Documents de travail

«Data Production and the coevolving AI trajectories: An attempted evolutionary model»

Auteurs

Andrea Borsato, Andre Lorentz

Document de Travail n° 2022 – 09

Mars 2022

Bureau d'Économie Théorique et Appliquée BETA

www.beta-economics.fr

@beta_economics

Contact :
jaoulgrammare@beta-cnrs.unistra.fr

Data production and the coevolving AI trajectories: An attempted evolutionary model

Andrea Borsato and André Lorentz *

March 11, 2022

Abstract

This paper contributes to the understanding of the relationship between the nature of data and the Artificial Intelligence (AI) technological trajectories. We develop an agent-based model in which firms are data producers that compete on the markets for data and AI. The model is enriched by a public sector that fuels the purchase of data and trains the scientists that will populate firms as workforce. Through several simulation experiments we analyze the determinants of each market structure, the corresponding relationships with innovation attainments, the pattern followed by labour and data productivity, and the quality of data traded in the economy. More precisely, we question the established view in the literature on industrial organization according to which technological imperatives are enough to experience divergent industrial dynamics on both the markets for data and AI blueprints. Although technical change behooves if any industry pattern is to emerge, the actual unfolding is not the outcome of a specific technological trajectory, but the result of the interplay between technology-related factors and the availability of data-complementary inputs such as labour and AI capital, the market size, preferences and public policies.

JEL Classification: L10, L60, O33, O38.

Keywords: Artificial Intelligence, Data Markets, Industrial Dynamics, Agent-based Models.

*Université de Strasbourg, Université de Lorraine, CNRS BETA, 67000 Strasbourg. Contact the corresponding author at: Andrea Borsato, Office 323 Avenue de la Forêt-Noire, 67000 Strasbourg, France. E-mail address: aborsato@unistra.fr.

Contents

1 Introduction and Motivation

We can interpret last 150 years of economic history as a complex process driven by automation. Artificial Intelligence (AI), defined as "machines or agents that are capable of observing their environment, learning, and based on the knowledge and experience gained, taking intelligent action or proposing decisions" (Annoni et al., 2018), seems a natural candidate for a further and important breakthrough in scientific and economic progress, as the steam engine or the electricity were in the past (Aghion et al., 2017; Cockburn et al., 2018; LeCun et al., 2015). Besides the potential capability to further speed up automation, AI entails profound changes in the structure of scientific systems. There is indeed a growing body of literature that emphasizes its capability to re-shape the way knowledge itself is produced *within* and *between* many scientific domains (Agrawal et al., 2019; Bianchini et al., 2020). From this point of view, AI affects two main channels through which the production of knowledge occurs. On the one hand, the *search* process unveils how AI spurs an explosion of data that individual researchers find hard to stay abreast of. On the other hand, AI impacts on the *discovery* process, which relates to the proper combination of existing wisdom to get new knowledge.

The dramatic improvements in AI technologies have fueled a plethora of academic works that investigate the manifold relationships AI weaves with the economic systems. The nature of AI as a General Purpose Technology (GPT) or as a Method of Invention (IMI) livens up the debates between Agrawal et al. (2019); Bianchini et al. (2020); Cockburn et al. (2018); Klinger et al. (2020, 2018); Vannuccini and Prytkova (2021). Conversely, Acemoglu et al. (2020); Acemoglu and Restrepo (2019); Aghion et al. (2019); Bordot (2022); Brynjolfsson et al. (2019); Domini et al. (2021); Furman and Seamans (2019), among the many others, elaborate on the effects of the introduction of AI-based systems in terms of economic growth and employment, while Ahmed and Wahed (2020); Armstrong et al. (2016); Gundersen and Kjensmo (2018); Hagendorff and Meding (2020); Nagaraj et al. (2020); Traub et al. (2019) deal with ethical issues related to uneven, or unequal, access to AI resources.

In this paper we focus on the fuel of AI: *data*. The huge increase in data collection and data availability unleashes new market opportunities in which their collection, storage and treatment are only some example of data-fueled business models (Vannuccini and Prytkova, 2021; Yu et al., 2021). Looking at data as the *new oil* (Economist, 2017; Savona, 2019; Varian, 2019) and AI systems as data-hungry, we analyze their relationships and mutual influences. In which way does their interaction determine the corresponding technological trajectories? How do they evolve? Do trajectories diverge, converge, or do they co-exist and move in parallel? What are the underlying *selection mechanisms*? In which way does the public sector enter the game and what could it do? We frame this broad research agenda and try to answer to these questions with an evolutionary agent-based model (ABM) in which firms are data producers that compete on different data and AI markets. Moreover, they engage in the labour market to hire best AI-building scientists. The economy is demand-led since the public sector and another exogenous entity fuel the purchase of data. Moreover, both firms and the government are involved in the training of scientists.

This model offers also a contribution to the understanding of industrial-dynamics issues raised by Malerba et al. (1999, 2001) and Malerba and Orsenigo (1995, 1996b). In particular, our results suggest that technological imperatives are not enough to experience divergent industrial dynamics. Although the economy needs technical change to live and flourish, yet this is not the driving force. The pattern followed by each industry unfolds not as the outcome of a specific technological trajectory, but as the result of the interplay between that very trajectory and the availability of data-complementary inputs, i.e., labour and AI capital, market size, consumer preferences and government policies.

The paper is organized as follows: Section II reviews the relevant literature; Section III elaborates on the nature of data and works as a prelude to the ABM we devise in Sections IV and V; last Section concludes.

2 Relation with the literature

Latest, potentially groundbreaking, improvements in Artificial Intelligence have fostered the research interest in several directions. Among the many fields that deal with the impacts of Artificial Intelligence on the economy, we disentangle three main branches. The first concerns to the relationships between AI and economic growth and employment. Introducing AI in the production of goods and services, and trying to reconcile the ongoing automation with the constancy of factors shares, Aghion et al. (2019) envisage a cost disease *à la* Baumol in which growth is determined not by what the economy is good for but by what is essential and yet hard to improve. Moreover, AI might discourage future innovations if a fast rate of creative destruction limits returns to innovations. However, they underline that AI can obviate to the role of population in generating economic growth, through new ideas.[1] Furman and Seamans (2019) and Acemoglu et al. (2020) investigate the implications of Artificial Intelligence on US labour market. Although their empirical analyses confirm a surge in AI activity after 2014, and estimate consistent and robust changes in the skills demanded by high exposure establishments, these works do not detect any negative rela-

[1]This finding is in agreement with Acemoglu and Restrepo (2017) which find no negative relationship between aging and economic growth; by contrast, countries undergoing rapid demographic changes are more likely to adopt new automation technologies.

tionship between AI exposure and employment or wage levels at the aggregate level. In other terms, albeit some sectors experience labour market upheaval following the introduction of new automation systems, AI technologies are still in their infancy and have spread only through a limited part of the US economy. With respect to these studies, our paper elaborates on a different issue, that is the threat of labour shortage and the importance of public policies in sustaining the growing and training of those very scientists that devise AI technologies. Adopting neoclassical terms for the sake of the argument, and agreeing with Aghion et al. (2019), growth is limited by what is hard to improve, i.e., labour supply, whose *marginal productivity* becomes infinite in this setting.

The second line of research is about Schumpeterian evolutionary theories on industrial organization. The pioneering works by Schumpeter (1934, 1942) and Nelson and Winter (1982) have been enriched by plentiful analyses at the core of the evolutionary literature on technological change and industrial dynamics.[2] The common idea this tradition puts forth is that the history of a number of industries is characterized by a succession of eras associated with specific dominant technologies. The rise of a new technology with a design far from established and satisfactory fosters an aggressive competition between new firms that dismantles users of previous technologies. When the technological design stabilizes in a later phase of its life cycle, the industry tends to concentrate because of dynamic economies of scale on both the supply side and the demand side.

The earlier works stressed supply side factors. For instance, Malerba and Orsenigo (1995, 1996b) suggest that technological regimes denoted by conditions of opportunity, appropriability and cumulativeness in the knowledge space determine the patterns of innovative activities among technological classes within industries. Furthermore, technological advances are associated with degrees of asymmetries among innovators, their stability in the ranking and to the economic size of companies. All these determinants are industry specific.

In agreement with Malerba et al. (2007), we show that demand side dynamics are equally important. In this mentioned article, the authors highlight the emergence of a dominant design as a result of network externalities and bandwagon effects. Our model instead provides a different mechanism according to which the same technological trajectory leads to market structures and patterns of innovation that differ across industries. Therefore, technology-related factors are not enough, or not crucial, to explain the divergent dynamics on both the markets for data and AI. Technological change, whilst essential to engender whatsoever pattern, must interact with other determinants to generate that specific market structure.

[2]Discussing all the relevant articles may divert our focus. For the sake of simplicity, we refer to Dosi (1982, 1988), Malerba and Orsenigo (1995, 1996b, 2002), Malerba et al. (1999, 2007), and Silverberg and Verspagen (1994, 1995). Dosi and Nelson (2010) offer a literature review of uncommon clearness.

Moreover, advances in technologies are not associated with concentration by default, but fierce competition can be typical of a long-run dynamics as well.

Finally, the third line of research relates to the advancement of Artificial Intelligence, its potential GPT-ness and the role of data. We have already mentioned that current performance in AI has propped an interesting debate with contrasting views on whether we should consider it as a GPT or even and IMI. On the one hand, Agrawal et al. (2019); Bianchini et al. (2020); Cockburn et al. (2018) and Klinger et al. (2020, 2018) share the idea that advances in machine learning and neural networks appear to have great potential as a research tool in problems of classification and prediction. Additionally, AI-based learning may be able to speed up the automation in discovery across many domains and to expand the playbook in the sense of uncovering the set of tasks which could be feasibly addressed. The approaches the scientific community adopts when framing issues may result drastically altered. However, such a description of AI-based technologies has been criticized. As Vannuccini and Prytkova (2021) point out, AI-based systems are (still) not as pervasive as many deem. In fact, they found application in a narrow set of industries, such as ICT, professional and scientific services, and finance and insurance. And even in the sectors whereby we observe the largest implementation of AI tools, this technology looks ”superimposed” on existing systems, and the related adoption consists of replacing capital in the execution of certain tasks. Moreover, AI technologies are such a complex phenomenon that describing them as GPT may look belittling, somehow paradoxically. A GPT usually consists of a single upstream source with many edges spreading towards a multitude of downstream sectors, where it contributes to the introduction of complementary innovations. Though it is undisputed that AI induces further downstream innovations, this technology actively participates in the creation of new technologies, *guiding* the role of invention and innovation (Agrawal et al., 2019; Vannuccini and Prytkova, 2021). This role is broader than what usually undertaken by a GPT.
Regardless of the contrasting viewpoints, these authors agree on the importance of huge data availability for the correct functioning of AI systems. As clearly stated by Cockburn et al. (2018, p. 15): ”Because the performance of deep learning algorithms depends critically on the training data that they are created from, it may be possible, in a particular area, for a specific company (either an incumbent or start-up) gain a significant, persistent, innovation advantage through their control over data that is independent of traditional economies of scale or demand-side network effects”. This behavior could result in a balkanization of data within each sector, not only reducing innovative productivity within the sector, but also reducing spillovers back to the deep learning GPT sector, and to other application sectors. Although in our model, data are in principle infinitely reproducible, their price and

availability to agents affect the market structure and suggests that the proactive development of institutions and policies that encourage data sharing and openness is likely to be an important determinant of economic gains from the development and application of deep learning.[3] We complete this ample review of the literature with a discussion on the nature of data that helps us frame, and the reader grasp, the main novelty of our analysis.

3 On the nature of data

We define data as ”a collection of observations of some things” (Chen et al., 2021, p. 2), that can be represented for simplicity as bit strings, e.g., a sequence of zeros and ones. Data might include information about anything: medical records, sensor readings from self-driving cars, consumer tastes, and so on. Chen et al. (2021), Jones and Tonetti (2020), and Savona (2019) attribute to data two general properties: *non-rivalry* and *non-separability*. On the one hand, data are infinitely usable by any number of firms simultaneously; on the other hand, the use of personal data affects the source of data itself. Non-rivalry rises concerns about its ownership since it would be very inefficient to assign exclusive ownership to few agents only. The same set of data may indeed be used by a multitude of agents simultaneously, with no peril of rights infringement between them.[4]

The manifold nature of data can be further highlighted by a concrete example. Duolinguo is a software based on AI mechanisms that allows for improvements in learning and understanding foreign languages. If we were the owners of this software and if we had many customers, then we would notice that some of our clients learn verbs before adverbs, and vice-versa. At a further step, we would be able to combine this knowledge with other information about, for instance, learning timing. This process helps us produce further data based on the initial information. Through these data, we make forecasts and verification, for example on how people behave with respect to nouns and adjectives, *given* the previous timing. The outcome is a further and possibly greater amount of data which belongs to the process of learning and understanding themselves.

This example helps us clarify the following. Firstly, data can be conceived as both the *input* and the *output* of a hypothetical production function based on AI technologies. As inputs, data are capital, labour, or a licensable asset of the firm. Data are capital, because their collection, treatment, and analytics consist of intangible investments that heap on national

[3]Focusing on China, Yu et al. (2021) investigate on how data shape actors relations in data-driven innovation systems. Findings reveal that data are fundamentally different from conventional resources and controlling data impacts upon business value creation, knowledge development and regulation formation.

[4]A corollary regards to the actions data producers could take: they may in principle exploit the information contained in the data so as to negatively affect data subjects' privacy. Concerns about privacy and security-related issues are in Chen et al. (2021, ch. 4).

accounts. Data are labour, for we can treat them as user possession that should benefit its owners (Arrieta-Ibarra et al., 2018). And then, data can be treated as a producers' intellectual property "to the extent that they result from their individual investments in their own personal knowledge" (Savona, 2019, p. 14).
Secondly, data might be an *endogenous* product of Artificial Intelligence, and this feature impacts on their value. We can argue that data do not quickly wear out and have the potential to perform its required function over a lengthy period, if not discarded for any reason. Therefore, they display some features of *durable* goods: the information content does not decay or diminish in value after prolonged use in the production of further data. Furthermore, data have value *if and only if* they are coupled with a technology which draws and manages the information embodied, i.e., only when they are employed in production. Put differently, data have value *as long as* they contribute to generate value. These features may mean that the smallest data might have value today and lose it tomorrow, to acquire again some value afterwards, if not discarded and properly stored.[5]
Thirdly, we could look at data with two different lenses. Data are *generic* when we do not focus on (the quality of) the specific information contained and they are just numbers we pile up in large quantities. For instance, the accumulation of a big bulk of data allows the user to develop and fine-tune any algorithm with which processing data themselves. Such accumulation increases the *efficiency* in production in subsequent periods. On the other hand, data are thought of as *specialized* when we focus on (the quality of) the specific information content they embody. Users pool different groups or sources of data to increase the knowledge in any field of research. Obviously, the increase in knowledge takes the form of further data that serve as input for the "production" of knowledge at subsequent steps of the analysis.

This brief discussion on the nature of data is necessary to understand the way we consider data since the next Section. More precisely, looking at data as both generic and specialized allows to shed light on the research agenda we introduced above.

4 An ABM for AI and data trajectories

We investigate the relationships that join the production and the nature of data, in terms of both input and output of an AI-based production process, with the technological trajectories in AI systems. We set up an economic system in which several agents compete and interact on several markets (Fig. 1). Precisely, we have:

[5]The discussion on the value of data involves increasing returns to scale arguments too: check Posner and Weyl (2019) and Savona (2019) for details. However, there is not consensus on this field: see Chen et al. (2021) and Varian (2019) for differing opinions.

Figure 1. Chart of the model

- A collection N of firms owned by capitalists: they produce data, undertake innovative search and compete in the market for data and AI. They are then data providers, i.e., organisations which own data and offer them to others for a fee (Spiekermann, 2019).
- A public sector, PS, whose activity concerns to the growing and training of scientists, and to the purchase of data.
- A collection S of scientists: they are the labour force which firms hire to build the AI capacity.
- An exogenous entity, D, from the outside which fuels the demand for data.

Our agents can be conceived as *system builders*, i.e., actors which endeavor "to extend the reach of the system and perform the socio-technical integration necessary to its development" (Vannuccini and Prytkova, 2021, p. 15). Each period of the simulation envisages the interaction on three markets:

- The differentiated market, M_{data}, in which firms compete on selling generic or specialized data.
- The labour market, M_{lab}, in which firms contribute to the public training of scientists through sponsorship and hire according to their needs.
- The market for technologies, M_{AI}, in which firms trade their AI blueprint.

A general feature that permeates all the markets is the Schumpeterian competitive environment in which firms get selected by fitness. Selection operates through the introduction of novelties, that might emerge as both process and product innovations, and firms will survive, or not, according to their ability to learn and adopt best productive techniques or organizational practices (Dosi and Nelson, 2010).

The remaining discussion is about what happens on the inside of each market. We detail them with the several actions the agents adopt.

4.1 The market for data, M_{data}

We define a data marketplace "as a digital platform on which data products are traded" (Spiekermann, 2019, p. 210). To describe the first market, M_{data}, we shall look again at Fig. 1 and Tab. 1. The chart shows how M_{data} is linked to three agents: on the supply side the N firms, on the demand side the exogenous entity D, the public sector PS, and the same N firms. The latter can be competitors and customers. Focusing on the production side, this place of interaction constitutes the *differentiated* market for data. Firms might in principle compete on two different *sides*, m: a side in which they produce and sell generic data, m_g, and another side in which they compete for selling specialized data, m_s. Producing generic

Dimension	Data Marketplace	
	m_g	m_s
Market Positioning	Data Providers	Data Providers
Market Access	Open	Open
Data Transformation and Integration	Raw, Generic Data	Quality, Specialized data
Platform Architecture	Decentralised	Decentralised

Table 1. Classification of data marketplace, M_{data}

data consists in providing customers with a large volume of data which, in principle, does not aim at conveying or processing any specific informational content. It is clear that in such a field, firms face what we denote as *price competition*.

On the other side of M_{data}, firms try to compete on the production and sale of specialized data, whose quality helps customers increase their knowledge about any scientific domain. Therefore, competition selects firms on the basis of their ability to provide clients with data of a certain level of quality and performance. Quality here reflects the capability of conveying precise information about a certain field of research. Furthermore, it mirrors not only the goodness of the product sold today but also the potential to engender further knowledge in the future.

Tab. 1 provides some further information. The list of features is based on the simple taxonomy by Spiekermann (2019). For what concerns the market positioning, both sides of the market are operated by data providers, hence the same actors involved in the trading. Secondly, we do not assume any form of entry barrier. Then, data transformation and integration differentiate data in the way they are processed and prepared for sale. Firms trade generic or raw data in m_g, while firms in m_s specialize in trading with particular data, in which an assurance warrants on corresponding performance and quality. Finally, although we assume the existence of a general infrastructure or repository that allows for the storage of data, and we do not deny that firms exchange their datasets, the marketplace architecture is defined as decentralized, because the data content remains in the hands of the suppliers. Such a decentralization hence favors some form of data sovereignty.

4.1.1 Firm's production function

For the sake of simplicity, firms combine several inputs in fixed quantities. Such combination takes the form of a Leontief production function as in Eq. (1):

$$Y_{i,t}^{P,m} = min\left[A_{i,t-1}^{m} \cdot S_{i,t-1}^{m}; B_{i,t-1} \cdot DS_{i,t-1}^{m}; C_{i}^{m} \cdot K_{i,t-1}^{m}\right] \tag{1}$$

in which $Y_{i,t}^{P,m}$ represents the flow of data, for the i^{th} firm at time t, competing on the m side of the market. The flow of data at time t is the result of a production cycle started one period before, at $t-1$. This production is the outcome of the combination of three inputs: first, labour, $S^m_{i,t-1}$ in the form of trained scientists; second, the data stock as result of the accumulation and storage of previously produced data, $DS^m_{i,t-1}$; third, capital, $K^m_{i,t-1}$ that consists of the AI technology, made up of hardware and data-processing algorithms, what Ahmed and Wahed (2020) label as *compute*; $A^m_{i,t-1}$ and $B_{i,t-1}$ represent, respectively, the productivity of labour and data, whereas C^m_i is the inverse capital-output ratio.

The nature of any input requires further clarification. First, although the functional form is similar along m, the production process behind is different according to the output. For instance, the production of a given volume of generic data requires the entrepreneurs to combine the compute $K^{m_g}_{i,t}$ with the accumulated data in quantities different from the one entailed by the production of specialised data. The related capital coefficient, $c^{m_g}_i$, is generally different from $c^{m_s}_i$. The same holds for the other inputs too. Second, for what concerns labour, scientists are heterogeneous in skills and can be divided between *generic* and *specialized* scientists. These are used not only in the production of data, which is a task performed by the AI capital stock, but in the production of the AI capacity and for R&D. Labour productivity is a function of this learning curve:

$$A^m_{i,t} = A^m_{i,t-1} \cdot \left[1 + A_0 \cdot \left(1 - e^{-\lambda \cdot \left[\frac{S_{t-1}}{S_t} \cdot \sum_{z=0}^{t-1} \bar{u}^m_{i,z} + \left(1 - \frac{S_{t-1}}{S_t}\right) \cdot \bar{u}_{i,z}\right]}\right) \cdot \left(\frac{\tilde{A}^m_{i,t}}{A^m_{i,t-1}} - 1\right)\right] \tag{2}$$

in which A_0 is a coefficient, $\tilde{A}_{i,t}$ is the maximum attainable value the scientist may reach when employed in the i^{th} firm, λ is steepness of the curve, and $\bar{u}^m_t$ is the actual utilization of scientists, equal to:

$$\bar{u}^m_{i,t} = \frac{Y^m_{i,t}}{a^m_{i,t} \cdot S^m_{i,t}}$$

The limit posed by $\tilde{A}_{i,t}$ is a function of the innovative search each entrepreneur undertakes to introduce process innovations. Observing the market from the scientist's perspective, we understand that working in different firms means having access to different technologies, different organizational structures, different "ways of doing things", etc. This heterogeneity results in different opportunities to learn. Each scientist, indeed, learns how to master the firm's technologies as long as she keeps working for that firm. When the firm grabs the fruits of its innovative search, the technological vintage shifts upwards. The scientist which is employed in that firm then benefits from an increase in the learning opportunities.

Another reasoning applies to the second input, $DS^m_{i,t}$. The earlier discussion on the nature of data suggests that the i^{th} entrepreneur produces the demanded flow of data at time t with all the bulk of data she has been possessing until $t-1$. The higher this stock of data, the higher the flow of producible new data. Moreover, as for the case for labour productivity, the productivity of the overall stock of data, $B_{i,t}$, is a function of process innovations, e.g., results of the continuous innovative search.

For what regards the nature of capital, $K^m_{i,t}$, we assume that it consists of AI-based technologies. An AI technology is essentially composed of some sibling domains: hardware, software, and algorithms.[6] This complex ensemble allows for, and constrains, the actual production and the processing of data. Moreover, each specific design defines the technological style of the AI technology as it ”emerges from the particular choice and combination of its elements, given their relative importance and the specific role they play in the whole system” (Vannuccini and Prytkova, 2021, p. 15). In Eq. (1), what we define as compute is only circulating capital. We take into account the inherent complementarity with labour and data. AI machines embody the efficiency in the processing of data, $B_{i,t}$, and we assume that each unit of labour must be combined with a proportional unit of AI capital:[7]

$$K^m_{i,t} \propto S^m_{i,t} \tag{3}$$

We can sum up this discussion by answering to this question: is there a structure in the bundle of data the i^{th} firm produces at any period t? If we look at any physical capital, we know there is a difference between a machine cutting metals and a machine working with chemical reagents. In the case of data, we could have a string providing the number of inhabitants in a given city, and another with the size of that city. The reference to a specific city reveals the underlying structure in the data. In this respect, if we consider data as fully substitutable among itself, we admit that combining the number of inhabitants in two different cities provides the same information than combining the number of inhabitants and the size of a unique city, for instance. The assumption of fully substitutability among data looks quite restrictive and *ad hoc*. A possible way to bypass this issue is the production function itself. Although the functional form is the same, the underlying productive processes are different: producing generic data requires a peculiar combination of inputs which differs from the production of specialized data.

4.1.2 Firm's output decisions

Regardless the endowment of data is, the i^{th} firm can produce a flow of new data which is constrained by the technology and labour endowments they possess at a given production cycle. Data is then a flexible input, whose amount changes from time to time, while the

[6] Traub et al. (2019) is an interesting examination of AI components.

[7] We suppose a firm as integrated, such that it builds its own required capital stock. In addition to this, the cost of capital is assumed away from the analysis: it can be either considered as negligible or as embodied in the wage rate, such that entrepreneurs hire workers endowed with physical capital.

capital stock and the manpower are available in strict fixed quantities, at least in the short run.

Firms produce *on demand* and do not store any inventory. They nonetheless form their sales expectations in an adaptive way to smooth short-term volatility. Expectations are essential to endow entrepreneurs with a planned data capacity to fulfil forthcoming demand. The stock of data stored, $DS^m_{i,t}$, follows an adjustment procedure of the form:

$$DS^m_{i,t} = \varepsilon^D \cdot DS^m_{i,t-1} + \left(1 - \varepsilon^D\right) \cdot \left(1 + \eta^m_{i,t}\right) \cdot DS^m_{i,t-1}$$

$$DS^m_{i,t} = \left[1 + \eta^m_{i,t} \cdot \left(1 - \varepsilon^D\right)\right] \cdot DS^m_{i,t-1} \tag{4}$$

in which ε^D is a coefficient while $\eta^m_{i,t}$ expresses entrepreneurial willingness to reach a normal data capacity utilization rate, u_n, as in Eq. (5):

$$\eta^m_{i,t} = \eta_0 \cdot \left(\frac{Y^m_{i,t-1}}{B_{i,t-1} \cdot DS^m_{i,t-1}} - u_n\right) \tag{5}$$

Identifying with $Y^{d,m}_{i,t}$ the demand faced by the i^{th} entrepreneur at time t, the amount of data delivered in the next period is:

$$Y^m_{i,t} = min\left[Y^{d,m}_{i,t-1}; Y^{P,m}_{i,t-1}\right] \tag{6}$$

Eq. (6) leaves room for the possibility that a given firm is not ready or sufficiently equipped for a prompt satisfaction of consumer's demand. In this case, the firm has two options to satisfy consumers. First, the entrepreneur can fulfill orders by adding to its production the stock of data stored. Yet, when even this option is not sufficient, she can only go through the market and buy what becomes second-hand data. Eq. (7) expresses the actual sales, $Y^{a,m}_{i,t}$ as outcome of this reasoning:[8]

$$Y^{a,m}_{i,t} = min\left[Y^{d,m}_{i,t}; Y^{P,m}_{i,t} + \sum_{i=1}^{N} DS^m_{i,t-1}\right] \tag{7}$$

4.1.3 Pricing and competitiveness

The pricing of data can be tricky: even though production requires labour, it is assumed to be used to build the AI capacity. The data production process *per se* does not actually re-

[8]If the economy's potential does not meet aggregate demand, the economy accumulates backlogs that will be added to next period demand. Moreover, at this point of the analysis somebody might envisage a contradiction in our reasoning: on the one hand, we assume no inventories, on the other hand we speak about firm's expectations and the possibility to not satisfy demand. This contradiction is only illusory. The production of data and the subsequent sale does not mean that the seller transfers the good to the buyer and loses it. In fact, the i^{th} firm keeps part of its production on a repository. What is stored there is in fact an inventory, but it does represent an inventory if inputs, not of outputs.

quire labour, since this task merely consists in launching the AI.[9] Besides the cost of labour employed in the development of the AI capital stock, we have to account for the *storage* cost of data.

For the sake of simplicity, the unit storage cost corresponds here to a fixed maintenance of the repository cost. Last but not least, if a firm suffers from a storage shortage, or data production capacity shortage, it then buys data from competitors in order to satisfy its demand. Pricing then depends the amount of data the single entrepreneur takes from her online server and the amount that she has to buy:

$$p^m_{i,t} = \alpha^m_{i,t} \cdot \left(1 + \mu^m_{i,t}\right) \cdot \left(\frac{\gamma}{B_{i,t}} + \frac{w^m_{i,t}}{A^m_{i,t}}\right) + \left(1 - \alpha^m_{i,t}\right) \cdot \left(1 + \mu^m_{i,t}\right) \cdot \bar{p}^m_{t-1} \tag{8}$$

in which $p^m_{i,t}$ is the unit price of output, $\mu^m_{i,t}$ is the mark-up over storage cost γ, unit labour cost $\frac{w^m_{i,t-1}}{A^m_{i,t-1}}$, and $\bar{p}^m_{t-1}$ is last period weighted average market price, equal to:

$$\bar{p}^m_{t-1} = \sum_{i=1}^{N} p^m_{i,t-1} \cdot \sigma^m_{i,t-1}$$

in which $\sigma^m_{i,t-1}$ represents the market share achieved by the entrepreneur in the previous period. The variable $\alpha^m_{i,t}$ allows for a weighted average in the price equation, and it is equal to:

$$\alpha^m_{i,t} = \begin{cases} \frac{Y^{P,z}_{i,t-1}}{Y^{d,z}_{i,t-1}} & \text{if} \quad Y^{P,z}_{i,t-1} < Y^{d,z}_{i,t-1} \\ 1 & \text{otherwise} \end{cases} \tag{9}$$

Entrepreneurs revise mark-ups period by period according to the market share *differential*:

$$\mu^m_{i,t} = max\left(\bar{\mu}; \mu^m_{i,t-1} \cdot \left[1 + v \cdot \left(\frac{\sigma^m_{i,t-1}}{\bar{\sigma}^m_{t-1}} - 1\right)\right]\right) \tag{10}$$

in which $\bar{\sigma}^m_{t-1}$ is the average share on each side of the market, $\bar{\mu}$ is the minimum mark-up and v is a parameter. The average market share can be expressed as the Herfindahl index:

$$\bar{\sigma}^m_{t-1} = \sum_{i=1}^{N} \left(\sigma^m_{i,t-1}\right)^2$$

We capture differences in efficiency between firms by computing some measures of competitiveness, which in turn impacts on the rate of change of the market share:

[9] The process is similar to the Hicksian production function with capital and labour, in which capital does not substitute for labour, but allows the latter to produce. In the present case, we have the opposite mechanism: the i^{th} entrepreneur hires scientists to build enough capacity as to transform data in new data, for labour produces the algorithm and related machines.

$$\Delta\sigma_{i,t}^{m} = \chi \cdot \frac{\left(E_{i,t}^{m} - \bar{E}_{t}^{m}\right)}{\bar{E}_{t}^{m}} \cdot \sigma_{i,t}^{m} \tag{11}$$

in which $\Delta\sigma_{i,t}^{m}$ is the instantaneous rate of change in the market share, expressed as function of the percentage discrepancy between firm's fitness, $E_{i,t}^{m}$, and the average competitiveness on the market, $\bar{E}_{t}^{m}$. The latter is computed as the weighting average between current competitiveness and past market share:

$$\bar{E}_{t}^{m} = \sum_{i=1}^{N} E_{i,t}^{m} \cdot \sigma_{i,t-1}^{m} \tag{12}$$

Since a firm might compete on both sides of the market, its fitness differs accordingly. The sale of generic data implies the ability to meet the demand and to sell production at low prices:

$$E_{i,t}^{m_g} = \frac{\alpha_{i,t}^{e_1,m_g}}{p_{i,t}^{m_g}} \tag{13}$$

in which e_1 is an elasticity. In contrast, on the quality side of the market customers choose between different suppliers by discriminating them on the basis of the quality of data sold too:

$$E_{i,t}^{m_s} = \frac{\alpha_{i,t}^{e_1,m_s} \cdot q_{i,t}^{e_2,m_s}}{p_{i,t}^{m_s}} \tag{14}$$

in which e_2 is again a constant.

4.1.4 Process and product innovations

Firms strive to introduce both process and product innovations. Process innovations result in two possible outcomes, namely increases in the productivity of data and, or, maximum labour productivity levels. The rise in the productivity of data means that one can produce more data from a lower bundle of data than before. This ability emerges through the development of better AI technologies. Improvements are driven by computation per time, greater predictive power, and all the other sorts of efficiency-enhancing solutions that decrease the replication of data or its processing.[10] Think about a self-driving car, whose AI system has to process a lot of data and information before deciding whether to turn "left" or "right". As you fine-tune the algorithm, the AI system might infer "left" or "right" with less information than before. The second process innovation is embodied in the hardware

[10] As a matter of clarification, the constraints on the improvements of the algorithm have a systemic nature (Vannuccini and Prytkova, 2021): when the algorithm and the hardware are complements, the evolution of the former is the result of a strategic choice in which the developers design the feature of superior algorithms on the basis of their current hardware production plans, and vice-versa.

and software parts of the capital stock, and enables scientists to reach a higher maximum productivity level, $\tilde{A}_{i,t}$. This process innovation is not linked to the development of data processing technology, but simply to the computers embodying it, and it is connected to scientists learning processes and opportunities. On the other hand, entrepreneurs competing on the quality side of the market have to explore the innovative possibilities to introduce in the market a bulk of specialised data with higher quality.[11] Firms spend on innovative search a percentage of their past profits, $\pi_{i,t-1}$. We are aware of the empirical regularity that sales affect R&D expenditures more than profits (Ciarli and Lorentz, 2010; Dosi et al., 2010, 2006). However, modelling a credit market is beyond the scope of the paper, and therefore we constrain R&D investments to the available resources, i.e., profits. Eq. (15) helps us clarify how each firm splits its R&D funds, $rd_{i,t}$:

$$rd_{i,t} = \xi_0 \cdot \pi_{i,t-1} = rd^{\vartheta}_{i,t} + rd^{q}_{i,t} \tag{15}$$

in which innovative investments, as percentage ξ_0 of past profits $\pi_{i,t-1}$, are composed of funds dedicated to process, $rd^{\vartheta}_{i,t}$, and product innovations, $rd^{q}_{i,t}$. We devise a procedure through which the single capitalist is willing to invest on R&D according to relative amount of revenues from each (side of the) market. The more successful an entrepreneur is, the more willing to further undertake innovative search on that side of the market. Formally:

$$\varphi^{\vartheta}_{i,t} = \frac{p^{m_g}_{i,t-1} \cdot Y^{m_g}_{i,t-1} + p^{M_{AI}}_{i,t-1} \cdot Y^{M_{AI}}_{i,t-1}}{\sum_m p^{m}_{i,t-1} \cdot Y^{m}_{i,t-1}} \tag{16}$$

Eq. (15) is re-written as:

$$rd_{i,t} = \varphi^{\vartheta}_{i,t} \cdot rd_{i,t} + \left(1 - \varphi^{\vartheta}_{i,t}\right) \cdot rd_{i,t}$$

We model the R&D routine in Nelson and Winter (1982). Firstly, each firm draws an "access to innovation" round from a standard uniform distribution. If this number is contained in the interval $\left[0; P^{\vartheta}_{inn,i,t}\right]$, R&D is successful. We set $P^{\vartheta}_{inn,i,t}$ as:

$$P^{\vartheta}_{inn,i,t} = 1 - \exp\left[-\varepsilon \cdot \frac{S^{m_g}_{rd,i,t-1}}{S^{m_g}_{i,t-1}}\right] \tag{17}$$

in which ε is the effectiveness of R&D, and the following ratio represents the share of scientists performing innovative search for process innovation in the previous period.[12]

[11] The production of new data might hence enlarge the *quality* and the *productivity* of the whole ensemble. There is an inherent difference between the productivity of a bundle of data and its quality. The first is a characteristic that benefits data providers with enhanced efficiency in production through better AI technologies; thus, it is something which stays inside the firm. Conversely, the quality of the data concerns to the benefits in terms of information content the users draw from the single data.

[12] Our take entails the belief that novelties, new ideas, innovation, are brought into the firms by newcomers, i.e., fresh hired scientists.

The innovative search might be successful for both types of process innovation, just one of them, or none. Improvements in process innovations take place through additive schedules. Labour productivity evolves as in Eq. (18)

$$\tilde{A}_{i,t} = \tilde{A}_{i,t-1} + \lambda^{\tilde{A}}_{i,t} \tag{18}$$

The improvement element, $\lambda^{\tilde{A}}_{i,t}$, is equal to the maximum between zero and a random draw from a truncated normal distribution, $\varepsilon^{\tilde{A}}_{i,t}$:

$$\lambda^{\tilde{A}}_{i,t} = max\left(\varepsilon^{\tilde{A}}_{i,t}; 0\right) \tag{19}$$

Similarly for data productivity as in Eq. (20):

$$B_{i,t} = B_{i,t-1} + \lambda^{B}_{i,t} \tag{20}$$

in which $\lambda^{B}_{i,t}$ is again determined as the maximum between zero and a normal random draw, $\varepsilon^{B}_{i,t}$:[13]

$$\lambda^{B}_{i,t} = max\left(\varepsilon^{B}_{i,t}; 0\right) \tag{21}$$

Product innovations occur with identical mechanisms. The access to innovation depends on a random draw from a standard uniform distribution as above. If this draw is contained in the interval $\left[0; P^{q}_{inn,i,t}\right]$, the firm may improve the quality of its data. We indicate such quality with $q_{i,t}$, and this variable evolves according to the usual rule:

$$q_{i,t} = q_{i,t-1} + \lambda^{q}_{i,t} \tag{22}$$

in which, again, $\lambda^{q}_{i,t}$ is a random variable drawn from a truncated normal distribution:[14]

$$\lambda^{q}_{i,t} = max\left(\varepsilon^{q}_{i,t}; 0\right) \tag{23}$$

4.2 The labour market, M_{lab}

The second place of interaction is the market for scientists, M_{lab}. Such market involves three types of agents. On the supply side, we have scientists, S, and the public sector, PS, whose scope consists of growing and training the labour force. On the demand side instead, there are the N enterprises, which now compete to hire best productive workers. Scientists are

[13] Eq. (17) highlights a further mechanism: entrepreneurs hire workers to build their own AI means of production and these workers build the same vintage of AI through time, unless newcomers arrive at the firm, possibly with better organizational *ideas* on how to improve technical vintage and labour productivity. Therefore when technical change unfolds, the elder generation of employees have to learn and adapt to new vintages.

[14] For the sake of simplicity, achievements in quality are commonly shared among stored data.

heterogeneous in skills and are divided between generic and specialized scientists used on both their respective production and R&D. Given the previous demand $Y^{d,m}_{i,t-1}$, the labour productivity $A^{m}_{i,t-1}$, the amount of R&D funds and the wage rate, we define the desired labour demand $S^{m,\iota}_{i,t}$ as:

$$S^{m,\iota}_{i,t} = \varepsilon^{S} \cdot S^{m,\iota}_{i,t-1} + \left(1 - \varepsilon^{S}\right) \cdot \left(1 + u^{S}\right) \cdot \frac{Y^{d,m}_{i,t-1}}{A^{m}_{i,t-1}} + \frac{rd^{m}_{i,t}}{w^{m}_{i,t}} \tag{24}$$

in which ε^{S} and u^{s} are parameters. We re-arrange this expression from Ciarli et al. (2010) and to explain it, we shall think about a firm that faced a peak in demand at $t-1$, such that its capacity was not sufficient to satisfy it. At period t, the same firm adjusts labour demand so as to update production capacity in terms of scientists. Moreover, this adjustment helps the firm deal with future peaks with a buffer of workers proportional to u^{S}. Entrepreneurs try to hire more than what needed. At the same time, when capacity is higher than demand, the firm start firing workers. Nevertheless, in this case, entrepreneurs fire less than what needed, so as to prevent that capacity falls short of demand.[15] Once the entrepreneur sets the desired labour demand, she opens an amount of new vacancies equal to:

$$S^{m}_{V,i,t} = max\left[0; \left(\varepsilon^{S} - 1\right) \cdot S^{m,\iota}_{i,t-1} + \left(1 - \varepsilon^{S}\right) \cdot \left(1 + u^{S}\right) \cdot \frac{Y^{d,m}_{i,t-1}}{A^{m}_{i,t-1}} + \frac{rd^{m}_{i,t}}{w^{m}_{i,t}}\right] \tag{25}$$

Opening vacancies does not mean filling them. Scientists have a structure of preferences about the workplace. Following Almudi et al. (2012), they choose which firm to join according to monetary as well as non-monetary criteria. The monetary determinant is based on the wage rate. Non-monetary factors reflect scientist's expectations of future gains in learning. Formally:

$$\psi^{m}_{i,t} = \psi_0 \cdot w^{m}_{i,t} + (1 - \psi_0) \cdot \tilde{A}_{i,t} \tag{26}$$

in which $\psi^{m}_{i,t}$ represents firm's attractiveness, computed as a weighted average between monetary and non-monetary factors, while ψ_0 is a coefficient. In a situation in which the aggregate supply of scientists is not sufficient to satisfy the aggregate demand for labour, not every firm covers its open vacancies. In this case, scientists will be allocated according to firm's attractiveness. The amount of filled vacancies is equal to:

$$S^{m,f}_{d,i,t} = min\left[S^{m}_{V,i,t}; \psi^{m,r}_{i,t} \cdot LS^{m}_{t}\right] \tag{27}$$

in which LS^{m}_{t} is the labour supply in terms of scientists, and $\psi^{m,r}_{i,t}$ is the relative attractiveness, computed as:

[15] In some extent Eq. (24) presents scientists as funds *à la* Georgescu-Roegen, for the hiring of scientists consists of investments a firm does to ameliorate its capabilities and gain further market power.

$$\psi_{i,t}^{m,r} = \frac{\psi_{i,t}^{m}}{\sum_{i=1}^{N} \psi_{i,t}^{m}}$$

Firm workforce is then:

$$S_{i,t}^{m} = S_{i,t-1}^{m} + S_{d,i,t}^{m,f} \tag{28}$$

Less-competitive entrepreneurs have thus two options to fill their vacancies, and both impact upon firm's attractiveness. First, they could increase R&D funds. Secondly, they could increase the wage rate in proportion to the level of unfilled vacancies. The wage rate is computed as a function of actual productivity growth and the percentage amount of unfilled vacancies:[16]

$$w_{i,t}^{m} = w_{i,t-1}^{m} \cdot \left(1 + \chi_1 \cdot \frac{\Delta A_{i,t-1}^{m}}{A_{i,t-1}^{m}} + \chi_2 \frac{S_{V,i,t-1}^{m} - S_{d,i,t-1}^{m,f}}{S_{V,i,t-1}^{m}}\right) \tag{29}$$

in which χ_1 and χ_2 are parameters.

The supply of scientists depends on the amount of funds the economic system spends for their training, which is funded partly by firms through sponsorship, and partly by the public sector. For what concerns to firms, price-competitive firms are more interested in training generic scientists, while quality-competitive firms prefer training and hiring specialized scientists. Each enterprise, competing in one or both sides of M_{data}, spends a percentage from past profits:

$$T_{i,t}^{m} = \xi_1 \cdot \pi_{i,t-1} \tag{30}$$

in which $T_{i,t}^{m}$ are sponsorship and ξ_1 is a constant. The public sector may adopt pro-firms and anti-firms policies. If there is a strong imbalance in the relative supply of scientists, e.g., a higher amount of generic scientists than specialized, or vice-versa, the public sector may fuel, or counteract, firms propensity to invest in generic scientists. For the sake of simplicity, the government draws funds from outside the system and trains pools of scientists as in Eq. (31):

$$rd_{PS,t} = rd_{PS,t}^{m_g} + rd_{PS,t}^{m_s} = \beta_t \cdot rd_{PS,t} + (1 - \beta_t) \cdot rd_{PS,t} \tag{31}$$

in which $rd_{PS,t}$ is about public funds split between generic and specialized scientists; β_t is a variable that reflects public sector's willingness to accommodate or counteract firms' aims and is expressed in the following way:

[16] It is worth noting that the wage rate may differ between types of scientists within the same firm.

$$\beta_t = \beta_0 \cdot \frac{\sum_{i=1}^{N} T_{i,t-1}^{m_g}}{\sum_{i=1}^{N} \sum_m T_{i,t-1}^{m}} + (1 - \beta_0) \cdot \left(1 - \frac{\sum_{i=1}^{N} T_{i,t-1}^{m_g}}{\sum_{i=1}^{N} \sum_m T_{i,t-1}^{m}}\right) \tag{32}$$

with $\beta_0 = 1$ when the public sector supports firms desires, and $\beta_0 = 0$ in the other case. The supply of scientists is a function of the investment in their training:

$$\Delta LS_t^m = \left(\sum_{i=1}^{N} T_{i,t-1}^{m} + rd_{PS,t-1}^{m}\right)^k \tag{33}$$

with k a parameter. The available supply of scientists at market level, LS_t^m, is the sum of three components: the available workers at time $t-1$, the newcomers ΔLS_t^m, and the difference between fired and hired scientists, ΔS_t^m:

$$LS_t^m = LS_{t-1}^m + \Delta LS_t^m + \Delta S_t^m \tag{34}$$

4.3 The market for AI technologies, M_{AI}

This subsection deals with the transmission mechanism of technological innovations. In particular, it is about the second-hand market for AI. We observe the presence of three main actors: the N firms on the supply-side whereas the exogenous entity D, the public sector and the same N firms are on the demand side. Surviving or successful corps might indeed find profitable to trade the blueprint of their AI to potential customers, that we identified with M_{data} competitors or agents from the outside. We stress two crucial points that distinguish M_{AI} from M_{data}. Firstly, fine-tuning the technological apparatus does not result in process innovations any more. M_{AI} looks at it as a *product* innovation. Secondly, what they actually sell on this market is not (only) their AI technology but the *production function* itself: selling the blueprint of any AI-based technology results in the sale of all the knowledge available to the firm about the creation, implementation, and mastering of AI, and how to combine it with accumulated data as to produce new output. Looking at the way we defined the production function in Eq. (1), AI embodies what the literature often denotes with "social technologies". This expression stands for how "knowledgeable people act and interact where the effective coordination of interaction is key to accomplishment" (Nelson and Sampat, 2001, p. 40). Technologies seldom involves simple activities undertaken by isolated individuals. They rather nest intrinsic elements in organizational structures that "capture the system of norms, beliefs, and social practises shaping the "ways of doing things" " (Dosi and Nelson, 2010, p. 61).[17]

[17] Obviously, different (social) technologies differ in the way they set up the division of labour and the coordination of the many tasks. These differences stand both between and within the borders of any firm. Moreover, such differences may prove to be more, or less, efficient as circumstances change, reflecting variations in opportunities and contexts (Nelson and Sampat, 2001).

In this picture, trading AI means selling own internal *secrets* on production and efficiency techniques. Firms fuel competitors' imitative strategy. To partially avoid this risk, firms do not trade their latest equipment, but an old vintage of it. If the i^{th} firm succeeds in improving its AI at time t, then it will trade on the market the technology available at the previous period. The sale follows a standard mark-up rule applied to the cost of maintenance of the technology, set equal to γ:

$$p_{i,t}^{M_{AI}} = \left(1 + \mu_{i,t}^{M_{AI}}\right) \cdot \gamma \tag{35}$$

in which $p_{i,t}^{M_{AI}}$ is the unit price and $\mu_{i,t}^{M_{AI}}$ the corresponding mark-up, whose functional form mimics Eq. (10).

Differently from the data market in which we employed the replicator dynamics for the market shares, here we assume and devise a raw matching mechanism that allows for direct interaction between firms (Ciarli et al., 2010).[18] In particular, firms search on this market for AI technologies that might help them enhance their competences and organizational capabilities. Searching on the market implies that each producer has a certain probability to be selected as a supplier. Once every customer has selected a supplier and decided to buy or not, demand from firms is simply aggregated.

For the sake of simplicity, we assume the probability of being selected as a simple function of $\tilde{A}$, B, and $p^{M_{AI}}$:

$$\Lambda_{i,t} = f\left(\tilde{A}_{i,t-1}; B_{i,t-1}; p_{i,t}^{M_{AI}}\right)$$

with first derivatives $f'_{\tilde{A}} > 0$, $f'_B > 0$, and $f'_p < 0$. The functional form might expressed with the aid of an index $I_{i,t}$:

$$I_{i,t} = \left(\frac{p_{i,t-1}^{M_{AI}}}{1 + \bar{p}_{t-1}^{M_{AI}}}\right)^{\iota_p} \cdot \left(\frac{\tilde{A}_{i,t-1}}{1 + \bar{A}_{t-1}}\right)^{\iota_{\tilde{A}}} \cdot \left(\frac{B_{i,t-1}}{1 + \bar{B}_{t-1}}\right)^{\iota_B} \tag{36}$$

in which ι_p, $\iota_{\tilde{A}}$, and ι_B are preferences over capital price, labour and data productivity, respectively. Hence, the selection probability is:

$$\Lambda_{i,t} = \frac{I_{i,t}}{\sum_{i=1}^{N} I_{i,t}} \tag{37}$$

4.4 Demand

Modelling demand is simple. We have an exogenous entity, D, and the government, PS, whose interest consists of purchasing data and AI-based systems. The exogenous entity can be viewed as grouping several industries that use data to get information on varied

[18] To clarify, the demand from D and PS is allocated through a replicator mechanism as usual.

subjects. We could have pharmaceutical and health organisations, which are interested in both data and AI technology. This industry requires better performing techniques to process medical records, and visual recognition systems for a precise detection of human cancers. We have firms building autonomous vehicles, which need an ample collection of sensor readings and actions taken by expert drivers, so as to develop security control systems; in addition to this, such producers might need highly-performing algorithms to analyse data on car's energy consumption and to devise more efficient batteries. Still, we could simply have data on customers' preferences: sellers will be hungry to know these preferences to tailor their products accordingly. The related matching between supply and demand would be further enhanced.

The demand for data and AI capital from the exogenous entity and the government grows exogenously. However, firm's demand for AI capital requires some clarification. The problem here is that a firm either buys the technology or not. The purchase of a technology requires funds that are subtracted from profits, and hence from in-house innovative search. Following Eq. (15) and Eq. (30), we suppose that the resources not invested in internal R&D or in sponsorship heap on a reserve for the purchase of AI capital from competitors:

$$K_{i,t}^{M_{AI},R} = (1 - \xi_0 - \xi_1) \cdot \pi_{i,t-1} \tag{38}$$

in which $K_{i,t}^{M_{AI},R}$ is the reserve for the technology. The demand schedule, $K_{i,t}^{l,d}$ is therefore:

$$K_{i,t}^{M_{AI},d} = \begin{cases} 1 \quad \text{if} & \begin{cases} K_{i,t}^{M_{AI},R} \geq p_{-i,t}^{M_{AI}} \\ \tilde{A}_{i,t} \leq \tilde{A}_{-i,t}; B_{i,t} \leq B_{-i,t} \end{cases} \\ 0 \quad \text{otherwise} \end{cases} \tag{39}$$

Eq. (39) means that you may choose to buy an AI technology on M_{AI} if you can pay for it, or when the benefits in terms of productivity are greater than in-house technology's.

With respect to each market, the aggregate demand is then composed of three elements as in Eq. (40): demand from the public sector, $Y_{d,t}^{m,PS}$; demand from the exogenous entity, $Y_{d,t}^{m,D}$; and backlogs accumulated through time, BL_{t-1}^{m}, if any:

$$AD_t^m = Y_{d,t}^{m,PS} + Y_{d,t}^{m,D} + BL_{t-1}^{m} \tag{40}$$

Eventually, we compute the amount of revenues and profits from each firm's performance on three competitive markets:

$$\pi_{i,t} = \sum_m p^m_{i,t} \cdot Y^{a,m}_{i,t} + \sum_{i=1}^{N-1} p^{M_{AI}}_{i,t} + \sum_m \bar{p}^m_{t-1} \cdot min\left[DS^m_{i,t-1}; \sigma^m_{i,t} \cdot \sum_{j=1}^{N-1} id^m_{j,t}\right] + \\ -\gamma\left(1 + \sum_m DS^m_{i,t}\right) - \sum_m w^m_{i,t-1} \cdot S^m_{i,t-1} + \\ -p^{M_{AI}}_{-i,t} \cdot K^{d,M_{AI}}_{i,t} - \sum_{j=1}^{N-1} \bar{p}^m_{t-1} \cdot min\left[id^m_{i,t}; \sum_{j=1}^{N-1} DS^m_{j,t-1}\right] \tag{41}$$

Firms benefit from three different sources of revenues: those from the sale of data to customers, $\sum_m p^m_{i,t} \cdot Y^{a,m}_{i,t}$; those from the sale of AI blueprint, $\sum_{i=1}^{N-1} p^{M_{AI}}_{i,t}$; and finally, revenues from the sale of data to other firms, $\sum_m \bar{p}^m_{t-1} \cdot min\left[DS^m_{i,t-1}; \sigma^m_{i,t} \cdot \sum_{j=1}^{N-1} id^m_{j,t}\right]$, which are equal to the average price times the minimum between firm's own data stock and what required by all the others, $\sum_{j=1}^{N-1} id^m_{j,t}$.[19] On the cost side, we have the total expenditure to store accumulated data and technology, $\gamma\left(1 + \sum_m DS^m_{i,t}\right)$; the labour cost, $\sum_m w_{i,t-1} \cdot S^m_{i,t-1}$; the cost borne for the purchase of AI systems, $p^{M_{AI}}_{-i,t} \cdot K^{d,M_{AI}}_{i,t}$; and eventually, the cost of acquiring data from others' stock on the basis of what is actually available in the industry, $\sum_{j=1}^{N-1} \bar{p}^m_{t-1} \cdot min\left[id^m_{i,t}; \sum_{j=1}^{N-1} DS^m_{j,t-1}\right]$.

Before analysing simulation results, we shall clarify that once you developed the AI capital in the form of algorithm, software, etc., what you sell on M_{AI} as a product innovation leads to extra profits. For the R&D cost is already paid, each unit sold on the second-hand market will be pure profits. In addition to this, we envisage an extreme form of increasing returns to scale: you might sell your own data processing technology to just one client or to many of them and the production cost is the same. This technology is then infinitely reproducible at no cost after its development. Every single unit (i.e. the blueprint of the AI) sold on M_{AI} is a *free lunch*.

5 Simulation results

We undergo the model through computer simulations. We first run a benchmark scenario in which 50 firms compete during 5000 period simulations along 50 Monte Carlo runs. Tab. 2 gathers baseline parameter values. We pay particular attention to the growth rate of government expenditure and to wage sensitivity to unfilled vacancies. We select very low values not to overheat the system. For instance, if we considered a single time step as roughly corresponding a quarter, a growth rate in public spending equal to 0.5% coincides with a 2% growth each year. All the other parameters are selected in accordance to the literature on agent-based models (Gatti et al., 2011). Additionally, we set initial conditions such that

[19]For the sake of clarity, we define $id^m_{i,t} = max\left[0; Y^{d,m}_{i,t} - Y^{P,m}_{i,t}\right]$.

Parameter	Description	Value
T	Time	5000
MC	Monte Carlo runs	50
N	Number of firms	50
β_0	Government accommodation	1
γ	Unit storage cost	1
ε	Coefficient in the probability to innovate	1
ε^D	Data stock adjustment coefficient	0.5
ε^S	Labour market friction	0.5
η_0	Coefficient in normal data utilisation rate	0.5
$\iota_{\bar{A}}$	Preference over labour productivity	1
ι_B	Preference over data productivity	1
ι_p	Preference over price	−0.1
$\bar{\mu}$	Minimum mark-up	0.09
ξ_0	Share of profits re-invested in R&D	0.5
ξ_1	Share of profits re-invested in sponsorship	0.01
χ	Market share sensitivity to competitiveness	0.1
χ_1	Wage sensitivity to productivity growth	0.5
χ_2	Wage sensitivity to unfilled vacancies	0.01
ψ_0	Attractiveness sensitivity to wages	0.5
A_0	Coefficient in learning function	0.1
e_1	Fitness elasticity to unfilled demand	0.5
e_2	Fitness elasticity to quality	1
g^D	Exogenous growth in demand	0.005
g_{PS}	Government demand growth rate	0.005
k	Labour supply elasticity to public and private sponsorships	1
u_n	Normal data utilisation rate	0.75
u_s	Unused labour capacity	0.15
v	Mark-up sensitivity to market share	0.01

Table 2. Parameter setting

firms are perfectly homogeneous: the heterogeneity will emerge when the model unfolds as outcome of interactions and different decision rules.

5.1 Baseline scenario

We analyse the baseline model by presenting results from both Monte Carlo averages and single simulations. Monte Carlo averages unfortunately hide interesting phenomena which are only detectable through a closer inspection of single replications. For the sake of clarity, we present and analyze the emergent properties on industrial dynamics first, and secondly the dynamics and trajectories related to technical change.

5.1.1 Competition and industrial dynamics

Fig. 2 and Fig. 3 provide information on the market structure and its dynamics with respect to the three markets of interest. We observe indeed a different and peculiar pattern for each market. The market for generic data can be conceived as a competitive environment which evolves toward an oligopolistic structure. The first 2000 time periods exhibit an aggressive and fierce competition between firms, from which a handful of enterprises emerges: the average market share is in between 0.2 and 0.3 across Monte Carlo runs. Firms get selected through their ability to produce and sell large volumes of data at low prices, regardless of quality. Distinguishing factors are then labour productivity and data productivity, which

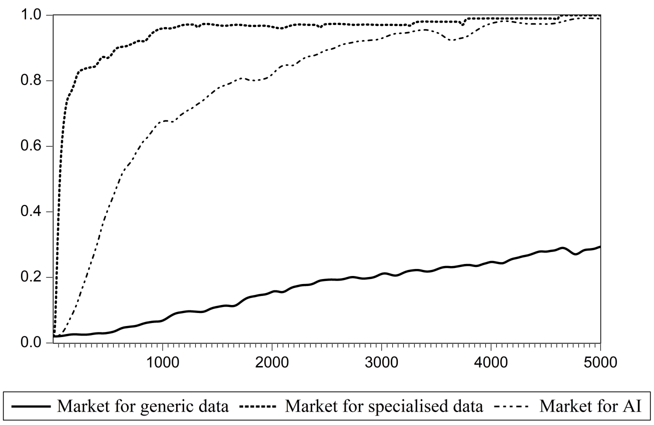


Figure 2. Baseline setting: market shares

result from process innovations.[20] In contrast, the market for specialised data clearly shows the usual pattern out of a replicator dynamics, with a strong tendency to a monopolistic structure once firms differentiate in their technologies (i.e., respective productivity) and quality. As suggested also in Fig. 3 (Panels B), the firm claiming all the market sells data of the highest quality and it is basically the only one which keeps on introducing product innovations in this market.

The remaining market, the one in which firms trade their AI systems and their *production function*, displays a deceiving average dynamics. If we looked at Fig. 2 only, we would see a tendency to monopoly, though not as fast as in the previous case. Yet, a closer inspection of single simulations reveals a leap-frogging structure (Fig. 3, Panel C). The dynamics depicts several waves of monopoly in which some firm becomes a temporary leader of the market.

We present additional statistics on industrial dynamics in Fig. 4. Panel A presents prices and mark-ups. The former are more concentrated around the average than the latter. For what concerns to the markets for generic data and AI technologies, coefficients of variation in prices converge to 0.2 and zero, respectively. These patterns are indicative to the market structure similar to (oligopolistic) competition in the market for generic data and to a leap-frogging dynamics in the AI market. The same applies with respect to the variation in mark-ups. For both markets, coefficients increase at the beginning then exhibit a decrease and finally converge to 0.5. A slightly different reasoning applies to the market for specialised data. Its monopolistic structure envisages a weak, but increasing, variation in

[20] Looking at single simulations, we found that often a batch of firms emerged as benefiting from higher productivity standards. However, benefits from increased productivity in terms of market shares led to higher mark-ups and higher wage rates. These matters counterbalance the gain in competitiveness originating from productivity standards greater than average, on the one hand, and allow for a reallocation of market demand to other firms. Competition is then restored. Related figures are available upon request.

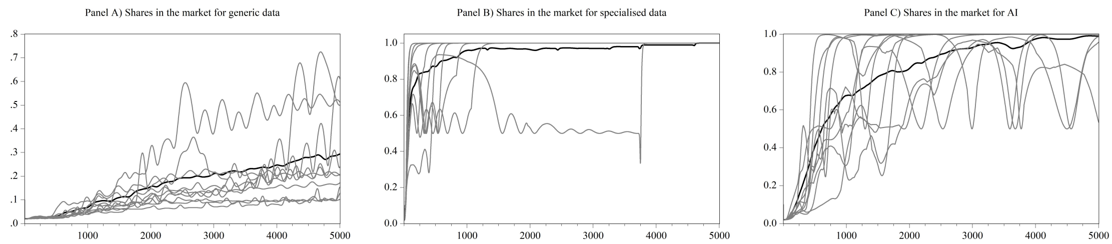


Figure 3. Baseline setting: single Monte Carlo runs

Note: gray lines represent Monte Carlo replications for the average variable; black lines are the average time series across Monte Carlo runs.

prices due to the premium for quality, while the mark-up's coefficient of variation converges to 0.8. This in turn bring about two remarks. Firstly, the growing difference in quality standards between the monopolist and all the others explains the magnitude of the coefficient. Secondly, once the best-quality firm becomes the established monopolist, it stops raising the mark-up limiting thereof the variation.

Panel B depicts the times series of inverse Herfindahl indexes for profits, R&D, revenues, and aggregate demand. In agreement with the above, we find that the index for revenues converges to unity. It means that, on average, firms get their largest source of revenues out of only one single market, out of three. Given the market structures, all but two firms take their revenues from the generic-data market. The other two are leaders, ultimate or not, in the other markets. This evidence clarifies why the inverse Herfindahl index for profits and R&D is below 1 and progressively declines, on average. Being competitive in only one market is often not sufficient to cover the losses experienced elsewhere, and to allow at the same time to undertake innovative search. In contrast, the index on aggregate demand is always close to 3. Since the growth rate of each demand component is the same, differences in monetary terms are due to prices only.

Finally, last panel in Fig. 4 suggests that labour shortage might characterise the market for specialised data, while being not a problem at all in the market for generic data. The benchmark setting, indeed, assumes the public sector as an accommodating agent and, since in the market for generic data many firms are competitive, they are able to fund, on aggregate, an absolute amount of sponsorship larger than what spent by the monopolist in the market for specialised data.

5.1.2 Technological change and technological trajectories

Since Nelson and Winter (1982), scholars have framed industrial dynamics in terms of Schumpeterian technological regimes. We can review this broad literature with the results

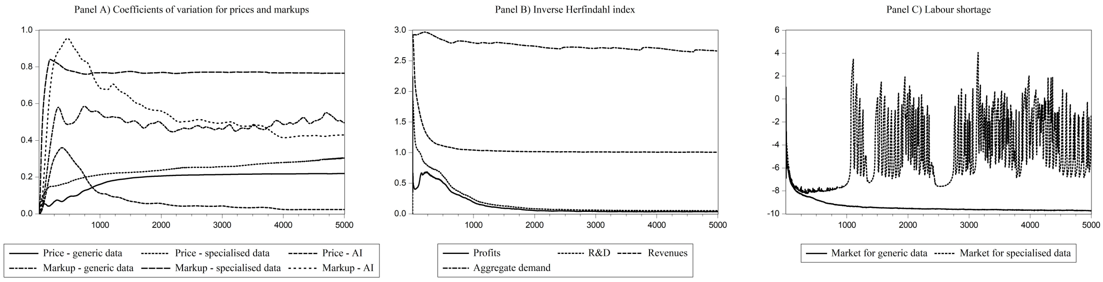


Figure 4. Baseline setting: further indicators

obtained by Malerba and Orsenigo (1995, 1996b, 2002) in their several studies on the topic. With reference to any industry or market, they suggest the existence of technological imperatives and technological regimes in terms of opportunity, appropriability, comulativeness and knowledge-based features. These industry-specific technology-related factors play a crucial role in determining the particular market structure that characterizes that very industry. For instance, the rise of a monopolistic structure in the market for specialised data would be the result of specific technological imperatives and mechanisms that differ from the technological regimes working on the market for generic data, since the competitive framework there in place. However, this line of reasoning is not fully satisfactory to understand the mechanisms at work in our model. To begin with, Fig. 5 and Fig. 6 represent the patterns of labour productivity, data productivity and data quality. Fig. 5 (Panel A) shows a similar increasing pattern for labour and data productivity, on average. Conversely, the upward trend followed by average quality is of a greater magnitude than the former. This feature recalls the old Schumpeterian argument: it is plausible to believe that the competitive environment in the market for generic data does not provide enough funds to firms for innovative search and the introduction of process innovations; the monopolistic setting in the market for specialised data allows, by contrast, for an active innovative undertake.

Panel B in Fig. 5 is about the coefficients of variation in productivity and quality. For what concerns to data and labour productivity, coefficients converge to a value close to 0.3, indicating a little variation around the average. Then, firms are not very different in their productivity levels. In contrast, the coefficient of variation is increasing towards a value around 2 when we consider quality. The monopolistic structure in the market for specialised data envisages a situation in which the monopolist keeps increasing the quality of its data through product innovations, while all the other firms do not innovate at all and their quality standards are low and not much greater than what set at $t = 0$. Therefore, the deviation around the (low) simple mean is very large and explains why the coefficient of variation is above 1.

We introduce a specific statistics, ABQ, that measures how different is the growth of three technological components: labour productivity, data productivity and quality. Denoting with $\sigma^2_{\tilde{A}}$, $\sigma^2_{\tilde{B}}$, and $\sigma^2_{\tilde{q}}$ the variances of the corresponding growth rates from a common average, the index is computed as:

$$ABQ = \sqrt{\frac{\sigma^2_{\tilde{A}} + \sigma^2_{\tilde{B}} + \sigma^2_{\tilde{q}}}{3}} \tag{42}$$

The pattern in Fig. 5 (Panel B) gravitates around zero, and this behaviour means that the economy does not focus on particular differences in the technological trajectories followed by labour productivity, data productivity or quality. All of them grow roughly at the same average rate. The dynamics in terms of productivity and quality, on the one hand, and the market structures, on the other hand, question the theory put forth by Malerba and Orsenigo (1996b). If we considered the market for generic data and the market for AI systems, and if we accepted their argument, we should believe that differences in the market structure are attributable to different technological regimes. But this is not the case: both markets share the same technology-related factors since firms get selected by corresponding labour and data productivity, i.e., by their ability to introduce process innovations. At the same time, we cannot believe that technological imperatives are enough to experience divergent industrial dynamics on the two submarkets for data. Both m_g and m_s deal with the same kind of product, and both quality and productivity grow at the same average rate, as revealed by ABQ. Although we need technical change for these diverse industrial structures to emerge, differences in the characteristics of these technological patterns, or the absence thereof, does not appear to be their main driving force. Though we need productivity gains to produce any industrial dynamics, its actual unfolding is not the outcome of a specific and peculiar technological trajectory, but it results from the interplay of the very technological factors with the availability of inputs in the form of labour and AI, on the one hand, and demand factors such as the market size, consumers preferences, and the government expenditures on the other.

To support this claim, we present in Tab. 3 to Tab. 5 the results of a battery of experiments in which we stress the influence of the sole technological factors on the overall behaviour of the model. The parameter ε is the exponent in the probability to innovate that we have already described in Eq. (17); Θ_j are instead the standard deviations of the normal distribution that governs $\lambda^{\tilde{A}}$, λ^{B} and λ^{q}. While the first parameter appears to control the frequency of appearance of technological changes, increasing the probability to innovate, the second, in turn, influence the amplitude of technological changes, when they occur.

Concerning the changes in ε, the higher its value, the higher the average levels of productivity (for both data and labour) and quality. Yet, this parameter does not alter the overall industrial structure: the market for generic data is always highly competitive, while the

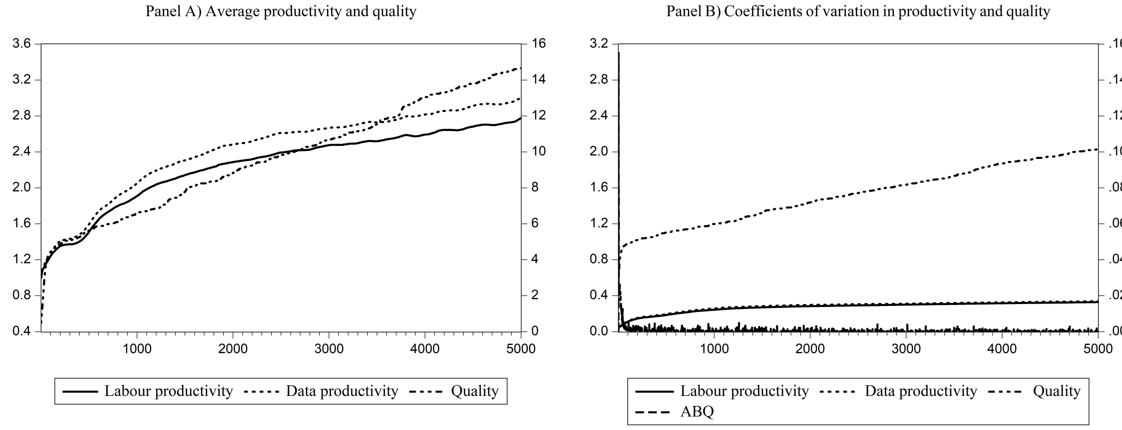


Figure 5. Baseline setting: technological indicators

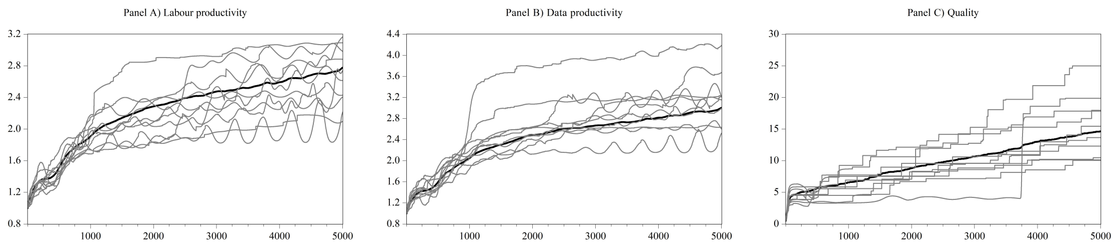


Figure 6. Baseline setting: patterns in productivity and quality

Note: gray lines represent Monte Carlo replications for the average variable; black lines are the average time series across Monte Carlo runs.

Index	$\varepsilon = 0$	$\varepsilon = 0.3$	$\varepsilon = 0.5$	$\varepsilon = 0.7$	$\varepsilon = 1$
$\bar{A}$	**1.000**	**1.488**	**1.778**	**1.953**	2.374
$\bar{B}$	**1.000**	**1.581**	**1.867**	**2.055**	2.574
$\bar{q}$	**0.500**	**3.727**	**5.638**	**6.858**	9.487
ABQ	**0.000**	**4.4X**10^{-4}	**4.7X**10^{-4}	**5X**10^{-4}	5.7X10^{-4}
$\bar{\sigma}^{m_g}$	**0.020**	**0.066**	**0.103**	**0.128**	0.178
$\bar{\sigma}^{m_s}$	**0.020**	**0.874**	0.935	0.949	0.938
$\bar{\sigma}^{m_{AI}}$	**0.020**	**0.686**	0.777	0.766	0.765
$\bar{\pi}_{IH}$	**0.584**	**0.235**	**0.178**	**0.153**	0.133
$\bar{rd}_{IH}$	**1.881**	**0.317**	**0.237**	**0.211**	0.186
$\bar{Y}^a_{IH}$	**2.908**	**1.106**	**1.074**	**1.069**	1.065
AD_{IH}	**2.909**	**2.757**	2.732	**2.767**	2.666
$lab^{m_g}_{short}$	**1.1X**10^{-3}	**1X**10^{-3}	**1X**10^{-3}	9.9X10^{-4}	9.9X10^{-4}
$lab^{m_s}_{short}$	**5.6X**10^{-4}	0.463	0.473	0.291	0.572
A_{cv}	**0.000**	**0.164**	**0.212**	**0.237**	0.297
B_{cv}	**0.000**	**0.176**	**0.222**	**0.253**	0.316
q_{cv}	**0.000**	**0.827**	**1.121**	**1.280**	1.518
$p^{m_g}_{cv}$	**0.000**	**0.097**	**0.134**	**0.158**	0.200
$p^{m_s}_{cv}$	**0.000**	**0.201**	0.227	0.227	0.276
$p^{m_{AI}}_{cv}$	**0.000**	**0.122**	0.092	0.079	0.072
$\mu^{m_g}_{cv}$	**0.000**	**0.450**	0.465	0.459	0.478
$\mu^{m_s}_{cv}$	**0.000**	0.749	0.768	0.777	0.757
$\mu^{m_{AI}}_{cv}$	**0.000**	**0.654**	0.580	0.527	0.543

Note: baseline scenario in red; black numbers are statistically different at 5% from the corresponding baseline.

Table 3. Experiment on ε

market for specialised data reaches a monopolistic setting that switches to leap-frogging when we consider the market for AI systems. Obviously, when ε equals zero, the model does not exhibit any dynamics. In doing so, we basically rule out the possibility to have any innovation in the system.

Similar patterns are confirmed when looking at the indicators after a change in Θ_j. Increasing its value raises labour and data productivity on average, and since the average market share sways around 0.4, a form of oligopolistic competition survives in the market for generic data. Finally, it is important to note that our measure ABQ remains very low, suggesting that, regardless of the value assumed by Θ_j, quality, labour and data productivity keep on growing at the same average rate.

Once we ascertained that strictly technology-related factors are combined with different industrial dynamics, the following sets of experiments try to single out further determinants that might lead to each market-specific patterns. These experiments concern to the elasticity of competitiveness with respect to the demand for and the quality of data, on the one hand, and to the role of the public sector and the elasticity of labour supply, on the other hand.

Index	$\Theta_{A,B}=0$	$\Theta_{A,B}=0.1$	$\Theta_{A,B}=0.5$	$\Theta_{A,B}=0.7$	$\Theta_{A,B}=1$
$\bar{A}$	**1.000**	2.182	**6.774**	**9.851**	**13.028**
$\bar{B}$	**1.000**	2.430	**7.300**	**9.765**	**14.758**
$\bar{q}$	9.931	9.801	**8.460**	**8.718**	**7.862**
ABQ	**3.6X**10^{-4}	5.5X10^{-4}	**7X**10^{-4}	**6.8X**10^{-4}	**6.8X**10^{-4}
$\bar{\sigma}^{m_g}$	**0.020**	0.171	**0.423**	**0.453**	**0.395**
$\bar{\sigma}^{m_s}$	0.942	0.949	0.951	0.953	**0.973**
$\bar{\sigma}^{m_{AI}}$	**0.020**	0.783	**0.924**	**0.919**	**0.918**
$\bar{\pi}_{IH}$	**0.419**	0.131	**0.056**	**0.053**	**0.057**
$\bar{rd}_{IH}$	**1.009**	0.187	**0.094**	**0.091**	**0.096**
$\bar{Y}^a_{IH}$	**1.910**	1.065	**1.039**	**1.037**	**1.037**
AD_{IH}	**2.849**	2.739	**2.425**	**2.377**	**2.257**
$lab^{m_g}_{short}$	**1.1X**10^{-3}	9.9X10^{-4}	**9.5X**10^{-4}	**9.3X**10^{-4}	**9.2X**10^{-4}
$lab^{m_s}_{short}$	**7.1X**10^{-3}	0.110	0.079	0.495	0.248
A_{cv}	**0.000**	0.264	**0.630**	**0.736**	**0.723**
B_{cv}	**0.000**	0.284	**0.665**	**0.704**	**0.778**
q_{cv}	1.532	1.540	1.435	1.467	**1.359**
$p^{m_g}_{cv}$	**0.000**	0.179	**0.360**	**0.408**	**0.494**
$p^{m_s}_{cv}$	**0.087**	0.228	**0.386**	**0.428**	**0.509**
$p^{m_{AI}}_{cv}$	**2.5E-08**	7.5E-02	**3.5E-02**	**2.7E-02**	**2.7E-02**
$\mu^{m_g}_{cv}$	**0.000**	0.483	**0.534**	**0.542**	0.488
$\mu^{m_s}_{cv}$	0.816	0.782	0.794	0.729	0.750
$\mu^{m_{AI}}_{cv}$	**0.000**	0.556	**0.432**	**0.411**	**0.354**

Note: baseline scenario in red; black numbers are statistically different at 5% from the corresponding baseline.

Table 4. Experiment on $\Theta_{A,B}$

Index	$\Theta_q=0$	$\Theta_q=0.1$	$\Theta_q=0.5$	$\Theta_q=0.7$	$\Theta_q=1$
$\bar{A}$	**2.230**	2.294	2.326	2.269	2.375
$\bar{B}$	**2.448**	**2.434**	2.511	**2.418**	2.575
$\bar{q}$	**0.500**	**1.395**	**4.778**	**6.981**	9.488
ABQ	**2.4X**10^{-4}	**4.5X**10^{-4}	**5.3X**10^{-4}	**5.2X**10^{-4}	5.7X10^{-4}
$\bar{\sigma}^{m_g}$	0.178	0.196	0.175	0.164	0.178
$\bar{\sigma}^{m_s}$	**0.179**	**0.766**	0.942	0.944	0.939
$\bar{\sigma}^{m_{AI}}$	**0.832**	**0.811**	0.773	**0.806**	0.765
$\bar{\pi}_{IH}$	**0.189**	**0.112**	0.134	0.132	0.133
$\bar{rd}_{IH}$	**0.278**	**0.232**	0.193	0.188	0.187
$\bar{Y}^a_{IH}$	**1.984**	**1.136**	1.072	1.064	1.065
AD_{IH}	**2.904**	**2.843**	2.654	**2.783**	2.666
$lab^{m_g}_{short}$	**9.7X**10^{-4}	**9.8X**10^{-4}	9.9X10^{-4}	9.9X10^{-4}	9.9X10^{-4}
$lab^{m_s}_{short}$	**6.4X**10^{-4}	0.402	1.379	0.208	0.574
A_{cv}	**0.267**	0.281	0.273	**0.270**	0.297
B_{cv}	**0.291**	**0.294**	0.297	**0.279**	0.316
q_{cv}	**0.000**	**0.299**	**0.974**	**1.260**	1.518
$p^{m_g}_{cv}$	**0.182**	**0.182**	0.193	**0.185**	0.200
$p^{m_s}_{cv}$	**0.182**	**0.225**	0.281	**0.228**	0.276
$p^{m_{AI}}_{cv}$	**0.059**	0.065	0.080	0.073	0.072
$\mu^{m_g}_{cv}$	0.468	0.484	0.475	0.468	0.478
$\mu^{m_s}_{cv}$	**0.467**	**0.700**	0.728	0.792	0.757
$\mu^{m_{AI}}_{cv}$	0.490	**0.475**	0.545	0.532	0.543

Note: baseline scenario in red; black numbers are statistically different at 5% from the corresponding baseline.

Table 5. Experiment on Θ_q

5.2 On the effect of demand elasticities

The first set of exercises focuses on the components of firm fitness in the market for data (i.e. the sensitivity of firms' demand), as described in Eq. (13) and Eq. (14). We test a range of values for the parameters e_1 and e_2. The former represents the elasticity of fitness to unsatisfied demand; the latter assesses the role of quality in the fitness ranking of firms.

5.2.1 On the influence of unsatisfied demand

Tab. 6 displays the results on the main statistics of interest for different values of e_1 combined with government policy of accommodating ($\beta_0 = 1$), or counteracting ($\beta_0 = 0$) firms' investment decisions. When the public sector accommodates firms willingness, we observe that increasing e_1 does not drastically change the overall picture of the economy with respect to the baseline configuration. Concerning the effect on the industrial dynamics, the competitive environment and the monopolistic structure still characterize the market for generic and specialized data, respectively. Conversely, a closer inspection of the case in which $e_1 = 0.3$ reveals that the market structure switches from the leap-frogging that was characterizing the benchmark configuration to permanent monopoly of the same firm after $t = 1250$. Furthermore, when the government counteracts entrepreneurial policies on sponsorship, the threat of labour shortage in the market for specialized data is relieved.
When we deal with the effects on technological trajectories, it is important to note, first, that interesting results emerge with a counteracting policy by the government. The index that measures the technological trajectory of the economy, ABQ, is often statistically significant from the benchmark average. However, and overlooking the very tiny absolute values, there is not a clear and linear pattern. Moreover, if the elasticity to unfulfilled demand did not significantly affect the average quality of data traded in the market when $\beta = 1$, in the present setting e_1 influences its pattern in a non-linear way and, when significant, values different from the benchmark lead to a decline in average quality. This outcome is due to the fact that a greater amount of demand comes out of the market for generic data and out of the market for AI systems. Therefore, an enhanced percentage of profits and R&D investments will be diverted from product innovations (affecting quality) towards process innovation (affecting data and labour productivity), therefore lowering the average quality. The diminished quality level subsequently reduces the variation across quality and prices for specialized data.[21]

[21] We should spend a few words on average labour productivity: the significant decrease corresponding to $e_1 = 0.3$ is correlated to the monopolistic structure envisaged in the market for AI systems. A reduced amount of revenues to invest in R&D resulted in no process innovations from the laggards, strengthening the monopolistic leader on the one hand, but diminishing the occurrence of innovation on the other.

Index	$\beta_0 = 0$					$\beta_0 = 1$				
	$e_1 = 0.3$	$e_1 = 0.5$	$e_1 = 0.7$	$e_1 = 0.9$	$e_1 = 1$	$e_1 = 0.3$	$e_1 = 0.5$	$e_1 = 0.7$	$e_1 = 0.9$	$e_1 = 1$
$\bar{A}$	2.292	2.257	2.306	2.311	2.313	**2.120**	2.259	2.289	**2.352**	2.274
$\bar{B}$	2.443	2.347	2.523	2.470	2.519	2.396	2.414	2.498	2.430	2.369
$\bar{q}$	**8.406**	8.597	9.854	**8.596**	9.863	9.505	9.827	9.440	9.389	9.795
ABQ	5.4X10^{-4}	**5.3X10^{-4}**	**5.3X10^{-4}**	**5.3X10^{-4}**	**5.2X10^{-4}**	5.9X10^{-4}	5.6X10^{-4}	**5.3X10^{-4}**	5.8X10^{-4}	5.4210^{-4}
$\bar{\sigma}^{m_g}$	0.156	0.178	0.150	0.178	0.199	0.159	0.170	0.155	0.164	0.170
$\bar{\sigma}^{m_s}$	0.939	0.940	0.957	0.940	0.949	0.947	0.953	0.956	0.945	0.957
$\bar{\sigma}^{m_{AI}}$	0.784	0.806	0.801	0.790	0.797	**0.828**	0.793	0.798	0.818	0.796
$\bar{\pi}_{IH}$	**0.137**	0.124	0.137	**0.138**	0.133	0.125	0.125	0.139	**0.142**	0.133
$\bar{rd}_{IH}$	**0.192**	0.177	0.189	0.188	0.184	0.175	0.182	0.194	0.191	0.184
$\bar{Y}^a_{IH}$	1.063	1.059	1.063	1.060	1.059	**1.055**	1.062	1.063	1.058	1.058
AD_{IH}	**2.815**	**2.790**	2.709	2.748	2.733	2.731	2.722	2.744	2.716	2.739
$lab^{m_g}_{short}$	**2X10^{-3}**	**1.4X10^{-3}**	**1.3X10^{-3}**	**1.3X10^{-3}**	1.9X10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}
$lab^{m_s}_{short}$	**4X10^{-4}**	**5X10^{-4}**	**5X10^{-4}**	**5X10^{-4}**	**5X10^{-4}**	1.490	0.877	0.319	0.387	0.479
A_{cv}	0.291	0.260	0.292	0.272	0.257	0.256	0.273	0.276	0.289	0.271
B_{cv}	0.298	0.273	0.302	0.297	0.279	0.293	0.290	0.295	0.285	0.290
q_{cv}	**1.415**	1.443	1.555	**1.424**	1.570	1.526	1.576	1.550	1.536	1.555
$p^{m_g}_{cv}$	0.187	0.174	**0.198**	0.185	0.181	0.177	0.183	0.193	0.180	0.185
$p^{m_s}_{cv}$	**0.220**	**0.207**	0.241	**0.225**	**0.217**	0.240	0.256	0.253	0.235	0.246
$p^{m_{AI}}_{cv}$	0.070	0.058	0.074	0.067	0.068	0.059	0.065	0.081	0.076	0.067
$\mu^{m_g}_{cv}$	0.489	0.473	**0.450**	0.488	**0.460**	0.480	0.479	0.462	0.495	0.473
$\mu^{m_s}_{cv}$	0.726	0.725	0.786	0.778	0.774	0.755	0.749	0.784	0.786	0.763
$\mu^{m_{AI}}_{cv}$	0.546	**0.480**	0.548	0.578	0.533	0.545	0.559	0.545	0.522	0.551

Note: baseline scenario in red; black numbers are statistically different at 5% from the corresponding baseline.

Table 6. Experiment on e_1

5.2.2 On the influence of the quality of data

Tab. 7 refers to the results after a change in e_2. With regards to the industrial dynamics, the higher e_2, the higher the tendency to reach a concentrated structure in the market for specialised data. At the same time, for values of e_2 below 0.3, this market structure switches towards a leap-frogging dynamics in which no firm manages to remain in a monopoly position. It is, however, important to note a difference depending on the government policy: when $\beta_0 = 1$, we observed a leap-frogging dynamics with ongoing monopolistic waves. In contrast, when β_0 turns to 0, leap-frogging is substituted by a duopolistic structure in which the two best-quality firms share the market after a period of intense competition. Interestingly, these two winners do not sell data of the same quality, but one provides strictly better data. When the elasticity under analysis is lower than unity and hence not large enough, the *best* firm does not manage to claim all the market for specialised data. The rise to duopoly explains also the increased value assumed by the inverse Herfindahl index for profits, R&D, revenues and demand with low values of e_2. The disappearance of pure and frozen monopolistic structures in every market raises the average number of sources of funds. When analyzing the availability of scientists, we find that labour shortage may occur in the market for specialized data only and as long as the public sectory accommodates entrepreneurial policy. Moreover, we find a negative relationship between e_2 and the supply of labour.[22]

[22]The survival of some form of competition through innovation in every market explains the increase in the inverse Herfindahl index referring to aggregate demand, revenues and R&D. At the same time, the lower the elasticity with respect to quality, the lower the variation in the price and mark-up applied to specialised data, when significantly different to the benchmark.

Index	$\beta_0 = 0$					$\beta_0 = 1$				
	$e_2 = 0.3$	$e_2 = 0.5$	$e_2 = 0.7$	$e_2 = 0.9$	$e_2 = 1$	$e_2 = 0.3$	$e_2 = 0.5$	$e_2 = 0.7$	$e_2 = 0.9$	$e_2 = 1$
$\bar{A}$	2.271	2.239	2.280	**2.389**	2.373	2.368	2.281	2.274	2.264	2.249
$\bar{B}$	2.546	2.480	2.410	2.514	2.578	2.456	2.409	2.512	2.483	2.444
$\bar{q}$	7.856	8.410	9.476	8.836	8.982	7.938	8.713	9.464	9.973	9.090
ABQ	**$8.2\text{X}10^{-4}$**	$5.8\text{X}10^{-4}$	$5.7\text{X}10^{-4}$	$5.4\text{X}10^{-4}$	$5.6\text{X}10^{-4}$	**$9.2\text{X}10^{-4}$**	**$6.9\text{X}10^{-4}$**	$5.8\text{X}10^{-4}$	**$6.1\text{X}10^{-4}$**	$5.6\text{X}10^{-4}$
$\bar{\sigma}^{m_g}$	0.184	0.178	0.184	0.175	0.178	0.176	0.161	0.202	0.157	0.173
$\bar{\sigma}^{m_s}$	**0.659**	**0.828**	**0.867**	0.917	0.926	**0.670**	**0.776**	0.938	0.930	0.943
$\bar{\sigma}^{m_{AI}}$	0.767	0.794	0.801	0.787	**0.770**	0.769	0.797	0.812	0.799	0.808
$\bar{\pi}_{IH}$	**0.108**	0.131	0.129	0.134	0.133	0.126	0.125	0.122	0.125	0.131
$\bar{rd}_{IH}$	**0.211**	0.192	0.182	0.190	0.186	**0.230**	0.189	0.179	0.182	0.182
$\bar{Y}^a_{IH}$	**1.118**	**1.073**	1.064	**1.067**	1.064	**1.120**	**1.075**	1.066	1.065	1.060
AD_{IH}	**2.858**	**2.862**	**2.793**	**2.814**	2.717	**2.827**	**2.775**	**2.747**	2.720	2.680
$lab^{m_g}_{short}$	**$1.9\text{X}10^{-3}$**	$4.5\text{X}10^{-3}$	**$1.8\text{X}10^{-3}$**	**$1.5\text{X}10^{-3}$**	**$1.2\text{X}10^{-3}$**	**$9.8\text{X}10^{-4}$**	$9.8\text{X}10^{-4}$	$9.9\text{X}10^{-4}$	$9.9\text{X}10^{-4}$	$9.9\text{X}10^{-4}$
$lab^{m_s}_{short}$	**$3.83\text{X}10^{-4}$**	**$4\text{X}10^{-4}$**	**$4.2\text{X}10^{-4}$**	**$4.3\text{X}10^{-4}$**	**$4.6\text{X}10^{-4}$**	2.993	1.257	**0.189**	0.378	0.888
A_{cv}	0.282	0.264	0.276	**0.299**	**0.296**	0.276	0.280	0.265	0.268	0.273
B_{cv}	0.313	0.303	0.289	0.303	0.316	0.283	0.294	0.292	0.290	0.303
q_{cv}	1.515	1.430	1.562	1.468	1.460	**1.662**	1.555	1.555	1.617	1.506
$p^{m_g}_{cv}$	0.193	0.191	0.187	0.193	0.199	0.180	0.187	0.189	0.182	0.189
$p^{m_s}_{cv}$	**0.220**	**0.218**	**0.216**	**0.223**	**0.242**	**0.239**	0.274	**0.230**	0.254	0.296
$p^{m_{AI}}_{cv}$	0.078	0.068	0.076	0.068	0.072	0.083	0.076	0.068	0.071	0.071
$\mu^{m_g}_{cv}$	**0.470**	0.470	0.469	**0.456**	0.486	0.472	**0.470**	0.490	0.489	0.492
$\mu^{m_s}_{cv}$	**0.616**	**0.633**	0.708	0.744	0.755	**0.536**	**0.612**	0.744	0.745	0.726
$\mu^{m_{AI}}_{cv}$	**0.552**	0.553	**0.556**	**0.538**	0.532	0.525	**0.562**	0.525	0.514	0.493

Note: baseline scenario in red; black numbers are statistically different at 5% from the corresponding baseline.

Table 7. Experiment on e_2

On the technological side, we do not find any noticeable variation in labour productivity, data productivity and quality. The influence of e_2 on the technological trajectory is circumscribed to $\beta_0 = 1$, in which higher values of this elasticity are associated with lower value of ABQ.

5.3 On the influence of the Public Sector

The second battery of experiments concerns to the role of the government. We focus, first, on the growth rate of public spending, g_{PS}, that applies both to the purchase of data and to the training of scientists. In a second exercise, we focus on k, the elasticity that links the supply of labour with public and private sponsorships as in Eq. (34). Results are in Tab. 8 and Tab. 9. In Fig. 7 and Fig. 8 we combine different values for both g_{PS} and k.

5.3.1 On the effect of public spending

As usual, we begin to elaborate on the industrial dynamics as in Tab. 8. The government affects the market structure by reinforcing tendencies already in place. On the one hand, albeit concentration increases with higher growth rates g_{PS}, competition among a narrow set of oligopolists is preserved in the market for generic data. On the other hand, the monopolistic structure is positively related with the rate g_{PS}. Conversely, we find that higher growth in public spending is associated with a lower average share in the market for AI systems when $\beta_0 = 0$, although we argue that this pattern is neither clear nor enough to alter the overall leap-frogging dynamics. The second feature we underline is about labour shortage: increasing the growth of public spending in the training of scientists obviously

Index	$\beta_0 = 0$				$\beta_0 = 1$			
	$g_{PS} = 0.005$	$g_{PS} = 0.01$	$g_{PS} = 0.015$	$g_{PS} = 0.02$	$g_{PS} = 0.005$	$g_{PS} = 0.01$	$g_{PS} = 0.015$	$g_{PS} = 0.02$
$\bar{A}$	2.332	**2.674**	**3.570**	**4.303**	2.374	**2.858**	**3.539**	**4.344**
$\bar{B}$	2.442	**2.860**	**3.735**	**4.452**	2.574	**3.007**	**3.742**	**4.469**
$\bar{q}$	9.445	**12.910**	**18.851**	**23.340**	9.487	**13.173**	**18.272**	**24.154**
ABQ	**5.4X**10^{-4}	**6.4X**10^{-4}	**7X**10^{-4}	**7.2X**10^{-4}	5.7X10^{-4}	**6.3X**10^{-4}	**6.9X**10^{-4}	**7.3X**10^{-4}
$\bar{\sigma}^{m_g}$	0.165	**0.225**	**0.252**	**0.292**	0.178	**0.225**	**0.248**	**0.276**
$\bar{\sigma}^{m_s}$	0.944	0.965	**0.977**	**0.978**	0.938	0.957	**0.971**	**0.975**
$\bar{\sigma}^{M_{AI}}$	0.792	**0.791**	0.749	**0.720**	0.765	0.784	0.753	0.727
$\bar{\pi}_{IH}$	0.137	0.139	0.130	**0.149**	0.133	0.138	0.144	**0.151**
$\bar{rd}_{IH}$	0.192	0.191	0.186	**0.200**	0.186	0.193	**0.198**	**0.204**
$\bar{Y}^a_{IH}$	1.064	1.064	1.066	1.069	1.065	1.066	1.069	1.070
AD_{IH}	**2.792**	2.703	2.704	2.654	2.666	**2.734**	2.665	2.643
$lab^{m_g}_{short}$	**1.4X**10^{-3}	**1.1X**10^{-3}	**1.2X**10^{-3}	**1.2X**10^{-3}	10^{-3}	**9.2X**10^{-4}	**9.3X**10^{-4}	**9.4X**10^{-4}
$lab^{m_s}_{short}$	**4.3X**10^{-4}	**4X**10^{-4}	**4X**10^{-4}	**4.1X**10^{-4}	0.572	**0.048**	**5X**10^{-4}	**5.8X**10^{-4}
$\tilde{A}_{cv}$	0.287	0.304	**0.401**	**0.414**	0.297	**0.333**	**0.374**	**0.427**
B_{cv}	**0.290**	0.321	**0.415**	**0.410**	0.316	0.336	**0.396**	**0.432**
q_{cv}	1.499	**1.824**	**2.373**	**2.616**	1.518	**1.897**	**2.242**	**2.633**
$p^{m_g}_{cv}$	0.194	0.203	**0.256**	**0.279**	0.200	**0.219**	**0.260**	**0.300**
$p^{m_s}_{cv}$	**0.223**	**0.239**	0.281	**0.294**	0.276	**0.247**	0.292	**0.322**
$p^{M_{AI}}_{cv}$	0.086	0.073	0.080	0.076	0.072	0.082	0.075	0.081
$\mu^{m_g}_{cv}$	0.484	0.494	0.484	**0.445**	0.478	0.483	**0.498**	0.460
$\mu^{m_s}_{cv}$	0.751	0.793	0.797	**0.816**	0.757	0.770	**0.839**	**0.810**
$\mu^{M_{AI}}_{cv}$	0.556	0.586	**0.593**	**0.603**	0.543	0.558	0.551	0.568

Note: baseline scenario in red; black numbers are statistically different at 5% from the corresponding baseline.

Table 8. Experiment on g_{PS}

relaxes the labour constraint and the economy works with no shortage of input.[23]

The most important outcome from this exercise involves technology-related factors. The greater the average labour and data productivity, and the average quality with respect to the baseline scenario. When the growth rate switches from the initial 0.5%, average productivity and quality raise from 2.4 and 9.5 on average to 4.4 and 24, respectively. The relationship is always positive and statistically significant, with the exception of labour productivity and quality when we compare the benchmark to the couple $\{g_{PS} = 0.005; \beta_0 = 0\}$. In addition to this, results point to an increase in the variation of both productivity and quality indexes, possibly associated with a market structure that presents a reinforcing pattern toward concentration. In this respect, the only differences between the scenario in which the government accommodates and the one in which it contrasts are the average values assumed by quality, labour and data productivity. Even if the pattern is the same, these values look slightly lower in the latter case than in the former setting. Finally, a positive relation between ABQ and g_{PS} seems to exist. Yet, if, from a quantitative point of view, the components of the technological trajectory start diverging in growth terms, related absolute values are always that little to be worth of consideration.

5.3.2 On the elasticity of labour supply

This experiment concerns the elasticity of the supply of labour to public and private sponsorship (Tab. 9). As discussed for the baseline scenario, there might be episodes of labour

[23] Though statistically significant from the baseline for most of the cases, we do not find any remarkable pattern for what concerns to the coefficients of variation in prices and mark-ups after changes to g_{PS}.

shortage in the market for specialised data, albeit the ratio between labour demand and its supply is below unity on average. To analyse the role of k, let us first consider what happens when we lower its value. For $\beta_0 = 1$, a smaller k entails a strong shortage of workers both in the market for generic data and in the market for specialised data. The presence of labour shortage completely changes the structure in every market. What could be thought of as an oligopolistic competition in the market for generic data approaches to full competition, with an average market share fluctuating around 3%. We argue something similar with respect to the market for specialised data: even if the average share declines from 93% to 32%, a closer inspection of simulations envisages that a leap-frogging dynamics rules in the short run, leaving the stand to a very aggressive competition in the long period, in which no firm emerges as winner. Conversely, the AI market reveals an intense leap-frogging in the first quarter of time simulations, before switching to monopoly in a later phase: all in all, the average market share looks slightly, but significantly, reduced.[24]

From what said, we see that a fiercer competitive environment enhances, on average, the number of markets from which firms draw their resources. In addition to this, besides significant improvements in data productivity when $k = 0.8$, the competitive setting in the market for specialised data is positively correlated with the average quality of data traded. For what concerns to the coefficients of variation, we find a common pattern for the aggregates of interest. In other terms, switching from $k = 1$ to $k = 0.8$ leads to a statistically significant increase in the average variation of productivity and prices, for instance, with respect to the baseline setting. However, this variation experiences a sizeable reduction if compared to the benchmark when $k = 0.6$. Exceptions are constituted by mark-ups and by the prices applied to specialised data, whose variation is affected by the initial and sustained leap-frogging market structure.

The outcomes above generally hold from a qualitative perspective when we combine a counteracting government policy to a reduction in the elasticity k. The main difference concerns to average quality, whose magnitude is not statistically significant from the baseline: the reason lies in the greater deviation from the average among Monte Carlo runs, that greatly lowers the value of the t-test.[25]

We leave as last point the analysis of gradual increases from $k = 1$ to $k = 1.4$. The first consideration is that we do not observe neither straightforward nor statistically significant differences with respect to the benchmark setting, also when associated with differences in β_0. The main exceptions are an increased coefficient of variation for labour and data pro-

[24] Figures available upon request.

[25] Looking at sample simulations, we noticed that in the market for generic data there is a turnover between full competition and leap-frogging, and full competition again. The point can be related to the fact that whenever a firm gains a prominent position in the market because of the introduction of a novelty, the problem of labour shortage does not allow for an ever-lasting monopolistic position in the market. In order to satisfy the demand for data, that firm must buy from others, thus re-establishing competition.

Index			$\beta_0 = 0$					$\beta_0 = 1$		
	$k = 0.6$	$k = 0.8$	$k = 1$	$k = 1.2$	$k = 1.4$	$k = 0.6$	$k = 0.8$	$k = 1$	$k = 1.2$	$k = 1.4$
$\bar{A}$	2.243	2.358	2.257	2.274	2.317	2.344	2.384	2.311	2.362	**2.471**
$\bar{B}$	**2.230**	2.462	2.347	**2.586**	**2.576**	2.415	**2.526**	2.425	**2.576**	2.494
$\bar{q}$	9.026	8.833	8.597	9.718	9.377	**10.256**	9.163	9.113	8.509	9.079
ABQ	**8.8X10^{-4}**	**6.5X10^{-4}**	5.3X10^{-4}	5.4X10^{-4}	5.4X10^{-4}	**7.3X10^{-4}**	**7.4X10^{-4}**	5.3X10^{-4}	5.2X10^{-4}	5.4X10^{-4}
$\bar{\sigma}^{m_g}$	**0.024**	**0.115**	0.178	0.155	0.202	**0.031**	0.171	0.165	0.181	0.186
$\bar{\sigma}^{m_s}$	**0.380**	**0.954**	0.940	0.944	0.953	**0.323**	**0.882**	0.934	0.945	0.949
$\bar{\sigma}^{M_{AI}}$	**0.571**	0.805	0.806	0.765	0.787	**0.700**	0.784	0.798	0.772	0.795
$\bar{\pi}_{IH}$	**0.896**	**0.352**	0.124	0.135	0.134	**0.862**	**0.255**	0.127	0.126	0.130
$\bar{rd}_{IH}$	**1.024**	**0.403**	0.177	0.191	0.189	**1.093**	**0.314**	0.183	0.182	0.185
$\bar{Y}^a_{IH}$	**1.266**	1.062	1.059	1.064	1.066	**1.370**	1.064	1.063	1.065	1.063
AD_{IH}	**1.658**	**2.111**	2.790	2.784	2.746	**1.837**	**2.192**	2.772	2.765	2.786
$lab^{m_g}_{short}$	**15.254**	**11.422**	**1.4X10^{-3}**	**3.2X10^{-4}**	**1.1X10^{-4}**	**25.617**	**10.531**	9.9X10^{-4}	**2.8X10^{-4}**	**9X10^{-5}**
$lab^{m_s}_{short}$	**1968.797**	3.028	**4X10^{-4}.5X10^{-4}**	**1.6X10^{-4}**	**6X10^{-5}**	**762.404**	**11.774**	7.7X10^{-3}	0.129	5.3X10^{-2}
$\tilde{A}_cv$	**0.227**	0.278	0.260	**0.277**	0.271	**0.205**	**0.300**	0.267	**0.289**	**0.294**
B_{cv}	**0.229**	0.293	0.273	0.316	0.301	**0.225**	**0.323**	0.281	**0.312**	0.298
q_{cv}	**1.019**	1.538	1.443	1.540	1.480	**0.852**	1.597	1.490	1.449	1.467
$p^{m_g}_{cv}$	**0.047**	**0.240**	**0.174**	0.194	0.189	**0.093**	**0.330**	0.186	0.193	0.189
$p^{m_s}_{cv}$	**0.921**	**0.754**	0.207	0.224	**0.224**	**0.792**	**1.139**	0.219	0.227	0.218
$p^{M_{AI}}_{cv}$	**0.101**	0.072	**0.058**	0.085	0.073	0.070	**0.057**	0.065	**0.083**	0.069
$\mu^{m_g}_{cv}$	**0.670**	**0.498**	0.473	0.469	0.455	**0.757**	**0.521**	0.465	0.483	0.469
$\mu^{m_s}_{cv}$	**0.827**	**0.716**	0.725	0.797	0.776	0.798	**0.641**	0.784	0.747	0.764
$\mu^{M_{AI}}_{cv}$	**0.629**	0.531	0.480	**0.601**	**0.556**	**0.582**	**0.537**	0.490	**0.585**	**0.545**

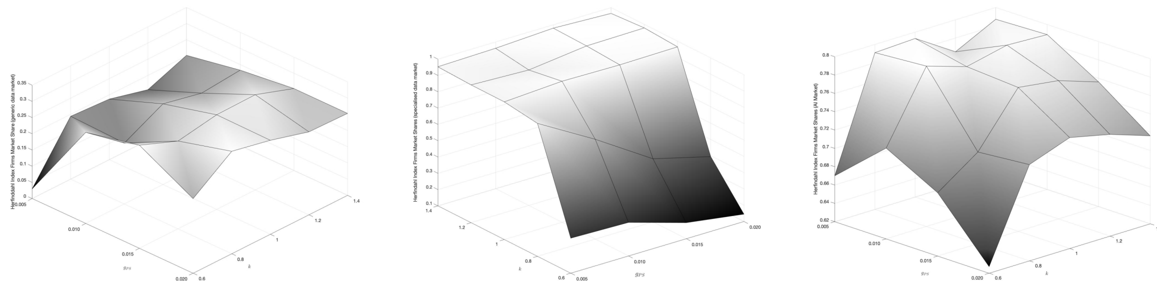
Note: baseline scenario in red; black numbers are statistically different at 5% from the corresponding baseline.

Table 9. Experiment on k

ductivity when $\beta_0 = 1$, and a rise in average data productivity when $k = 1.2$ and $\beta_0 = 0$. To conclude, this exercise does not offer particular insights on technological factors if not with reference to specific combinations of parameters. Yet, we suggest a negative, though uncertain, relationship between the elasticity k and ABQ. The greater the former, the lower the latter, whilst we shall remark once more that average values remain very small in absolute terms.

To deepen the analysis and somehow recap last two experiments, we have also performed a further exercise with combinations of several values for g_{PS} and k while leaving $\beta_0 = 1$. Fig. 7 presents the patterns followed by the Herfindahl index in firms market shares. Concerning the market for generic data, different combinations of g_{PS} and k let the average share follow a hump-shaped pattern. Although competition is always a robust feature of this sub-market, the environment seems more concentrated when the growth rate in public spending is in between 0.01 and 0.015, and the elasticity k is slightly larger than unity. Conversely, the lower k, the fiercer the competition and the lower the Herfindahl index. When we analyse the market for specialized data, we notice the public spending is able only to strengthen the ongoing decrease in the elasticity k. The impelling labour shortage enables the economy to re-allocate demand among a large number of firms. Conversely, the market for AI technologies involves a tent shape which does not alter the overall leap-frogging dynamics. From a quantitative point of view, higher value of k are associated with a higher weighted average market share. On the other hand, a hump-shaped relationship links the same average share to the growth rate in public spending.

Focusing on the technological indicators, Fig. 8 represents the patterns about labour pro-



Note: pictures concern to the Herfindhal index of the shares in the market for generic data (left), specialized data (centre) and AI market (right); experiments consider $\beta_0 = 1$.

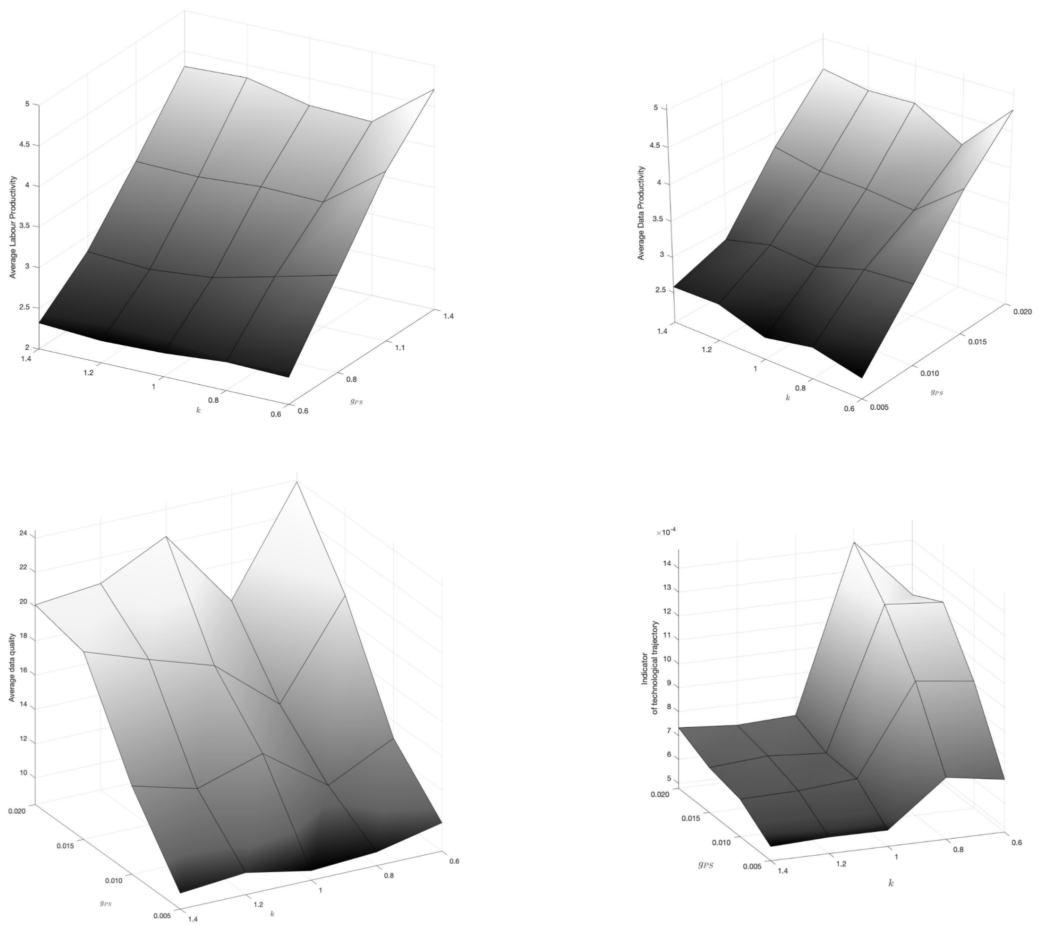
Figure 7. Experiment on g_{PS}: market shares

ductivity, data productivity, quality and ABQ. In this context, the government expenditure matters more: the higher g_{PS}, the higher all the statistics represented in the chart. The elasticity k exerts a noticeable influence only when the corresponding value approaches to unity. In contrast, k has a clear quantitative effect on ABQ. Precisely, a negative relationship exists between these two variables of interest.

From what said, the growth rate in public expenditure and the elasticity of labour supply to public and private investments affect the system through different channels when jointly assessed. On the one hand, the elasticity k impacts on the market structure and the lower its value, the stronger the competitive pressure. On the other hand, the growth rate in public spending is positively associated with a better performance in terms of productivity and data quality.

6 Conclusions

The dramatic increase in the collection and production of data along with the sizable improvements in AI-based technologies raised questions on whether we are at the dawn of a new industrial revolution in which digital technologies are at the helm. As suggested both by academic scholars and economic press, Artificial Intelligence is data-hungry, and wherever there is AI, there are data, and perhaps vice-versa. This (over?) abundance of data unveils new market opportunities where the collection, treatment and storage of data are only few examples of future data-fuelled business models (Vannuccini and Prytkova, 2021). In this paper we focused on this fuel of AI and analyzed in which way the nature of data and AI technologies, both in terms of inputs and outputs, mutually influence the corresponding technological trajectories and what are the selection mechanisms at work. Inspired by the Schumpeterian and evolutionary wisdom, we have devised an agent-based simulation model in which firms compete on several markets by producing and selling data. Moreover, they undertake innovative search to ameliorate the Artificial Intelligence capacity, which



Note: graphics refer to labour productivity (top left), data productivity (top right), data quality (bottom left) and ABQ (bottom right); experiments consider $\beta_0 = 1$.

Figure 8. Experiment on g_{PS}: technological indicators

is essential to increase both the efficiency in production and the quality of data traded. The demand-led framework is further augmented by the presence of a public sector with a twofold objective: on the one hand, it fuels the purchase of data and AI technologies, while at the same time it grows and trains generations of scientists with the option of accommodating or counteracting entrepreneurial willingness on which kind of scientists to supply.
Among the insights we were able to draw, there is the evidence that, even though the three markets of interest share a common technological trajectory, the interaction with the peculiarities of each marketplace gives rise to different and precise market structures. This result somewhat contrasts with the established view according to which industrial dynamics and patterns of innovative activity can be described by supply-side technology-related factors (Malerba and Orsenigo, 1996a). Our framework suggests that dynamics might be quite more complex and related to the interaction with demand components and the availability of inputs. Indeed, even if data inputs are infinitely reproducible basically at no cost, the market structure and the thrust of an industry depend also on the availability of *secondary* inputs such as labour, in the form of scientists.
Once we ascertained that there are limits to the growth of an industry despite the infinite reproducibility of the core good involved, a flow of open questions emerges. For example, they may re-open debates on the value of labour, whose *marginal utility* could have become infinite if we had adopted a neoclassical setting. Or questions related to the value and the storage of data: even if they are infinitely reproducible, their physical storage raises concerns on environmental issues such as energy consumption. In this respect, data might represent the new land as a means of production, with inherent diminishing returns in storage. These and further research questions will beef up our future research on the topic.

Acknowledgements

This research has received financial support from the French National Research Agency [reference: DInnAMICS -ANR-18-CE26-0017-01].

References

Acemoglu, D., Autor, D., Hazell, J., and Restrepo, P. (2020). AI and Jobs: Evidence from Online Vacancies. Technical report, National Bureau of Economic Research.

Acemoglu, D. and Restrepo, P. (2017). Secular stagnation? The effect of aging on economic growth in the age of automation. *American Economic Review*, 107(5):174–79.

Acemoglu, D. and Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2):3–30.

Aghion, P., Jones, B. F., and Jones, C. I. (2017). Artificial intelligence and economic growth. Technical report, National Bureau of Economic Research.

Aghion, P., Jones, B. F., and Jones, C. I. (2019). *9. Artificial Intelligence and Economic Growth*. University of Chicago Press.

Agrawal, A., McHale, J., and Oettl, A. (2019). *5. Finding Needles in Haystacks: Artificial Intelligence and Recombinant Growth*. University of Chicago Press.

Ahmed, N. and Wahed, M. (2020). The De-democratization of AI: Deep Learning and the Compute Divide in Artificial Intelligence Research. *arXiv preprint arXiv:2010.15581*.

Almudi, I., Fatas-Villafranca, F., and Izquierdo, L. R. (2012). Innovation, catch-up, and leadership in science-based industries. *Industrial and Corporate Change*, 21(2):345–375. Publisher: Oxford University Press.

Annoni, A., Benczur, P., Bertoldi, P., Delipetrev, B., De Prato, G., Feijoo, C., Macias, E. F., Gutierrez, E. G., Portela, M. I., and Junklewitz, H. (2018). Artificial Intelligence: A European Perspective. Technical report, Joint Research Centre (Seville site).

Armstrong, S., Bostrom, N., and Shulman, C. (2016). Racing to the precipice: a model of artificial intelligence development. *AI & society*, 31(2):201–206. Publisher: Springer.

Arrieta-Ibarra, I., Goff, L., Jiménez-Hernández, D., Lanier, J., and Weyl, E. G. (2018). Should we treat data as labor? Moving beyond" free". In *aea Papers and Proceedings*, volume 108, pages 38–42.

Bianchini, S., Müller, M., and Pelletier, P. (2020). Deep Learning in Science. *arXiv preprint arXiv:2009.01575*.

Bordot, F. (2022). Artificial Intelligence, Robots and Unemployment: Evidence from OECD Countries. *Journal of Innovation Economics Management*, 37(1):117–138. Publisher: De Boeck Supérieur.

Brynjolfsson, E., Rock, D., and Syverson, C. (2019). *1. Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics*. University of Chicago Press.

Chen, L., Bolton, P., Holmström, B. R., Maskin, E., Pissarides, C. A., Spence, A. M., Sun, T., Sun, T., Xiong, W., and Yang, L. (2021). Understanding Big Data: Data Calculus in the Digital Era.

Ciarli, T. and Lorentz, A. (2010). Product variety and changes in consumption patterns: The effects of structural change on growth. In *13th ISS Conference, Aalborg*.

Ciarli, T., Lorentz, A., Savona, M., and Valente, M. (2010). The effect of consumption and production structure on growth and distribution. A micro to macro model. *Metroeconomica*, 61(1):180–218. Publisher: Wiley Online Library.

Cockburn, I. M., Henderson, R., and Stern, S. (2018). The impact of artificial intelligence on innovation. Technical report, National bureau of economic research.

Domini, G., Grazzi, M., Moschella, D., and Treibich, T. (2021). Threats and opportunities in the digital era: automation spikes and employment dynamics. *Research Policy*, 50(7):104137. Publisher: Elsevier.

Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research policy*, 11(3):147–162. Publisher: Elsevier.

Dosi, G. (1988). Sources, procedures, and microeconomic effects of innovation. *Journal of economic literature*, pages 1120–1171. Publisher: JSTOR.

Dosi, G., Fagiolo, G., and Roventini, A. (2010). Schumpeter meeting Keynes: A policy-friendly model of endogenous growth and business cycles. *Journal of Economic Dynamics and Control*, 34(9):1748–1767. Publisher: Elsevier.

Dosi, G., Llerena, P., and Labini, M. S. (2006). The relationships between science, technologies and their industrial exploitation: An illustration through the myths and realities of the so-called 'European Paradox'. *Research policy*, 35(10):1450–1464. Publisher: Elsevier.

Dosi, G. and Nelson, R. R. (2010). Technical change and industrial dynamics as evolutionary processes. *Handbook of the Economics of Innovation*, 1:51–127. Publisher: Elsevier.

Economist, T. (2017). The world's most valuable resource is no longer oil, but data. *The Economist: New York, NY, USA*.

Furman, J. and Seamans, R. (2019). AI and the Economy. *Innovation policy and the economy*, 19(1):161–191. Publisher: University of Chicago Press Chicago, IL.

Gatti, D. D., Desiderio, S., Gaffeo, E., Cirillo, P., and Gallegati, M. (2011). *Macroeconomics from the Bottom-up*, volume 1. Springer Science & Business Media.

Gundersen, O. E. and Kjensmo, S. (2018). State of the art: Reproducibility in artificial intelligence. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32. Issue: 1.

Hagendorff, T. and Meding, K. (2020). The big picture: ethical considerations and statistical analysis of industry involvement in machine learning research. *arXiv preprint arXiv:2006.04541*.

Jones, C. I. and Tonetti, C. (2020). Nonrivalry and the Economics of Data. *American Economic Review*, 110(9):2819–58.

Klinger, J., Mateos-Garcia, J., and Stathoulopoulos, K. (2020). A narrowing of AI research? *arXiv preprint arXiv:2009.10385*.

Klinger, J., Mateos-Garcia, J. C., and Stathoulopoulos, K. (2018). Deep learning, deep change? Mapping the development of the Artificial Intelligence General Purpose Technology. *Mapping the Development of the Artificial Intelligence General Purpose Technology (August 17, 2018)*.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444. Publisher: Nature Publishing Group.

Malerba, F., Nelson, R., Orsenigo, L., and Winter, S. (1999). 'History-friendly'models of industry evolution: the computer industry. *Industrial and corporate change*, 8(1):3–40. Publisher: Oxford University Press.

Malerba, F., Nelson, R., Orsenigo, L., and Winter, S. (2001). History-friendly models: an overview of the case of the computer industry. *Journal of Artificial Societies and Social Simulation*, 4(3):6. Publisher: Journal of Artificial Societies and Social Simulation.

Malerba, F., Nelson, R., Orsenigo, L., and Winter, S. (2007). Demand, innovation, and the dynamics of market structure: The role of experimental users and diverse preferences. *Journal of Evolutionary Economics*, 17(4):371–399. Publisher: Springer.

Malerba, F. and Orsenigo, L. (1995). Schumpeterian patterns of innovation. *Cambridge journal of Economics*, 19(1):47–65. Publisher: Oxford University Press.

Malerba, F. and Orsenigo, L. (1996a). The dynamics and evolution of industries. *Industrial and Corporate change*, 5(1):51–87. Publisher: Oxford University Press.

Malerba, F. and Orsenigo, L. (1996b). Schumpeterian patterns of innovation are technology-specific. *Research policy*, 25(3):451–478. Publisher: Elsevier.

Malerba, F. and Orsenigo, L. (2002). Innovation and market structure in the dynamics of the pharmaceutical industry and biotechnology: towards a history-friendly model. *Industrial and corporate change*, 11(4):667–703. Publisher: Oxford University Press.

Nagaraj, A., Shears, E., and de Vaan, M. (2020). Improving data access democratizes and diversifies science. *Proceedings of the National Academy of Sciences*, 117(38):23490–23498. Publisher: National Acad Sciences.

Nelson, R. R. and Sampat, B. N. (2001). Making sense of institutions as a factor shaping economic performance. *Revista de Economía Institucional*, 3(5):17–51. Publisher: Revista de Economía Institucional.

Nelson, R. R. and Winter, S. G. (1982). An evolutionary theory of economic change. *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship*.

Posner, E. A. and Weyl, E. G. (2019). *Radical Markets*. Princeton University Press.

Savona, M. (2019). The value of data: Towards a framework to redistribute it.

Schumpeter, J. A. (1934). The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle (1912/1934). *Transaction Publishers.–1982.–January*, 1:244.

Schumpeter, J. A. (1942). *Capitalism, socialism and democracy*. routledge.

Silverberg, G. and Verspagen, B. (1994). Learning, innovation and economic growth: a long-run model of industrial dynamics. *Industrial and Corporate Change*, 3(1):199–223. Publisher: Oxford University Press.

Silverberg, G. and Verspagen, B. (1995). An evolutionary model of long term cyclical variations of catching up and falling behind. *Journal of Evolutionary Economics*, 5(3):209–227. Publisher: Springer.

Spiekermann, M. (2019). Data marketplaces: Trends and monetisation of data goods. *Intereconomics*, 54(4):208–216. Publisher: Springer.

Traub, J., Quiané-Ruiz, J.-A., Kaoudi, Z., and Markl, V. (2019). Agora: Towards An Open Ecosystem for Democratizing Data Science & Artificial Intelligence. *ArXiv, abs/1909.03026*.

Vannuccini, S. and Prytkova, E. (2021). Artificial intelligence's new clothes? From general purpose technology to large technical system. *From General Purpose Technology to Large Technical System (April 7, 2021). SWPS*, 2.

Varian, H. (2019). *16. Artificial Intelligence, Economics, and Industrial Organization*. University of Chicago Press.

Yu, Z., Liang, Z., and Wu, P. (2021). How data shape actor relations in artificial intelligence innovation systems: an empirical observation from China. *Industrial and Corporate Change*, 30(1):251–267. Publisher: Oxford University Press.