eta-1} \chi^{\text{in}}(A_{it}, p, P(t)). \qquad \text{(R\&D costs)}$$

Normalizing the total labor force to one, the equilibrium wage is given by[25]

$$w = \frac{\eta-1}{\eta} \left(\sum_{i=1}^{n} (1-\tau_{it})^{\eta-1} A_{it}^{\eta-1}\right)^{\frac{1}{\eta-1}}, \qquad \text{(F.1)}$$

and aggregate income is

$$I = \frac{\left(\sum_{i=1}^{n} (1-\tau_{it})^{\eta-1} A_{it}^{\eta-1}\right)^{\frac{\eta}{\eta-1}}}{\sum_{i=1}^{n} (1-\tau_{it})^{\eta} A_{it}^{\eta-1}}.$$

Hence

$$\Omega = \frac{1}{\sum_{i=1}^{n} (1-\tau_{it})^{\eta-1} A_{it}^{\eta-1}},$$

and

$$I\Omega = \frac{\left(\sum_{i=1}^{n} (1-\tau_{it})^{\eta-1} A_{it}^{\eta-1}\right)^{\frac{1}{\eta-1}}}{\sum_{i=1}^{n} (1-\tau_{it})^{\eta} A_{it}^{\eta-1}}.$$

[25] In the special case of a uniform tax ($\tau_{it} = \tau$), the wage rate w equals $\frac{\eta-1}{\eta}(1-\tau)\widetilde{A}$, where $\widetilde{A} = \left(\sum_{i=1}^{n} A_{it}^{\eta-1}\right)^{\frac{1}{\eta-1}}$.

The planner's problem can then be written as a function of productivity only

$$\max_{\{\tau_{it},s_{it}\}} \mathbb{E}\left[\sum_{t=1}^{T} \overline{W}_t\right], \qquad \text{(objective function)}$$

$$\text{s.t. } \overline{W}_t = \frac{\left(\sum_{i=1}^{n}(1-\tau_{it})^{\eta-1}A_{it}^{\eta-1}\right)^{\frac{1}{\eta-1}}}{\sum_{i=1}^{n}(1-\tau_{it})^{\eta}A_{it}^{\eta-1}} \left(\sum_{i=1}^{n}\left(1+\frac{2\eta-1}{\eta}(1-\tau_{it})\right)(1-\tau_{it})^{\eta-1}A_{it}^{\eta-1}\right.$$

$$\left. -\frac{\kappa}{\eta}\sum_{i=1}^{n}(1-s_{it})\left(\overline{A}(t)^{\theta}A_i(t)^{1-\theta}\right)^{\eta-1}\chi^{\text{in}}(A_{it},p,P(t))\right), \qquad \text{(social welfare)}$$

$$\forall i:\ (1-\tau_{it})^{\eta}A_{it}^{\eta-1} \geq (1-s_{it})\kappa\left(\overline{A}(t)^{\theta}A_i(t)^{1-\theta}\right)^{\eta-1}\chi^{\text{in}}(A_{it},p,P(t)), \qquad \text{(incentive constraint)}$$

$$\sum_{i=1}^{n}(1-\tau_{it})^{\eta-1}A_{it}^{\eta-1}\tau_{it} \geq \frac{\kappa}{\eta}\sum_{i=1}^{n}\left(\overline{A}(t)^{\theta}A_i(t)^{1-\theta}\right)^{\eta-1}\chi^{\text{in}}(A_{it},p,P(t))s_{it}. \qquad \text{(budget constraint)}$$

Moreover, with firm-specific taxes, the equilibrium wage satisfies

$$w = \frac{\eta-1}{\eta}\left(\sum_{j=1}^{n}(1-\tau_{jt})^{\eta-1}A_{jt}^{\eta-1}\right)^{\frac{1}{\eta-1}},$$

which yields

$$\left(\frac{\eta-1}{\eta w}\right)^{\eta-1} = \frac{1}{\sum_{j=1}^{n}(1-\tau_{jt})^{\eta-1}A_{jt}^{\eta-1}}.$$

Profits are given by

$$\pi_{it} = \frac{I}{\eta}(1-\tau_{it})^{\eta}\left(\frac{\eta-1}{\eta w}\right)^{\eta-1}A_{it}^{\eta-1} - (1-s_{it})c_{it}.$$

Hence,

$$\pi_{it} = \frac{I}{\eta}\frac{(1-\tau_{it})^{\eta}A_{it}^{\eta-1}}{\sum_{j=1}^{n}(1-\tau_{jt})^{\eta-1}A_{jt}^{\eta-1}} - (1-s_{it})c_{it}.$$

Eliminating also I using $I = \left(\sum_{j=1}^{n}(1-\tau_{jt})^{\eta-1}A_{jt}^{\eta-1}\right)^{\frac{\eta}{\eta-1}} / \left(\sum_{j=1}^{n}(1-\tau_{jt})^{\eta}A_{jt}^{\eta-1}\right)$, then gives

$$\begin{aligned}\pi_{it} &= \frac{1}{\eta}\frac{\left(\sum_{j=1}^{n}(1-\tau_{jt})^{\eta-1}A_{jt}^{\eta-1}\right)^{\frac{1}{\eta-1}}}{\sum_{j=1}^{n}(1-\tau_{jt})^{\eta}A_{jt}^{\eta-1}}(1-\tau_{it})^{\eta}A_{it}^{\eta-1} - (1-s_{it})c_{it}\\ &= \frac{I\Omega}{\eta}(1-\tau_{it})^{\eta}A_{it}^{\eta-1} - (1-s_{it})c_{it}.\end{aligned}$$

Recall that we defined the scale factor Ψ as $\left(\sum_{j=1}^{n}(1-\tau_{jt})^{\eta-1}A_{jt}^{\eta-1}\right)^{\frac{1}{\eta-1}} / \sum_{j=1}^{n}(1-\tau_{jt})^{\eta}A_{jt}^{\eta-1}$ (see Equation (B.19) in Supplementary Appendix B.2), so that profits can be written compactly as

$\pi_{it} = \Psi\,(1-\tau_{it})^{\eta} A_{it}^{\eta-1} - (1-s_{it})c_{it}$.[26]

The expected profit from imitation is given by

$$\mathbb{E}_i^{\text{im}}\left[\pi_i(t+\Delta t)\right| A_i\,(t)\,, P(t)] = q\,(1-F_{a_{it}}(t))\,\Psi(1-\tau_{it})^{\eta} A_{it}^{\eta-1}\widetilde{A}^{\eta-1} + (1-q\,(1-F_{a_{it}}(t)))\,\Psi(1-\tau_{it})^{\eta} A_{it}^{\eta-1},$$

In terms of log-productivities $a_{it} = \log A_{it}$, $\overline{a}(t) = \log \overline{A}(t)$ and $\log \widetilde{A} = \widetilde{a}$, we can write

$$\mathbb{E}_i^{\text{im}}\left[\pi_i(t+\Delta t)\right| a_{it}, P(t)] = \Psi(1-\tau_{it})^{\eta} e^{(\eta-1)a_{it}}\left(1+q\,(1-F_{a_{it}}(t))\left(e^{(\eta-1)\widetilde{a}}-1\right)\right).$$

The expected profit from innovation is given by

$$\mathbb{E}_i^{\text{in}}[\pi_i(t+\Delta t)|\, A_{it}, p_i(t), P(t)] = p_i(t)\Psi(1-\tau_{it})^{\eta} A_{it}^{\eta-1}\widetilde{A}^{\eta-1} - (1-s_{it})\Psi\kappa\left(\overline{A}(t)^{\theta} A_{it}^{1-\theta}\right)^{\eta-1}$$
$$+ (1-p_i(t)) \times \left\{\delta\left[q\,(1-F_{a_{it}}(t))\,\Psi(1-\tau_{it})^{\eta} A_{it}^{\eta-1}\widetilde{A}^{\eta-1} + (1-q\,(1-F_{a_{it}}(t)))\,\Psi(1-\tau_{it})^{\eta} A_{it}^{\eta-1}\right]\right.$$
$$\left. +(1-\delta)\Psi(1-\tau_{it})^{\eta} A_{it}^{\eta-1}\right\}. \tag{F.2}$$

In terms of log-productivities, this can be written as

$$\mathbb{E}_i^{\text{in}}[\pi_i(t+\Delta t)|\, a_{it}, p_i(t), P(t)] = p_{it}(t)\Psi(1-\tau_{it})^{\eta} e^{(\eta-1)(a_{it}+\widetilde{a})} - (1-s_{it})\Psi\kappa e^{(\eta-1)\theta\overline{a}(t)} e^{(\eta-1)(1-\theta)a_{it}}$$
$$+ (1-p_i(t)) \times \left\{\delta\left[q\,(1-F_{a_{it}}(t))\,\Psi(1-\tau_{it})^{\eta} e^{(\eta-1)(a_{it}+\widetilde{a})} + (1-q\,(1-F_{a_{it}}(t)))\,\Psi(1-\tau_{it})^{\eta} e^{(\eta-1)a_{it}}\right]\right.$$
$$\left. +(1-\delta)\Psi(1-\tau_{it})^{\eta} e^{(\eta-1)a_{it}}\right\}.$$

Finally, denoting by

$$\widehat{\kappa} = \frac{\kappa}{e^{(\eta-1)\widetilde{a}}-1},$$

we can then write the indicator function for whether firm i pursues imitation as

$$\chi^{\text{im}}(a,p,P) = 1-\chi^{\text{in}}(a,p,P) = \begin{cases} 1 & \text{if } p < \frac{(1-\delta)q(1-F_a)+(1-\tau(a))^{-\eta}(1-s(a))\widehat{\kappa}e^{\theta(\eta-1)(\overline{a}-a)}}{1-\delta q(1-F_a)}, \\ 0 & \text{otherwise.} \end{cases} \tag{F.3}$$

[26] In the special case of a uniform tax ($\tau_{it} = \tau$), we have $\Psi = \frac{1}{\eta}(1-\tau)^{1-\eta}\left(\sum_{i=1}^{n} A_{it}^{\eta-1}\right)^{\frac{1}{\eta-1}-1}$ (see Equation (B.20) in Supplementary Appendix B.2) and profits given by $\pi_{it} = \frac{1}{\eta}\left(\sum_{i=1}^{n} A_{it}^{\eta-1}\right)^{\frac{1}{\eta-1}-1}(1-\tau)A_{it}^{\eta-1} - (1-s_{it})c_{it}$.