Small Bus Econ (2022) 59:893–923
https://doi.org/10.1007/s11187-021-00558-6

Capacity constraints as a trigger for high growth

Alex Coad · Clemens Domnick · Florian Flachenecker · Peter Harasztosi · Mario Lorenzo Janiri · Rozalia Pal · Mercedes Teruel

Accepted: 27 May 2021 / Published online: 9 November 2021

Abstract High-growth enterprises (HGEs) have a large economic impact but are notoriously hard to predict. Previous research has linked high-growth episodes to the configuration of lumpy indivisible resources inside firms, such that high capacity utilisation levels might stimulate future growth. We theorize that firms reaching critically high capacity utilisation levels reach a "trigger point" involving either broad-based investment in further growth or shrinking back to previous levels. We analyze EIBIS survey data (matched to ORBIS) which features a question on time-varying capacity utilisation. Overcapacity is a transitory state. Firms enter into overcapacity after a period of the rapid growth of sales and profits, and the years surrounding overcapacity have higher employment growth rates. Firms operating at overcapacity make incremental investments (e.g. capacity expansion, process improvements and modern machinery) rather than investing in R&D and new product development. We find support for the "fork in the road" hypothesis: for some firms, overcapacity is associated with launching into massive investments and subsequent sales growth, while for other firms, overcapacity is negatively related to both investments and sales growth.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11187-021-00558-6.

A. Coad (✉)
Waseda Business School, Waseda University, Tokyo, Japan
e-mail: alex.coad@waseda.jp

C. Domnick · M. L. Janiri
European Commission, Joint Research Centre - European Commission, Inca Garcilaso 3, 41092 Seville, Spain
e-mail: Clemens.DOMNICK@ec.europa.eu

M. L. Janiri
e-mail: Mario.JANIRI@ext.ec.europa.eu

F. Flachenecker
Joint Research Centre, European Commission, Rue du Champ de Mars 21, 1050 Brussels, Belgium
e-mail: florian.flachenecker@ec.europa.eu; f.flachenecker@ucl.ac.uk

F. Flachenecker
University College London, Central House, 14 Upper Woburn Place, London WC1H 0NN, UK

P. Harasztosi · R. Pal
European Investment Bank, Luxembourg, Spain
e-mail: p.harasztosi@eib.org

R. Pal
e-mail: r.pal@eib.org

M. Teruel (✉)
Departament d'Economia, GRIT & ECO-SOS, Facultat d'Economia i Empresa, Universitat Rovira i Virgili, Av. Universitat 1, 43204 Reus, Spain
e-mail: mercedes.teruel@urv.cat

Plain English Summary Operating above maximum capacity is like a "fork in the road": while some firms shrink back to stay within existing capacity constraints, others respond by launching into broad-based growth. We develop a theory of firm growth, according to which some firms are better positioned for subsequent growth, depending upon their state of capacity utilisation. Firms with plenty of slack capacity can easily grow within the bounds of existing capacity constraints. Firms that are already operating above maximum capacity cannot grow by drawing on slack resources, but instead, their growth requires a broad-based investment in many interconnected areas. Our statistical analysis shows that operating above maximum capacity is relatively rare and unlikely to persist and seems due to rising demand. Some firms respond to being above maximum capacity by slowing down and adapting to existing capacity constraints, while others treat overcapacity as a "trigger point" that launches them into subsequent growth. Our results are of interest to those seeking to understand and predict firm growth: investors, entrepreneurs, academics and policymakers.

Keywords High-growth enterprises (HGEs) · Firm growth · Investment · Capacity utilisation · Trigger points

JEL classifications L25 · L26

1 Introduction

High-growth enterprises (HGEs) have received substantial interest from policymakers, academics and investors since the seminal contribution of Birch (1979). A stable finding across countries and time periods is that about 4% of firms create about 50% of the jobs (Storey, 1994). Interest in HGEs has grown with awareness of their disproportionately large contribution to economic growth and job creation (Delmar et al., 2003; Henrekson & Johansson, 2010; Coad et al., 2014; McKenzie, 2017; Grover Goswami et al., 2019; Flachenecker et al., 2020). However, a problem for "gazelle hunters" is that it is notoriously difficult to predict which firms will become HGEs (Fischer & Karlan, 2015). Part of the problem is that HGE status is often a temporary episode rather than a time-invariant firm characteristic (Daunfeldt & Halvarsson, 2015; Grover Goswami et al., 2019); hence, firms may drift in and out of the subsample of "potential" HGEs.

The episodic nature of high-growth events draws on the stylized fact that firm growth is not a smooth process but takes place in lumps, bumps and jumps (Arata, 2019). Firm growth rates follow a heavy-tailed distribution (rather than a normal distribution), such that while most firms do not grow much in any year, a handful of firms will have very fast growth or decline in each year (Bottazzi & Secchi, 2006).

The challenge for policymakers, interested in supporting HGEs on the grounds of their impressive job creation potential, is to target their policy interventions to nudge firms into HGE episodes. The fundamental question for policymakers, therefore, is whether rapid growth can be triggered. If the "trigger points" of rapid growth can be better understood, then policy interventions that are targeted at these trigger points could be a cost-effective way to "nudge" hesitant firms that are at a critical decision point in their growth path to launch into rapid growth.[1]

This study takes a step back and theorizes about the reasons why firms might suddenly launch into a period of high growth. Building on Penrose (1959) and her vision of firms as constantly shifting configurations of lumpy resources, we suggest that firms at a critical state of maximum use of resources will be uniquely positioned at the window of opportunity to launch into high growth. Previous literature has discussed "trigger points" for firm growth (Brown & Mawson, 2013), as well as investigating links between capacity utilisation and firm growth on a theoretical (Coad & Planck, 2012) and empirical (Pozzi & Schivardi, 2016) level, but we fill a gap in the literature regarding how high capacity utilisation might serve as a critical decision point for firms' growth trajectories.

This study makes several novel contributions. First, amid a paucity of empirical evidence on firm-level capacity utilisation levels and their relationship to firm

[1] Such policy interventions could perhaps include access to finance (for financing the expansion) or providing technical support to remove uncertainty surrounding the expansion path. An alternative to direct policy interventions, however, could be attempts to set up an environment whereby private actors (e.g. private equity funds) undertake these actions.

growth, we explore a rich large-scale questionnaire dataset that includes information on capacity utilisation levels. More specifically, we match the EIBIS survey (European Investment Bank Survey of Investment and Investment Finance) to the ORBIS database maintained by Bureau van Dijk. This dataset provides rich evidence that enables comparisons across many EU member states. Of central interest is the question asking a firm if it is "[o]perating at its maximum capacity attainable under normal conditions?", with four responses: above maximum capacity; at maximum capacity; somewhat below full capacity; substantially below full capacity. Second, we develop a theory of firm growth and capacity utilisation and formulate our novel "fork in the road" hypothesis whereby firms operating above maximum capacity will either invest massively in subsequent growth or shrink back to previous production levels. Third, we investigate the antecedents, characteristics and consequences of operating above maximum capacity levels, focusing in particular on showing that entry into a state of overcapacity is preceded by the rapid growth of sales and profits, and firms in overcapacity have higher employment growth before, during and after being at overcapacity. This is consistent with firms being pushed into a critically high level of capacity utilisation by rising demand. Firms in a state of overcapacity tend to make incremental investments in capacity expansion for existing products, and modern machinery, rather than investing in R&D and new product development. We find some evidence that firms entering into overcapacity reach a decision point (or fork in the road), with some firms taking overcapacity as an opportunity to launch into subsequent sales growth, while for other firms overcapacity is linked to a decline in sales.

The paper is organized as follows. Section 2 discusses the previous literature and presents our research questions. Section 3 presents our data. Section 4 contains our analysis. Section 5 discusses our results, and Section 6 concludes.

2 Background

2.1 Firm growth as a lumpy bumpy process

Growth is not a smooth process but occurs in lumps, bumps and jumps. Nearly 100 years ago, Ashton (1926) considered the growth patterns of British textile firms and observes that "In their growth they obey no one law. A few apparently undergo a steady expansion... With others, increase in size takes place by a sudden leap" (Ashton, 1926, pp. 572–3).

Relatedly, Doms and Dunne (1998) observe that firm investment takes place in concentrated bursts, such that in most years, firms do not invest in their capital stock—but when they do, the investment is huge: "51.9% of plants in a year increase their capital stock by less than 2.5%, while 11% of plants in a year increase their capital stock by more than 20%" (p415). Furthermore, "on average, half of a plant's total investment over the 1973–1988 period was performed in just three years" (p417).

Recently, researchers have investigated the Laplace distribution of firm growth rates, which is a robust fact of firm growth (Arata, 2019; Bottazzi & Secchi, 2006; Stanley et al., 1996). The Laplace distribution is heavy tailed compared to the Gaussian, which means that while most firms hardly grow from 1 year to the next, it is not unusual for a handful of firms, in each year, to experience relatively fast growth. The Laplace distribution of firm growth rates can be explained in terms of firms being composed of lumpy and interdependent resources, bundled together in multiples that do not match up, hence leading to excess capacity in various dimensions, yet striving towards a full utilisation of their resources, such that (depending upon the arrangement of resources within the firm and the degree of slack) there may be critical junctures at which the firm can only grow by taking a large leap forward (Coad & Planck, 2012; see Fig. 1 below).

2.2 Firms as bundles of discrete resources

In her influential book on firm growth, Penrose (1959) put forward a theory of firms being composed of "resources" that are indivisible in nature and such that the configuration of these lumpy resources provides the impetus and the direction of further growth.

We may consider that present-day firms are complex bundles of resources including employees, machines, software, raw materials, land and buildings, IPR assets, capabilities, product ranges, distribution networks, etc. (Penrose, 1959). Firms combine

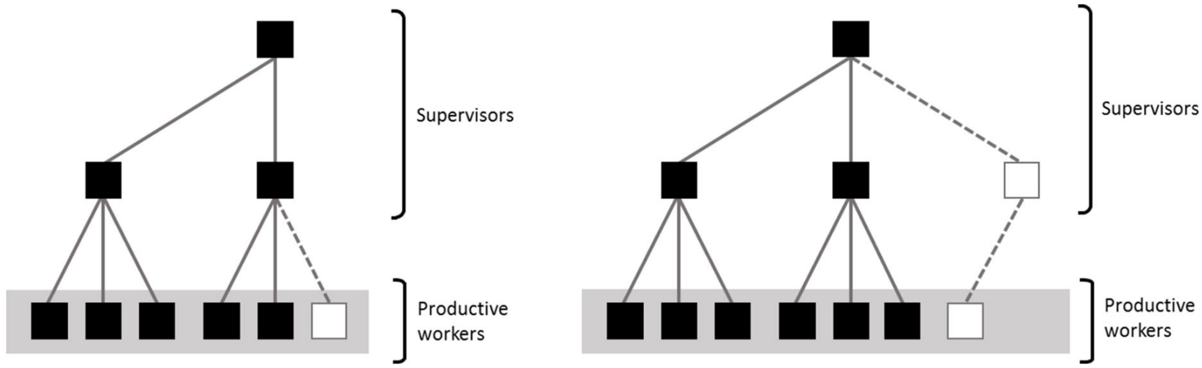


Fig. 1 Firms as bundles of discrete resources. In this model, each supervisor has a span of control of 3. Depending upon the "criticality" of the system, adding a production worker may lead to an increase in the number of supervisors further up the hierarchy. If there is some slack in the system, a production worker can be added, and new supervisors need not be added (see left). If, however, the attention of supervisors is already at full utilisation, the addition of a production worker will require the addition of a supervisor (see right). Source: based on Coad and Planck (2012)

these resources to assemble their productive capacity. These resources are affected by indivisibilities brought on by integer restrictions (Penrose, 1959), i.e. one cannot build half a building, buy half a machine or develop half a product. Penrose (1959, chapter 5) discussed the "balance of processes" to explain that the proper use of a collection of indivisible productive resources requires that the most efficient level of production corresponds to the least common multiple of these factors.[2]

According to Penrose, there are varying degrees of organisational slack, depending upon how the various resources are arranged in the firm. Firms are heterogeneous in their ability to respond to economic stimuli with the growth of production capacity (Pozzi & Schivardi, 2016). In some cases, production capacity can be expanded relatively easily, by adding one input (e.g. one production employee), and this new input can draw on existing slack (i.e. spare capacity) throughout the organisation. In other cases, adding input to an organisation that is already at a critical state of full utilisation of resources (i.e. with no organisational slack) will require broad-based investment in complementary resources. Some previous work has focused upon how growing firms can "change gears" to facilitate further growth by rearranging employees and tasks into a new organisational structure that involves the addition of hierarchical layers (Caliendo et al., 2015; Cruz et al., 2018). An important point is that there will be non-constant marginal costs of producing an extra unit, because the costs of expanding production will depend upon the degree of capacity utilisation (Butters, 2020).

In the case of a simple model of a single-product firm, Fig. 1 shows that adding a unit of production capacity has non-constant marginal costs that depend upon the degree of capacity utilisation in the organisation. In some cases, adding one production unit has cascading knock-on effects that trigger further growth.

The configuration of resources in a firm is continuously shifting. Firms are in a constant state of flux: employees come and go, the productivity of employees generally increases over time through learning effects and process innovations and heterogeneous productivity growth across tasks means that the sorting of tasks to employees is constantly being rearranged (for example if some tasks are automated using new software). As time goes by, the productivity of resources may change (e.g. due to learning or depreciation). The degree of organisational slack in the firm (as a complex system) is constantly evolving. In some cases, the configuration of resources in a firm, and opportunities brought on by the idiosyncratic balance of productive assets and capabilities at a particular point in time, can provide an impetus for further growth.

[2] This resembles the Leontief production function (Butters, 2020).

Adding a basic input to the firm can sometimes trigger an "avalanche" of further investments in productive capacity and firm growth. An analogy could be the "sandpile" model (Bak et al., 1993): randomly adding grains of sand to a sandpile will cause the sandpile to self-organize to approach a critical slope, and (depending upon the configuration of grains of sand in the pile) the addition of one more grain of sand could trigger avalanches of various sizes. In the context of business firms, an implication is that the marginal costs of growth are non-constant. At some times, it is easier for a firm to expand production by, e.g. 5% than at other times. These non-constant costs of growth are intimately related to a firm's state of capacity utilisation and degree of organisational slack and can be thought of as "trigger points" (Brown & Mawson, 2013) for high-growth episodes.

2.3 High-growth firms and "trigger points"

The concept of "trigger point" (Brown & Mawson, 2013) recognizes that high growth is episodic rather than a time-invariant trait of firms, and also, it is a concept of policy interest because policymakers are interested in targeting their interventions at these trigger points. Examples of trigger points from Brown and Mawson (2013) include discrete events such as new capital investments, new bank funding, changes in ownership (e.g. management buyouts or buy-ins; MBOs or MBIs) or boosts to sales coming from obtaining a new contract or customer. We could also add some further examples of trigger points: hiring a first employee (a daunting step which corresponds to doubling a firm's size; Coad et al., 2017; Fairlie & Miranda, 2017),[3] first steps into internationalisation, introducing a second product, building a second production plant, installing next-generation capital equipment, restructuring a firm's hierarchical layers to be better positioned for subsequent expansion (Caliendo et al., 2015; Cruz et al., 2018) or perhaps overcoming a regulatory threshold for firm size (Schivardi & Torrini, 2008; Garicano et al., 2016; Bornhäll et al., 2017). The challenge is to find a juncture where intervention can have an impact via knock-on effects such as subsequent adjustments, investments and hires.

We argue that policy interventions targeted at these trigger points could be a cost-effective way to leverage a step change or a discontinuity in a firm's growth paths. However, for such a policy intervention to be effective, it is necessary to accurately identify trigger points and also to understand exactly are the needs of firms upon reaching these trigger points (e.g. access to finance, technical support or temporary relief from the burdens that accompany larger size).

Trigger points are times of discontinuity when the firm is of a certain size and critical configuration of resources and capacity utilisation. Trigger points are ephemeral, and therefore, the policymaker only has a short window of opportunity. If successful, however, the policymaker could seize a cost-effective opening to leverage a large impact. However, timing is crucial. Policy interventions targeted at trigger points should, on the one hand, move fast enough to arrive in time of need and, on the other hand, should also be withdrawn shortly after the trigger point circumstances have passed.

Considering the importance of timing, policymakers must quickly assemble the information required to decide upon the eligibility of candidate firms. This presumably places the onus of data collection upon firms to self-select into the pool of candidates for support, instead of placing the onus of data collection upon policymakers.[4] As a consequence, firms should be aware of available support schemes so that they know when to apply for support. Given the short timescale for intervention, firms applying for such

[3] The hiring of a first employee can be seen as a daunting, once-in-a-lifetime gamble, that effectively corresponds to a doubling of the firm's size. Indeed, rapid growth is risky and has been shown to increase the chances of failure (Coad et al., 2020; Zhou and van der Zwan, 2019). Rapid growth is also linked to higher costs such as higher interest rates (Rostamkalaei and Freel, 2016). Risk-averse or untrusting entrepreneurs may also underestimate the gains from having an extra employee. Policy might have a role in stimulating firms to hire their first employee. For example, a temporary (e.g. 2 years) employment tax freeze for firms that hire their first employee could give them the time to appreciate how useful the extra employee is, to the point that they do not downsize once the policy is finished. This could bring about further growth, because research shows that growth is an "acquired taste" in that past growth contributes to the desire for future growth (Wiklund and Shepherd, 2003).

[4] See however Brown and Mawson (2013, p. 289), who write: "undoubtedly the key to identifying trigger points is to monitor firms closely". However, policymakers at the national and international level (e.g. the European Commission) often do not have the attention nor resources to constantly monitor firms in the search for their trigger points; therefore, it does not seem feasible for these national and supra-national policymakers to offer bespoke policies conditional on the idiosyncratic circumstances of individual firms.

support should not have to prepare large amounts of supporting documentation. There is no time for a case-by-case detailed evaluation of a dossier. Therefore, a policy must be simple and based on a small number of clearly verifiable indicators (e.g. employee records, tax returns) to avoid cases of fraud.

2.4 Capacity constraints as a trigger point

This paper investigates whether firms operating in a critical state of overcapacity are prone to launch into a high-growth episode. A constraint on new firm growth is that their market entry is rarely met by avid consumers, but instead, they must overcome informational barriers and slowly develop a reputation in order to accumulate a customer base (Foster et al., 2016). At a certain point, growing demand may pressure firms to increase their level of capacity utilisation, which (taking into account indivisibilities in machinery and installed capital) may result in adding worker shifts at overtime pay rates (Nikiforos, 2013). When firms fluctuate around maximum capacity utilisation levels, they potentially reach a decision point or a fork in the road. Firms do not always react to economic stimuli or "shocks" to launch into further expansion, but when they do, they are more responsive to demand shocks[5] than to productivity shocks (Pozzi & Schivardi, 2016). Therefore, rising demand may push firms into a new regime of prolonged expansion. Figure 2 provides an illustration of this "fork in the road" hypothesis.

The following illustrative example may help: A factory floor only has space for a maximum of six machines, each machine requires only 3 operators per 24 h (e.g. 8-h shifts each), and the maximum span of control for bosses is to have one supervisor for up to 9 machine operators. If a firm that is making full use of its existing capacity (i.e. with six machines, 18 machine operators and 2 bosses) suddenly gets a new production opportunity, it can either increase its production (which would require broad-based new investments in new machine operators, hence new bosses and new machines, hence new factory space) or it can forego the opportunity to expand production

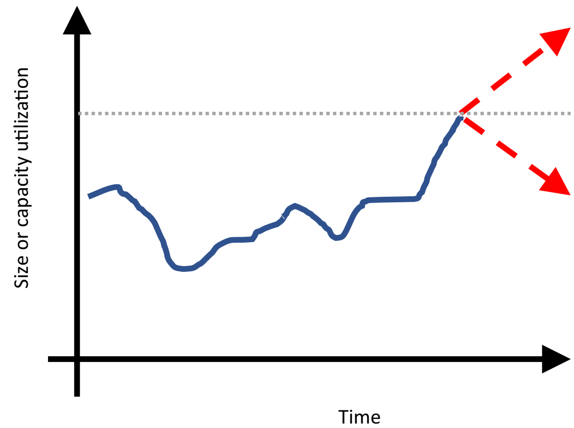


Fig. 2 The fork in the road hypothesis. Firms that reach a state of excessively high capacity utilisation will face a decision point either to continue their upward growth trajectory (which will need require large-scale investment) or shrink back to stay within existing capacity levels

(Coad & Planck, 2012). We suggest that such a firm would be at a fork in the road either it must invest massively in growth or it will stagnate and drift back from the state of full capacity utilisation.[6]

Box 1 illustrates how marginal costs of growth are not constant but depend upon the degree of capacity utilisation. In normal times, production can increase with minimal investments in raw materials and production employees. But, at critical junctures, growth can only proceed by jumps in production quantities and spikes in average costs and marginal costs, as firms must engage in broad-based investment to scale up their corporate infrastructure.

2.4.1 *The Fork in the Road hypothesis*

This paper investigates what happens when firms reach a critically high level of capacity utilisation. While some research into the behaviour of growing firms corresponds to the concept of trigger points (discussed above), nevertheless, there is a gap in

[5] Pozzi and Schivardi (2016) estimate idiosyncratic demand shocks as residuals of the demand function.

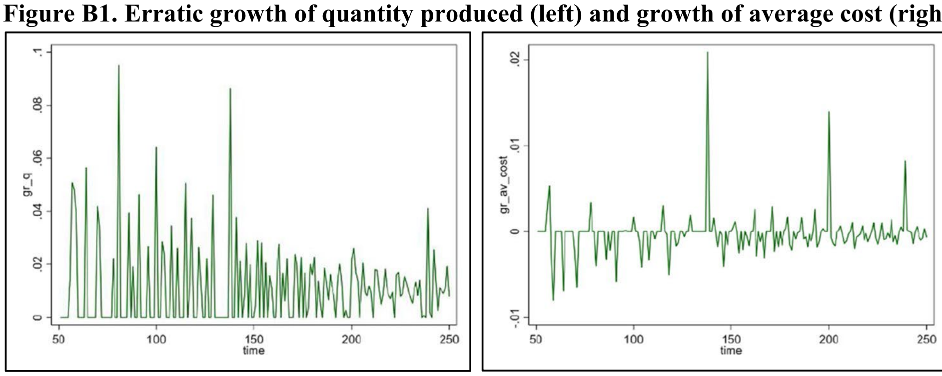
[6] Stagnation here need not correspond to remaining at 100% capacity, but falling back below 100% capacity, because many random shocks (e.g. customers leave, employees leave or fall ill, machines cannot always be used because they need to be repaired) would make it unlikely that the firm can permanently stay close to the 100% boundary.

BOX 1: Non-constant costs of growth

We revisit the simulation data in Coad and Planck (2012, p198ff), and extend it to obtain some novel results. A firm can produce a quantity of output Q that requires 5 inputs: raw materials, production employees, machines, supervisors, and production plants. Integer restrictions affect all of these inputs (e.g. one cannot grow by adding ½ a machine or $1/5$ of a supervisor), and each input has its fixed price (see Coad and Planck 2012 Table 1 for details). To keep things simple, demand = price × Q = $\beta - \varphi Q$, where $\varphi = 0.5$, and the demand parameter β starts at 100 and grows 1% in each period. In this setup, the integer restrictions affecting the inputs lead to erratic growth dynamics of total quantity produced, and changes in average cost. We focus on changes in average cost (total cost / quantity produced), rather than marginal cost, because the latter is particularly problematic: at any point in time, the marginal cost of producing the infra-marginal q^{th} unit is different from that of producing the $q+1^{th}$ unit, which in turn is different from the marginal cost of producing the $q+2^{th}$ unit, and so on. It is non-trivial to compare the marginal cost for e.g. cases where a firm grows by 2 units (at time t), then by 10 units (at time t+1), then by zero units at time t+2. Focusing on the average cost alleviates slightly these difficulties.

Figure B1 shows the growth rates of quantity and average cost, where growth rates are calculated using log-differences, and where data for the first 50 periods (initialisation) are discarded. Notice that, while demand is growing steadily and smoothly at 1% per period, the firm's growth of quantity produced is erratic (Figure B1, left) as a result of the idiosyncratic time-varying degree of capacity utilisation reflected in the erratic changes in average cost (Figure B1, right). Figure B1 (right) shows that, while most of the changes in average cost are negative (i.e. cost reductions through economies of scale), nevertheless sometimes the average cost jumps upwards, corresponding to broad-based expansionary investment events such as building a new production plant.

Figure B1. Erratic growth of quantity produced (left) and growth of average cost (right).



Source: Authors' calculations based on the Coad and Planck (2012, p198ff) simulation data.

our knowledge regarding empirical evidence on how capacity utilisation affects firm growth (partly due to a lack of high-quality data). We contribute novel theory and evidence with survey data from a large number of European countries.

Investigating this area seems worthwhile from a micro-economic perspective, because many growing firms are likely to find themselves running at over-capacity, and better understanding how to overcome these capacity-related challenges will enable firms to continue growing. The topic also seems interesting from a macro-economic perspective: If more firms are able to overcome episodes of high capacity utilisation, this could potentially result in more employment, more value added, more innovations, aggregate productivity growth through reallocation towards higher-productivity growing firms, etc. From a policy perspective, if timely support is available to firms being dragged over the jagged rocks of an uneven terrain of production possibilities, then HGEs can survive and thrive in order to realize potential macroeconomic benefits.

We therefore state our hypothesis hence:

> Fork in the road hypothesis: When firms reach a critically high state of capacity utilisation, some firms will respond by investing massively and launching into subsequent growth, while other firms will shrink back operations to a more comfortable level given the current capacity constraints. Given this heterogeneity, there is a particularly large variation in the growth rates taken by firms that reach a critically high state of capacity utilisation.

2.5 Capacity constraints and financial performance

Previous literature on capacity utilisation rates suggested that firms far below maximum capacity have low performance in terms of productivity and profits (e.g. Shen & Chen, 2017). However, when firms start operating above maximum capacity, there are rising costs such as "time compression diseconomies", costs for hiring additional machinery at suboptimal conditions and overtime pay. Operating above what is normally considered to be maximum capacity can also distort the use of inputs away from the optimal mix, to decrease reliance on inputs whose scale is fixed (e.g. heavy machinery), and to increase reliance on inputs that are more flexible in the short run (e.g. low-wage employment, intermediate inputs), which may increase the unit production costs as firms move above maximum capacity utilisation (ECB, 2007).[7]

Figure 3 illustrates these effects. Operating costs do not start from zero, because there are fixed costs (e.g. costs of keeping machines unused) that arise even when zero units are produced. Operating costs rise slowly until the point of maximum capacity, where they start rising exponentially. While profits (i.e. revenue minus costs) are higher for firms at maximum capacity compared to firms below maximum capacity, nevertheless, it is not entirely clear whether firms operating at a scale slightly above maximum capacity can obtain higher profits than those at maximum capacity. It is possible that the rising costs due to operating at overcapacity are insufficient (at least initially) to offset the rising revenues from selling additional units. Comparing the profits of these two categories (above maximum capacity vs at maximum capacity) is therefore an empirical question.

2.6 Measuring capacity utilisation: insights from previous literature

A key concept in our analysis is the degree of capacity utilisation at the firm. While it may be optimal to use less than the totality of installed capacity (Pozzi & Schivardi, 2016), especially if demand is volatile (Butters, 2020), nevertheless, a degree of capacity utilisation that is too low is inefficient in the sense that installed capacity is unused and may lead to extra maintenance costs, and these productive resources could be better used elsewhere in the economy. Low capacity utilisation generally corresponds to low productivity, because capacity is being used inefficiently (Butters, 2020). Capacity utilisation varies over the business cycle, with low capacity utilisation common in recessions (Baldwin et al., 2013; Basu, 1996). Low capacity utilisation may be advantageous, though, in the context of monopoly or oligopoly, because it can function as an entry deterrent since the excess capacity would mean that entrants would have difficulties being profitable (Nikiforos, 2013).

Overcapacity and capacity utilisation mean different things in different studies. The term "overcapacity" often refers to an excess of productive capacity (usually at an industry level) needed to satisfy the corresponding industry-level demand (Henderson & Cool, 2003), often in a context where "bandwagon" effects of imitating rivals (Henderson & Cool, 2003) or government subsidies (Liu et al., 2019; Zhang et al., 2016) can distort the signals regarding how much capacity is needed in the industry. In this vein, for example Shen and Chen (2017) investigate "overcapacity" in the sense that China's manufacturing industries (e.g. steel, coal, cement, glass) have excess production capacity, at an industry level, to satisfy the needs of consumers. In their case, the authors recommend that the appropriate response would be the elimination of surplus capacity via the exit of unviable "zombie" firms. Basu (1996) takes the growth of materials as a proxy for capacity utilisation and finds

[7] There is some evidence that higher capacity utilisation, due to rising demand, translates into firms increasing their prices and also their profit margins (ECB, 2007). We do not have data on firms' prices so we cannot investigate this.

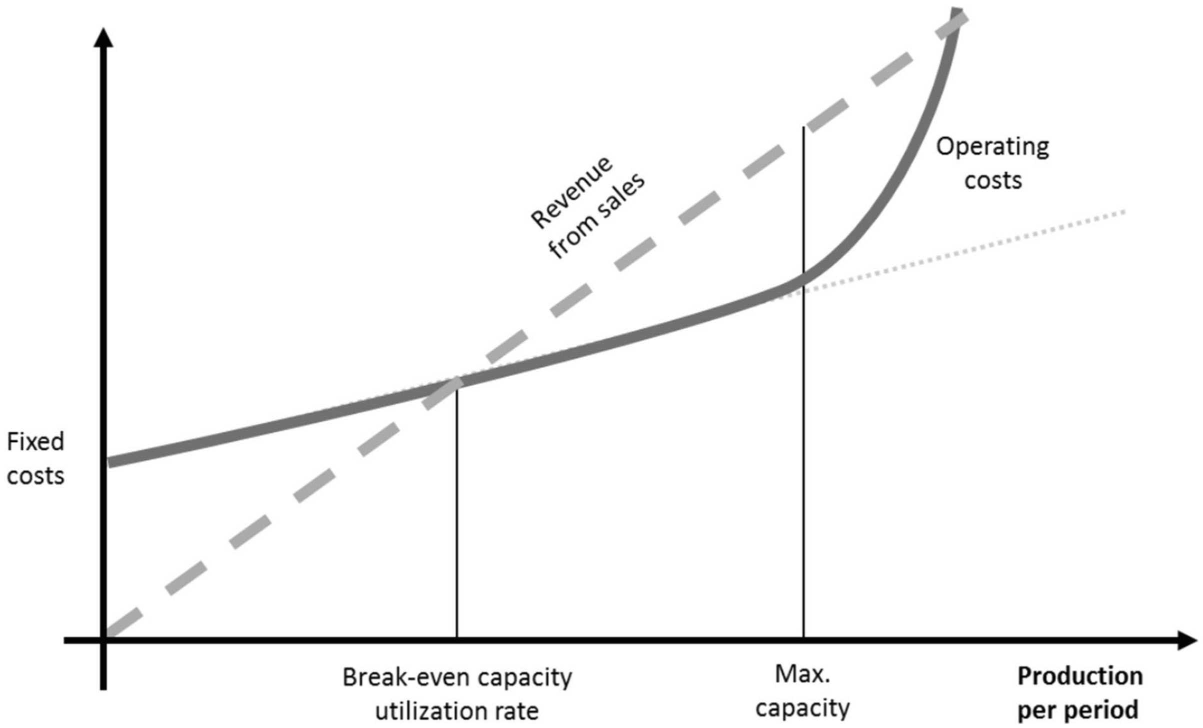


Fig. 3 Conceptual diagram of how profits are expected to vary with capacity utilisation per period. Red solid line: operating costs. Blue dashed line: revenue from sales

that variations in capacity utilisation help explain the procyclical nature of productivity over the business cycle.

This paper focuses on capacity utilisation at the firm level. Capacity utilisation has traditionally focused on physical capital (Berndt & Morrison, 1981), although capacity utilisation considerations can also be applied to other domains such as human resources and managerial attention (Ocasio, 1997). Firm-level capacity utilisation is easier to measure in some industries than others, depending on industry characteristics. In the airline industry, for example capacity utilisation is measured using the "load factor", i.e. the share of occupied seats relative to total seat miles flown (e.g. Baltagi et al., 1998; Butters, 2020; Dana & Orlov, 2014).[8] Baltagi et al. (1998) observe that deregulation helped US airlines to improve their degree of capacity utilisation, moving from an average load factor of 0.50 to 0.61 when comparing the post-deregulation period 1990–1994 with the period 20 years earlier. Dana and Orlov (2014) found that the US airline load factors increased from 62% in 1993 to 80% in 2007, with this increase being partly explained by increased online reservations from growing internet penetration.

Some studies have developed ways of measuring capacity utilisation that can be applied to a broader range of sectors. For example, Pozzi and Schivardi (2016) develop an indicator of capacity utilisation in the context of a structural economic model. Baldwin et al. (2013) take the ratio of capital income to value added as a proxy for capacity utilisation, based on the reasoning that variation in the ex post return to capital reflects variation in capacity utilisation.

Survey data on capacity utilisation is rare. Pozzi and Schivardi (2016), however, analyze such data in their structural estimation of firm growth dynamics. They focus on Italian firms in the textile and leather, metals and machinery sectors, because the market structure in these sectors is expected to be relatively close to monopolistic competition. The

[8] However, more technical indicators of capacity utilisation have been put forward for airlines, and they do not always closely correspond to the load factor (Baltagi et al., 1998).

focus of their paper is showing that demand shocks can more freely translate into firm growth than TFP (total factor productivity) shocks, because frictions (such as management capabilities and workforce human capital) hinder firms' abilities to draw upon TFP growth as a trigger for subsequent firm growth. Their survey data refers to "the maximum output that can be obtained using the plants at full capacity, without changing the organisation of the work shifts". They find (p614) that "the average degree of capacity utilisation is 81%, with a standard deviation of 13%; the 5th and the 95th percentile are 60% and 98%, respectively".

Interestingly, Pozzi and Schivardi (2016) observe the considerable variation of capacity utilisation within firms over time (p614): "the utilized capital variable displays plenty of within-firm variation: a variance decomposition reveals that within variation represents between 83% (for the textile sector) and 92% (for the metals sector) of the between variation". This latter finding is compelling in the context of HGE prediction, because, while most predictor variables used in HGE prediction are relatively time invariant and stable (e.g. founder's education, founder's previous business experience, geographic region, sector, legal form), high-growth status itself is episodic and short lived. As a result, various scholars have highlighted the need for time-varying explanatory variables when predicting firm growth (Coad & Srhoj, 2020; Geroski & Gugler, 2004; Storey, 2011).

Similar in meaning to the concept of capacity utilisation is the concept of "slack" resources. Cohen et al., (1972, p. 12) write: "Slack is the difference between the resources of the organization and the combination of demands made on it". George (2005, p664) pens a similar definition: "I introduce transient slack, defining it as excess resources available after resource demands for operations have been met". Slack resources may conceptually correspond to machinery or employees that are not being used to their full potential. Nohria and Gulati (1996) develop a composite measure of slack (based on working time and department's operating budget) and find a U-shaped relationship with innovation: some slack is good, but too much reduces innovation outcomes. Most of the previous literature, however, has operationalized slack resources in terms of financial slack (defined in terms of financial resources and financial demands on these resources). This is probably because non-financial resources are hard to measure and quantify. A potential drawback, though, of measuring slack in such financial terms is that highly productive firms (with considerable free cash flow) selling their products at high prices could be considered to have slack even if they are operating their production inputs at a lean and highly efficient level of capacity utilisation. One could therefore prefer concepts such as "productive slack" or "operational slack" instead of "financial slack" as it is often measured in the literature. However, in this paper, we prefer the terms "capacity utilisation" and "overcapacity".

A major problem for empirical work into capacity utilisation is that most datasets do not usually contain information on the level of capacity utilisation or the configuration of discrete resources within firms, despite their central importance in Penrose's theory of firm growth. For example, many administrative datasets collected by national statistical offices focus more on variables related to taxes and social security obligations and easily observed administrative variables (address, sector of activity, legal form). Instead, we investigate a rich data source that asks firms about their degree of capacity utilisation, to test our conjectures about capacity utilisation and high growth.

3 Data

3.1 Data description

Our analysis is based on the EIBIS dataset merged with the Bureau van Dijk ORBIS database. EIBIS is an EU-wide survey that gathers qualitative and quantitative information on investment activities by non-financial corporates, both SMEs (5–250 employees) and larger corporates (250 + employees), their financing requirements and the difficulties they face. Using stratified sampling, EIBIS aims to be representative across all 27 Member States of the EU, the UK and the USA, within countries, four firm size classes (micro, small, medium and large) and four sector groupings (manufacturing, services, construction and infrastructure). The survey is carried out through telephone (CATI) interviews in the local language. All interviewed firms are drawn from the BvD ORBIS database, which allows the survey answers to be linked to firms' financials and other administrative information, although firm information remains

Table 1 Capacity utilisation

	2016		2017		2018		2019		Total
Above maximum capacity	579	4.72%	590	4.85%	809	6.64%	828	6.22%	2,806
At maximum capacity	5766	47.04%	5769	47.45%	5644	46.30%	6897	51.85%	24,076
Somewhat below full capacity	4563	37.23%	4504	37.05%	4666	38.28%	4643	34.90%	18,376
Substantially below full capacity	1349	11.01%	1295	10.65%	1070	8.78%	935	7.03%	4649
Total	12,257	100.00%	12,158	100.00%	12,189	100.00%	13,303	100.00%	49,907

Columns contain frequencies (# of firms) and percentage shares for each year.

Table 2 Frequency of overcapacity status

		Years in overcapacity					
		0	1	2	3	4	Total
Number of times surveyed	1	22,571	1345				23,916
	2	5782	608	54			6444
	3	2409	329	49	10		2797
	4	1135	182	36	10	1	1364
	Total	31,897	2464	139	20	1	34,521
	Panel firms only	9326	1119	139	20	1	10,605

anonymous. Detailed methodology on the survey is available from IPSOS.[9] EIBIS has been shown to be a reliable data source with no systematic sampling bias (Brutscher et al., 2020).

We use four waves of the EIBIS survey (2016–2019) with information on over 34,500 firms with 50,651 observations. The panel structure of the survey is presented in Table OSM-1.1. Out of the 34,521 firms, 69% are surveyed once, 19% surveyed twice, 8% three times and 4% of the firms are surveyed in all four waves. All in all, 10,600 firms are surveyed multiple times. Additionally, we merge the EIBIS survey with financial variables derived from ORBIS (2014–2018). We use 2 years prior to the survey year to construct the change in variables. The length of the panel information is considerably increased based on the variables derived from ORBIS. Panel information of the financial variables is available for longer periods even for firms surveyed once such as 90% of firms have ORBIS-derived financial variables for at least two times and 73% of them (more than 17,000 firms) are observed in ORBIS five times or more. Still, few firms (1749) do not have financial variables derived from ORBIS for our sample of 2013–2018. The joint structure is described in Table OSM-1.2 for the period 2013–2018. The full coverage of ORBIS-derived variables is until 2017 due to a lag in data provision to ORBIS and this explains the drop from 28,596 in 2017 to 10,525 firms in 2018. Overall, 99,650 observations are available for firms for 2013–2018, including 2 years when they are not participating in the survey (see Table OSM-1.3 for a summary of the merged EIBIS-ORBIS dataset).

3.2 Variables definition and descriptive analysis

The main dependent variables of this paper are capacity utilisation and firm growth. Capacity utilisation corresponds to whether the firm operated in the latest financial year "at its maximum capacity attainable under normal conditions", with four responses: above maximum capacity; at maximum capacity; somewhat below full capacity and far below full capacity, as shown in Table 1.[10] In what follows, for conciseness,

[9] https://www.eib.org/attachments/eibis-methodology-report-2019-en.pdf

[10] By normal condition, it is meant the firm general practices regarding the utilisation of machines and equipment, overtime, work shifts, holidays, etc.

we sometimes refer to the category of "above maximum capacity" as "overcapacity".

In each year, out of the four categories, the least populated is "above maximum capacity", which varies from 4.6% in 2016 to 6.6% in 2018. The most populated category is "at maximum capacity" which corresponds to about 50% of responses. The category of firms "substantially below full capacity" corresponds to about 7%–11% of firms in each year. It is hard to imagine that this category corresponds to firms engaging in deliberate and rational decisions to operate far below capacity.[11] In line with previous literature, our analysis indicates that these firms do not make full use of their production capacity, as this state is associated with low productivity and poor financial performance, see Appendix OSM-1. Such firms could correspond to "zombie firms" (Shen & Chen, 2017).

Table 2 presents the frequency of overcapacity status. Fewer than 10% of years correspond to being in a state of overcapacity. Spending 2 years in overcapacity is very rare. Being in a state of overcapacity therefore appears to be a one-off event, rather than an enduring state. Figure 7 in the Appendix shows how capacity utilisation categories vary across countries.

Of central interest to our paper is the phenomenon of firm growth, which is measured either using annual growth rates in the years after capacity utilisation states or in terms of growth over a 3-year period (which corresponds to the usual definition of HGEs). The HGE dummy has the advantage of providing a simple binary indicator of whether a firm is an HGE, whereas annual growth rates have the advantage of providing richer information on the distribution of outcomes (since growth rates are continuous variables) as well as providing a finer-grained information (growth is calculated each year instead of over a 3-year period). These two indicators of firm growth (HGE dummy vs annual growth rates) are therefore seen as complementary.

[11] One possible motivation for operating far below the maximum capacity utilisation rates is to provide a credible threat of a price war to potential entrants (Nikiforos, 2013). This entry deterrent motivation could perhaps explain some, but probably not all, of the cases of firms that are operating far below maximum capacity levels.

When firm growth is measured over 1 year, the log difference is the preferred way to calculate a growth rate (Coad, 2009; Tornqvist et al., 1985). Growth of X, where $X \in \{Sales, Employment, Profit\}$, for firm i at time t, is calculated as:

$$GR\_X_{i,t} = \log\left(X_{i,t}\right) - \log(X_{i,t-1})$$

We measure HGEs by using the cumulative 3-year growth rate of more than 33% HGFs (from t-3 to t). More exactly, our measurement corresponds to the standard OECD-Eurostat definition of HGEs (Petersen and Ahmad, 2007): an enterprise with an average annualized turnover or employment growth greater than 10% (or alternatively 20%) per year over the past 3 years and having 10+employees at the beginning of the growth period. We use this approach in order to have a heterogeneous focus across size groups (Ferrando et. al, 2019).

Our HGE measurement relies on the EIBIS data, which collected in each year the information on the number of employees both in the current year and 3 years prior. In this way, even for firms with limited panel information, a 3-year growth can be calculated for the whole EIBIS sample period.[12]

Control variables include investment growth and details on the type of investments. We use alternative measurements of investment growth derived both from ORBIS (where investment is defined as a percentage change in fixed assets) and from EIBIS (as firms provide information on the total amount of investment). Nevertheless, two consecutive survey responses are needed to calculate investment growth, which limits considerably the number of observations. Growth rates are winsorized at the 5% and 95% levels. As an alternative way to capture the investment dynamic, a survey variable is used on investment compared to t-1 with three possible answers: above, the same or less than in the previous year. This later measurement, however, has the limitation of showing just the direction of change without indicating the size of the change.

We control also for investment types, by relying on the EIBIS variables of investments in (1) land, business buildings and infrastructure; (2) machinery and equipment; (3) research and development (including

[12] ORBIS financial data are available with a 1-year lag compared to the EIBIS.

Table 3 Transition matrix

Frequencies (# of firms) and percentage shares. For brevity, “above max.”, “at max.”, “bit below max.” and “far below max.” correspond to the categories “above maximum capacity”, “at maximum capacity”, “somewhat below full capacity” and “substantially below full capacity”, respectively.

		Capacity at $t+1$				
		Above max	At max	Bit below max	Far below max	Row total
Capacity at t	Above max	143	463	172	32	810
		17.65%	57.16%	21.23%	3.95%	
	At max	480	4513	1751	191	6935
		6.92%	65.08%	25.25%	2.75%	
	Bit below max	199	1811	3285	486	5781
		3.44%	31.33%	56.82%	8.41%	
	Far below max	32	233	624	622	1511
		2.12%	15.42%	41.30%	41.16%	
	Column total	860	7089	5889	1339	15,177
		5.67%	46.71%	38.80%	8.82%	

the acquisition of intellectual property); (4) software, data, IT networks and website activities; (5) training of employees and (6) organisation and business process improvements. Alternatively, we control for investment purposes, such as (1) investment for replacement, (2) for expanding capacity and (3) for developing new products, processes or services.

We consider also the internationalisation activity of the company, captured by two alternative dummy variables for (1) whether the firm is exporting directly and (2) whether the firm invested in another country. Moreover, our estimations include also a dummy variable that captures whether the firm is a subsidiary of another firm or if it is an independent company.

We control also for firm’s performance/profitability, captured by three alternative categorical variables: (1) the firm is generating (a) loss, (b) profit or (c) is at break-even point; (2) profit before tax as a share of turnover being in five different categories from below 2% to above 15%; and (3) business prospects with three alternative responses for (a) improving, (b) staying the same or (c) deteriorating.

We check also the innovativeness of the company, defined according to the introduction of new products, processes or services as being either (1) globally new, (2) new for the country or (3) those less radical innovators with products new only to the company. Additionally, a digitalisation dummy is derived from EIBIS according to the adoption of any of the listed new digital technologies. These include 3-D printing, robotics, big data and analytics, virtual reality, the internet of things, platform technologies and drones. This variable equals one if the firm either has implemented partially or organized the entire business around any of the technologies listed. This variable is only available for the year 2018 from EIBIS 2019.

As additional explanatory variables, dummy variables are created on financing from grants[13] and whether the firm had an energy audit in the last 3 years.

It is well known that firm size is a major predictor of firm behaviour and performance; therefore, we control for firm size in our regressions. This is done by controlling for log_sales_{it} as well as the quadratic term $log\_sales_{it}{}^2$, to allow for a potentially non-linear influence of size (e.g. if size mainly affects performance when firms are below a critical size threshold (Sutton, 1997). Additionally, dummies for year, country and sector are also used as control variables.

Regarding the sector of activity, the EIBIS survey asks respondents about the “main sector of activity of this company”. Almost 30% of respondents are in manufacturing. Both the construction sector and the wholesale and retail trade sector account for over 20% of respondents each.

Summary statistics of these variables appear in Appendix Tables OSM-2 and OSM-3.

3.3 Methodology

Considering the paucity of previous research in this area, we describe a relatively unfamiliar phenomenon

[13] From the survey, grants could come from various sources, including “financial support or subsidies from regional and national government and funding provided by the European Commission”.

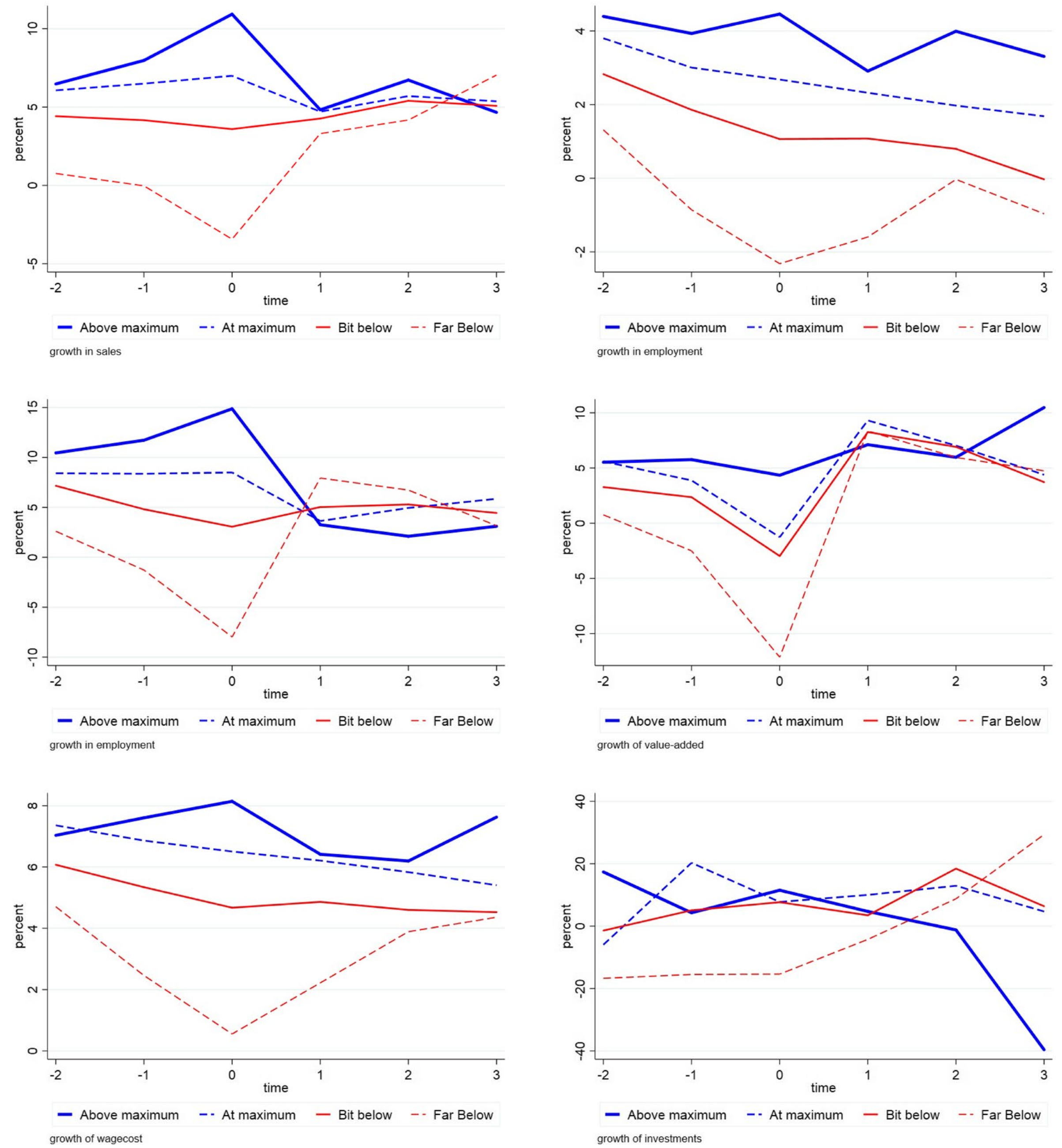


Fig. 4 Event history time-series plots. Mean growth rates for various capacity utilisation categories. Overcapacity is measured at time t=0. Growth rates of 6 variables: sales growth (top left), employment growth (top right), profits growth (centre left), value added growth (centre right), wage growth (bottom left), and growth of total investment (bottom right)

using descriptive techniques in an exploratory way. A variety of techniques are applied, with many results in the Online Supplementary Materials. We begin with descriptive statistics, to investigate the characteristics of firms in our sample, the frequencies of firms in capacity utilisation categories, and to see how these frequencies vary across EU member states. Transition matrices show the dynamics of entry and exit from various capacity utilisation states. Time-series plots show the dynamics of firms in the years before and after capacity

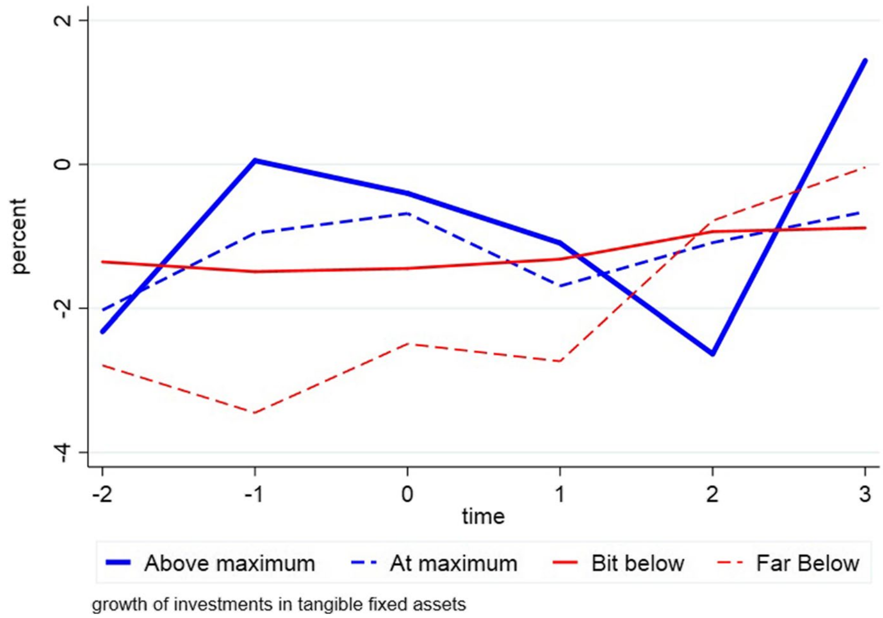


Fig. 5 Event history time-series plot. Mean growth rates for various overcapacity categories. Overcapacity is measured at time t=0. Growth rates of investment in fixed assets

utilisation states, for various indicators such as sales growth, profits growth and employment growth. Multivariate regressions can control for potentially confounding background factors to provide a clearer view of the relationships between capacity utilisation and growth. In addition, quantile regressions (Koenker & Bassett, 1978) explore heterogeneous responses in terms of growth paths after overcapacity, allowing us in particular to evaluate the fork in the road hypothesis.

4 Analysis

4.1 How do firms transition into and out of capacity utilisation states?

Table 3 shows the transition matrix for capacity categories. For those at maximum capacity or a bit below maximum capacity, at time *t*, these firms are most likely to stay in these same categories in $t+1$ (i.e. to remain positioned along the diagonal). In contrast, those operating above maximum capacity at time (*t*) are relatively unlikely to remain in the same category the next year but instead are more likely to operate at maximum capacity at time ($t+1$). This raises the question of whether those firms transitioning from above maximum capacity to at maximum capacity (about 3% of firms in total) increase their capacity or whether they maintain their original capacity levels while reducing output accordingly.

4.2 Dynamics of entering into and leaving overcapacity

Figure 4 shows the evolution of key variables relating to mean[14] firm growth and performance in the years before and after particular capacity utilisation states at time t=0. We focus on the years up to $t+3$, which aligns with standard indicators of HGEs that measure growth over a three-year period.

[14] Figures 4 and 5 show the results for mean growth rates in capacity utilisation categories. In our baseline graphs, presented here, we prefer the mean to the median, because the median employment growth in many cases is precisely 0.0000, due to integer restrictions in employee headcounts. However, the median may be of interest, because it is less sensitive to outliers than the mean (remember, though, that our growth rates variables have been already been winsorized to remove outliers; hence, sensitivity to outliers is less of a concern). Appendix OSM-5 contains the corresponding graphs when the median growth rate is taken, instead of the mean. Appendix OSM-5 also shows the corresponding graphs when the growth rate variables are “cleaned” or pre-processed via OLS regression to remove the potential influence of size, country, sector and year components. The results are overall similar, although a major difference concerns the growth of total investment at time $t=3$ for firms above maximum capacity (therefore, we prefer not to make any strong interpretations of the growth of total investment for overcapacity firms at time $t=3$).

Figure 4 shows many interesting findings. First, there is a rapid growth of sales, employment and profits in the years before being “above max capacity”, providing further evidence that entering a state of overcapacity is a rather fortunate state of affairs, resulting from rising sales, employment and profits in the preceding years. Profits growth is relatively high for overcapacity firms in the years before/during being “above max capacity”, although this quickly levels off to give a mediocre profit growth in the subsequent years.

Second, employment growth of “above maximum capacity” firms is higher than that of all other categories (before and after), at all periods. This suggests that firms operating at overcapacity are major job creators. Relatedly, wage growth (Fig. 4, bottom left) is always highest in the category of firms that are “above maximum capacity” at time *t*, both before and after being at overcapacity. There are various interpretations of this wage growth: overcapacity firms might be paying some kind of “overtime” premium or seeking to increase the skills of the workforce—but, the finding that overcapacity firms are creating more jobs and, relatively better-paying jobs, could make them a thought-provoking category for policymakers.

Third, the evolution of these variables allows some cautious speculation about the causal ordering of the variables. One possible interpretation could be that sales growth leads to employment growth as firms push back against overcapacity. This is consistent with vector autoregression evidence that sales growth causes employment growth (Moneta et al., 2013). It is also consistent with suggestions of the key role of rising demand on the capacity choices of new firms (Foster et al., 2016; Nikiforos, 2013).

Fourth, some variables have a mediocre performance in the years after overcapacity status. This could signal that reaching overcapacity is the crest of a wave of rising demand that cannot be sustained for long. For profit growth and total investment growth, the growth rates end up being the lowest among all categories a few years after the episode of being “above maximum capacity”. This could be because such firms undertook their investment in previous years instead. Indeed, growth of total investment (Fig. 4, bottom right) is relatively high (but not the highest) for those in the category “above maximum capacity”, in the years beforehand but drops off sharply afterwards. This drop in investment can be explored further, to see which type of investment is cut back. Figure 5 shows that the average growth of investment in fixed assets is relatively high (for overcapacity firms) during the year of overcapacity, then drops in the following 2 years, but picks up again to reach a high value for the period 3 years after the overcapacity event. This suggests that spending time in overcapacity is associated with high growth of investment in tangible fixed assets at the time of overcapacity, as well as 3 years later.[15] The sharp decrease in total investment observed in Fig. 4 (bottom right) therefore corresponds to investment types that are not tangible fixed assets (e.g. a drop in investment in intangibles or R&D), although the evidence so far is not conclusive because these unconditional line plots could be affected by confounders such as firm size and sector (this is explored below).

Regarding firm behaviour at the time of being at overcapacity, we earlier conjectured that firms above maximum capacity face a fork in the road either they invest massively for subsequent growth or scale back to stay within existing capacity limits. This would be reflected in a higher variation in growth rates for firms above maximum capacity: while some take the opportunity to launch into rapid growth, others cut back. Appendix OSM-10 investigates this conjecture and finds some interesting results for the investment. For firms above maximum capacity, there is a sharp increase in the variation in growth rates for total investment (i.e. all types of investment) while there is a clear decrease in the variation in growth rates of investment in tangible fixed assets. This could be taken as evidence that firms at a period of overcapacity will subsequently embark on a relatively homogenous strategy of high investment in tangible fixed assets (i.e. relatively high growth rates but with low variation in growth rates across firms), while such firms have high variation in their growth rates for investment in intangible assets, with some overcapacity firms investing heavily in intangibles while others neglect this area.

[15] We could speculate that this corresponds to a new wave of “replacement” or “expansionary” investment in fixed assets to follow through in alleviating previous capacity constraints.

Table 4 MNL (multinomial logit) regressions for the different capacity utilisation categories

	Overcapacity	At_max_cap	Bit_below	Overcapacity	At_max_cap	Bit_below
HGE dummy	0.974***	0.576***	0.329***	0.707***	0.370***	0.165
	[0.097]	[0.079]	[0.079]	[0.166]	[0.137]	[0.137]
log_sales	0.751***	0.501***	0.388***	0.309	0.234	0.287
	[0.167]	[0.096]	[0.095]	[0.319]	[0.221]	[0.217]
log_sales_sq	-0.021***	-0.011***	-0.009***	-0.008	-0.002	-0.005
	[0.005]	[0.003]	[0.003]	[0.010]	[0.007]	[0.007]
Profit status: loss	-1.796***	-1.689***	-1.158***	-1.916***	-1.632***	-1.223***
	[0.096]	[0.049]	[0.047]	[0.186]	[0.101]	[0.093]
Profit status: break-even	-1.088***	-0.928***	-0.562***	-	-	-
	[0.111]	[0.064]	[0.063]			
log_wagebill	0.075**	0.063***	0.059**	0.049	0.061	0.037
	[0.036]	[0.024]	[0.024]	[0.072]	[0.054]	[0.053]
Age: 2 ≤ age < 5 years	0.266	0.219	0.647	-0.176	0.028	0.003
	[0.559]	[0.387]	[0.403]	[1.328]	[1.146]	[1.141]
Age: 5 ≤ age < 10 years	0.040	0.086	0.401	-0.245	-0.000	-0.110
	[0.546]	[0.378]	[0.394]	[1.305]	[1.133]	[1.128]
Age: 10 ≤ age < 20 years	0.058	0.021	0.388	-0.438	0.103	0.013
	[0.542]	[0.375]	[0.391]	[1.300]	[1.129]	[1.124]
Age: 20+ years	-0.235	-0.186	0.275	-0.509	-0.102	-0.096
	[0.540]	[0.374]	[0.390]	[1.296]	[1.127]	[1.121]
Subsidiary	0.095	0.002	0.012	0.183	-0.031	-0.006
	[0.073]	[0.052]	[0.052]	[0.132]	[0.099]	[0.097]
Directly exported	-0.138**	-0.301***	-0.106**	-0.129	-0.207*	-0.115
	[0.069]	[0.047]	[0.047]	[0.140]	[0.106]	[0.105]
Invested in another country	-0.198*	-0.315***	-0.188**	-0.232	-0.334***	-0.179
	[0.112]	[0.080]	[0.078]	[0.173]	[0.128]	[0.124]
R&D investment dummy				-0.025**	-0.018**	-0.013
				[0.011]	[0.008]	[0.008]
IT investment dummy				-0.009	-0.014	-0.007
				[0.014]	[0.010]	[0.010]
Training investment dummy				0.024	0.014	0.008
				[0.016]	[0.011]	[0.011]
Business processes inv. dummy				0.006	-0.009	0.003
				[0.011]	[0.008]	[0.008]
Replacing capacity - investment (%)				0.005	0.001	0.003
				[0.004]	[0.003]	[0.003]
Expanding capacity - inv. (%)				0.011**	0.007**	0.007**
				[0.004]	[0.003]	[0.003]
New prod./process inv. (%)				0.000	-0.000	-0.000
				[0.004]	[0.003]	[0.003]
Investment wrt t-1: broadly stayed the same				-0.398***	-0.225**	-0.196**
				[0.124]	[0.093]	[0.092]
Investment wrt t-1: less than prev. year				-0.855***	-0.805***	-0.409***
				[0.164]	[0.111]	[0.107]
Innov: new to country				0.105	-0.079	0.023
				[0.163]	[0.119]	[0.117]

Table 4 (continued)

	Overcapacity	At_max_cap	Bit_below	Overcapacity	At_max_cap	Bit_below
Innov: new to global mkt				0.029	-0.227**	-0.251**
				[0.143]	[0.106]	[0.103]
Business prospects: same				0.065	0.279***	0.135
				[0.126]	[0.094]	[0.093]
Business prospects: deteriorate				-0.420**	-0.505***	-0.262**
				[0.171]	[0.119]	[0.115]
Profits before tax: 2% to 4%				0.373**	0.238**	0.118
				[0.178]	[0.121]	[0.118]
Profits before tax: 5% to 9%				0.368**	0.294**	0.115
				[0.178]	[0.122]	[0.118]
Profits before tax: 10% to 14%				0.591***	0.355**	0.083
				[0.205]	[0.146]	[0.143]
Profits before tax: 15% or more				0.595***	0.221	-0.225
				[0.204]	[0.145]	[0.142]
Observations	37,415	37,415	37,415	10,089	10,089	10,089

***p < 0.01, **p < 0.05, * p < 0.1. Standard errors are clustered at the firm level. Omitted baseline reference case for the dependent variable is "far below full capacity". Constant term, country, sector and year dummies are included in all of the estimations but not reported. In specification (2), "profit status: breakeven" is automatically dropped from the estimations, presumably due to the reduced number of observations.

4.3 Which factors are associated with entering into overcapacity?

Table 4 investigates the determinants of different capacity utilisation categories using multinomial logistic regressions, thereby considering the various capacity utilisation categories in an integrated econometric framework. The dependent variable corresponds to the four capacity utilisation categories, with "far below full capacity" as the omitted baseline case. Given the trade-off between a richer set of explanatory variables, at the cost of a smaller number of observations (because the explanatory variables are unequally affected by missing values), Table 4 presents two regression specifications. Regression model (2) has more explanatory variables and fewer observations than model (1).[16]

Table 4 shows that rapid growth contributes to overcapacity, because HGEs (where rapid growth is measured from *t-3:t*) are more likely to find themselves in the category of "above maximum capacity" at time t and to a lesser extent for the category "at maximum capacity".

Table 4 also shows that profitable firms are more likely to be "above maximum capacity". Making a loss is negatively related to being "above maximum capacity" or "at maximum capacity" but is positively associated with being "far below capacity".[17] Furthermore, firms above maximum capacity are less likely to be R&D investors and more likely to invest for motivations of "expanding capacity".

4.4 What routes do firms take out of overcapacity?

Table 5 investigates which variables are associated with firm growth in the following period. In line with the event study line plots in Fig. 4, being above maximum capacity is related to subsequent employment growth, but there is no statistical relationship with subsequent growth of sales or profits. Some possible explanations for why employment growth occurs later than sales and profits growth could be the following.

[16] It is possible that the sample used in model (2) is not representative of the sample used in model (1), due to sample selection effects. This could be one explanation why the coefficients for log(wage bill) are never significant in model (2), while they are significant in model (1). However, an alternative explanation for the change in coefficients on log(wage bill) from (1) to (2) could be due to mild multicollinearity with the newly entered explanatory variables.

[17] Regressions with full country dummies coefficients available upon request.

Table 5 Regression results. Dependent variables: growth ($t+1$) of sales, employment and profits

	f1_g_sales	f1_g_sales	f1_g_sales	f1_g_sales	f1_g_empl	f1_g_empl	f1_g_empl	f1_g_empl	f1_g_profit	f1_g_profit	f1_g_profit	f1_g_profit
State-of-art machinery/ equipmt	0.011**	0.014***	0.012**	0.005	0.018***	0.017***	0.014***	0.011	-0.003	0.002	0.005	0.017
	[0.005]	[0.005]	[0.005]	[0.009]	[0.004]	[0.004]	[0.004]	[0.007]	[0.016]	[0.016]	[0.016]	[0.028]
Innov: new to the firm	-0.011*	-0.009	-0.002	-	-0.003	-0.003	-0.001	-	-0.005	-0.007	-0.009	-
	[0.006]	[0.006]	[0.006]		[0.004]	[0.004]	[0.005]		[0.019]	[0.019]	[0.019]	
Innov: new to country	-0.003	-0.002	0.005	-0.001	0.001	0.001	0.003	0.003	0.019	0.018	0.014	-0.017
	[0.006]	[0.006]	[0.006]	[0.008]	[0.005]	[0.005]	[0.005]	[0.006]	[0.019]	[0.019]	[0.019]	[0.023]
Innov: new to global mkt	-0.002	-0.002	0.007	-	0.002	0.002	0.005	-	0.003	0.002	0.012	-
	[0.008]	[0.008]	[0.008]		[0.006]	[0.006]	[0.006]		[0.026]	[0.026]	[0.027]	
R&D investment (IHS)	0.025	0.071**	0.029	-8.685***	0.048**	0.048**	0.045*	-0.37	-0.117	-0.067	-0.111	1.955
	[0.031]	[0.031]	[0.032]	[3.211]	[0.023]	[0.023]	[0.024]	[2.471]	[0.096]	[0.098]	[0.099]	[9.915]
Above maximum capacity	0.002	0.01	0.008	-0.01	0.039	0.039***	0.032***	0.033***	-0.036	-0.03	0.001	0.005
	[0.009]	[0.009]	[0.009]	[0.016]	[0.006]	[0.007]	[0.007]	[0.012]	[0.028]	[0.029]	[0.029]	[0.050]
At maximum capacity	0.013**	0.02***	0.016**	0.008	0.035***	0.035***	0.028***	0.024***	-0.04*	-0.033	0.001	-0.022
	[0.006]	[0.006]	[0.006]	[0.012]	[0.004]	[0.004]	[0.005]	[0.008]	[0.021]	[0.021]	[0.021]	[0.038]
Somewhat below maximum capacity	0.006	0.012**	0.01	0.004	0.022***	0.022***	0.015***	0.004	-0.03	-0.026	0.001	-0.029
	[0.006]	[0.006]	[0.006]	[0.011]	[0.004]	[0.004]	[0.004]	[0.008]	[0.021]	[0.021]	[0.021]	[0.038]
Sales (in logs)		-0.061***	-0.072***	-0.071***		0.030***	0.029***	0.040***		-0.078**	-0.068**	-0.022
		[0.010]	[0.010]	[0.018]		[0.007]	[0.008]	[0.014]		[0.030]	[0.032]	[0.054]
Sales squared (in logs)		0.002***	0.002***	0.001**		-0.001***	-0.001***	-0.001**		0.002**	0.001	0.000
		[0.000]	[0.000]	[0.001]		[0.000]	[0.000]	[0.000]		[0.001]	[0.001]	[0.002]
Profit status: loss			-0.013**	0.001			-0.032***	-0.031***			0.293***	0.169***
			[0.005]	[0.009]			[0.004]	[0.007]			[0.028]	[0.051]
Profit status: break-even			-0.025***	-			-0.022***	-			0.036	-
			[0.007]				[0.005]				[0.024]	
Wagebill (in logs)			0.022***	0.021***			-0.008***	-0.008**			0.015**	0.013
			[0.002]	[0.004]			[0.002]	[0.003]			[0.007]	[0.013]
Age: 2 ≤ age < 5 years			0.008	0.029			0.000	0.052			0.140	0.421**
			[0.034]	[0.052]			[0.023]	[0.044]			[0.130]	[0.196]
Age: 5 ≤ age < 10 years			-0.007	0.013			-0.009	0.030			0.182	0.431**
			[0.033]	[0.049]			[0.022]	[0.042]			[0.127]	[0.188]

Table 5 (continued)

	f1_g_sales	f1_g_sales	f1_g_sales	f1_g_sales	f1_g_empl	f1_g_empl	f1_g_empl	f1_g_empl	f1_g_profit	f1_g_profit	f1_g_profit	f1_g_profit
Age: 10 ≤ age < 20 years			-0.025	0.003			-0.009	0.037			0.133	0.418**
			[0.033]	[0.048]			[0.022]	[0.041]			[0.127]	[0.187]
Age: 20+ years			-0.039	-0.015			-0.024	0.016			0.121	0.397**
			[0.032]	[0.048]			[0.022]	[0.041]			[0.126]	[0.186]
Subsidiary			-0.003	-0.003			-0.005*	-0.008*			-0.010	-0.009
			[0.004]	[0.006]			[0.003]	[0.004]			[0.012]	[0.019]
Directly exported			0.015***	0.017***			0.005**	0.003			0.021*	0.031
			[0.004]	[0.006]			[0.003]	[0.005]			[0.011]	[0.020]
Invested in another country			0.010	-0.003			-0.001	-0.006			0.010	0.023
			[0.006]	[0.008]			[0.005]	[0.006]			[0.018]	[0.024]
R&D investment dummy				0.092***				0.004				-0.020
				[0.034]				[0.026]				[0.105]
IT investment dummy				0.001				0.001*				0.001
				[0.001]				[0.000]				[0.002]
Training investment dummy				0.001				0.001				0.001
				[0.001]				[0.001]				[0.002]
Business processes inv. dummy				0.000				0.000				-0.002
				[0.001]				[0.000]				[0.002]
Replacing capacity - investment (%)				-0.005				0.002				0.023
				[0.023]				[0.018]				[0.070]
Expanding capacity - inv. (%)				0.038				0.042**				0.082
				[0.024]				[0.019]				[0.072]
Investment in R&D (%)				-0.020				0.019				0.030
				[0.023]				[0.018]				[0.070]
Investment wrt t-1: broadly stayed the same				-0.005				-0.020***				0.014
				[0.006]				[0.004]				[0.017]

Table 5 (continued)

	f1_g_sales	f1_g_sales	f1_g_sales	f1_g_sales	f1_g_empl	f1_g_empl	f1_g_empl	f1_g_empl	f1_g_profit	f1_g_profit	f1_g_profit	f1_g_profit
Investment wrt t-1: less than prev. year				-0.017**				-0.023***				-0.006
				[0.008]				[0.006]				[0.024]
Innov: new to country				-				-				-
Innov: new to global mkt				-0.008				-0.002				-0.038
				[0.009]				[0.007]				[0.028]
Business prospects: same				-0.028***				-0.011**				-0.027
				[0.006]				[0.004]				[0.018]
Business prospects: deteriorate				-0.066***				-0.017**				-0.157***
				[0.009]				[0.006]				[0.027]
Profits before tax: 2% to 4%				-0.002				0.010*				-0.091***
				[0.008]				[0.006]				[0.026]
Profits before tax: 5% to 9%				-0.003				0.008				-0.131***
				[0.008]				[0.006]				[0.026]
Profits before tax: 10% to 14%				-0.014				0.000				-0.159***
				[0.009]				[0.007]				[0.029]
Profits before tax: 15% or more				-0.012				0.010				-0.197***
				[0.011]				[0.008]				[0.032]
Observations	18,724	18,724	17,277	5846	14,912	14,521	13,342	4425	12,880	12,754	12,116	4106
R-squared	0.020	0.027	0.039	0.054	0.018	0.021	0.035	0.061	0.004	0.006	0.025	0.043

Robust standard errors are in brackets. ***p < 0.01, **p < 0.05, *p < 0.1. A constant, country, sector and year dummies are included in all of the regressions but not reported in detail.

We report 100x the coefficient for better readability for: state-of-art machinery/equipment share and investment shares (%) in replacing capacity, expanding capacity, and for investment in R&D

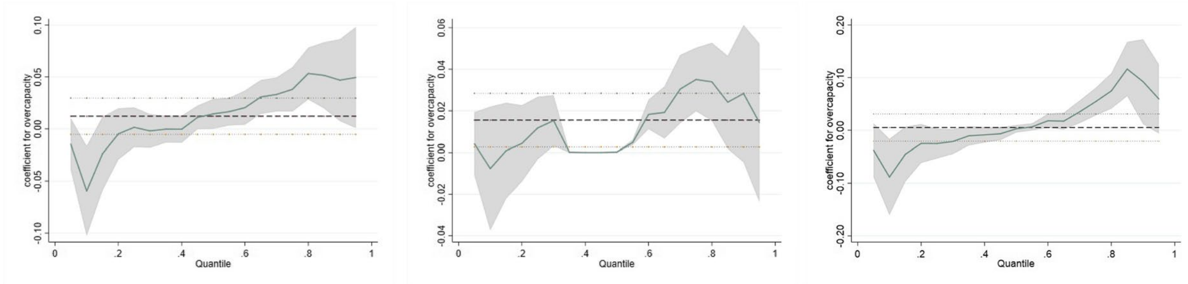


Fig. 6 Quantile regression results where the dependent variable is sales growth (left), employment growth (centre) or investment growth (right). Tables 6, 7 and 8 contain the corresponding regression output

First, it could be that the processes of hiring suitable employees take time. Second, the temporal lag could reflect the case that sales growth and profits growth are the causal drivers of subsequent employment growth (for example if the growth of sales and profits are a signal to executive decision-makers and investors, as well as means of generating cash flow, that are useful to justify the need for new hires). Third, it could be that existing employees are willing to accept the pressures of heavy workloads and overtime, but only in the short run, such that new employees are needed later on.

Appendix OSM-9 repeats the analysis in Table 5, but with a HGE dummy (for rapid growth in employment from t:t + 3) instead of the firm growth variables (t:t + 1). Although the number of observations is strongly reduced, nevertheless, we detect a positive and statistically significant relationship between operating above maximum capacity at time t and the subsequent probability of being an HGE in terms of employment growth (t:t + 3).

Table 5 also shows that some "usual suspects" (investment in IT, innovation, R&D) do not go far in explaining subsequent growth. The R^2 statistic from the growth rate regressions in Table 5 never rises above 6%, which is admittedly low but comparable to previous studies.[18]

Table 5 shows that capacity utilisation continues to play an important role in predicting growth of employment and also sales in the following period. One possible reason for this, in line with our "fork in the road" hypothesis, is that the average effect of overcapacity on subsequent growth masks the heterogeneity within the overcapacity category: that some firms take overcapacity as an opportunity to invest massively in a new growth trajectory, while others get "back to normal" (following the dynamics of mean reversion, perhaps) after a period of overcapacity. Quantile regression is a suitable econometric tool to investigate such heterogeneity in the effect of overcapacity across the growth rate distribution.

Figure 6 and Tables 6, 7 and 8 show the quantile regression results, where the dependent variable is the growth of sales, growth of employment or growth of investment, and the main explanatory variable of interest is an overcapacity dummy (which equals 1 for firms in a state of overcapacity). Figure 6 (left) shows that overcapacity is a significant predictor of subsequent sales growth and furthermore that the role of overcapacity depends on a firm's growth experience. For firms that have rapid sales growth in the year after overcapacity (at the upper quantiles of the conditional dependent variable), the overcapacity status is positive and statistically significant, hence associated with faster growth for these firms. However, for firms that have a decline in sales in the year after overcapacity, the overcapacity status is negative and statistically significant, which is consistent with the interpretation that overcapacity has a dampening role on sales growth for those firms that choose to shrink back after overcapacity. Similar results are found in Fig. 6 (right) for investment growth. These results are consistent with the "fork in the road" hypothesis.

For sales growth, the coefficient at the 90% quantile is 0.0316 (Table 6). This means that, all else equal, a rapid growth firm (at the 90% quantile of the conditional growth rates distribution) will have a (log difference) growth rate that is 0.0316 higher if it is operating above maximum capacity, compared to firms that are not operating above maximum capacity.

[18] See, e.g. Coad (2009, Table 7.1) for a review.

Table 6 Quantile regression results. Dependent variable: growth of sales in the following period (i.e. in *t:t+1*), where the predictor variables are measured at time *t*. The main explanatory variable of interest is the overcapacity dummy, which takes value 1 if firms report currently operating at above their maximum normal capacity

	10% quantile	25% quantile	50% quantile	75% quantile	90% quantile
overcapacity dummy	-0.042**	0.002	0.014**	0.030***	0.032**
	[0.016]	[0.008]	[0.005]	[0.008]	[0.014]
Sales (in logs)	0.032*	0.013	-0.031***	-0.120***	-0.167***
	[0.019]	[0.009]	[0.006]	[0.010]	[0.016]
Sales squared (in logs)	-0.002***	-0.001***	0.001***	0.003***	0.003***
	[0.001]	[0.000]	[0.000]	[0.000]	[0.001]
Age: $2 \leq$ age < 5 years	0.077	-0.011	-0.008	0.208	-0.057
	[0.060]	[0.029]	[0.020]	[0.032]	[0.054]
Age: $5 \leq$ age < 10 years	0.071	-0.002	-0.02	-0.012	-0.094*
	[0.060]	[0.029]	[0.020]	[0.031]	[0.052]
Age: $10 \leq$ age < 20 years	0.089	-0.011	-0.011	-0.037	-0.133**
	[0.058]	[0.028]	[0.020]	[0.030]	[0.051]
Age: 20 + years	0.106*	-0.009	-0.042**	-0.057*	-0.169***
	[0.058]	[0.028]	[0.019]	[0.030]	[0.050]
Wagebill (in logs)	0.079***	0.024***	0.009***	0.009***	0.016***
	[0.004]	[0.002]	[0.001]	[0.002]	[0.004]
Subsidiary dummy	-0.003	- 0.009**	- 0.007**	-0.002	0.003
	[0.008]	[0.004]	[0.003]	[0.005]	[0.008]
Exporter dummy	0.024***	0.008**	0.015***	0.023***	0.012*
	[0.007]	[0.003]	[0.003]	[0.004]	[0.007]
R&D dummy	-0.007	-0.003	0.004	0.011**	0.0126
	[0.009]	[0.004]	[0.003]	[0.005]	[0.008]
Constant	- 1.043***	-0.386***	0.256***	1.188***	1.870***
	[0.151]	[0.073]	[0.051]	[0.080]	[0.138]
Observations	22,681	22,681	22,681	22,681	22,681

Standard errors in brackets. These quantile regressions do not control for country, year and macro-sector fixed effects, because of computational issues of non-convergence. Key to significance stars: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

This increase in the sales growth rate of 0.0316 is not negligible, considering that sales growth has a mean of 0.05 and a standard deviation of 0.21 (see the summary statistics in Table OSM-2). At the lower quantiles of the conditional growth rate distribution, the coefficient is even larger in magnitude (-0.0417), indicating that firms above maximum capacity shrink faster than their counterparts in other capacity utilisation categories at the 10% quantile of the (conditional, log difference) sales growth distribution.

Fig. 6 (centre) shows the corresponding results for employment growth. The coefficients are positive and significant at the upper end of the conditional employment growth distribution, which indicates that overcapacity is associated with faster employment growth for those firms that experience rapid employment growth in the year after overcapacity. However, the results at the lower quantiles are indistinguishable from zero instead of being negative (as we observed in Fig. 6 for growth of sales and investment). The employment growth results only confirm one of the two prongs of the "fork": overcapacity is positively linked to growth for some growing firms, but overcapacity is not related to employment decline for the other firms. This "asymmetry" in the quantile regression results for employment growth (compared to sales growth) could be due to firing restrictions, that discourage layoffs, leading the regression coefficient of overcapacity to be prevented from taking negative values (associated with job destruction) and instead of taking values close to zero (corresponding to zero job growth from overcapacity).

Tables 6, 7 and 8 present the quantile regression results in more detail. Some of the explanatory variables are positively associated with growth rates at all quantiles, i.e. log(wage bill) and the exporter dummy in Table 6 (sales growth). Subsidiaries are generally associated with lower employment growth at all quantiles (Table 7).

In other cases, some variables play a "stabilizing" role in the sense that they are associated with higher growth of declining firms (at low quantiles) but with reduced growth of faster-growing firms (at the upper quantiles). Such variables are age (20 + years dummy; see Table 6) and also log(wage bill) in Table 8.

R&D investment is associated with employment growth in the upper quantiles in Table 7, suggesting that R&D investment is associated with employment growth for fast-growth firms. Firm size is significantly associated with growth, although the effect is non-linear, with the quadratic terms varying over the quantiles.

Further analysis seeks to investigate the characteristics of firms who, upon reaching the fork in the road, respond with either growth or decline. This is operationalized using interaction terms in a similar quantile regression framework. The results were far from statistically significant, however, no doubt linked to the small number of observations in the overcapacity category.[19]

5 Discussion of the findings

The discussion of our main results can be organized into six themes.

First, it appears that entering into a state of overcapacity is generally a felicitous state, rather than a costly mistake of misjudging one's ability to satisfy incoming orders. This in itself was not clear in the previous literature. At a firm level, the rapid growth of sales and profits is observed in the years before, and also during, entry into a state of overcapacity (Fig. 4).[20] Sales growth and profits growth return to normal levels after the overcapacity event, however, suggesting perhaps that the fortuitous swelling in demand must eventually come to an end. Employment growth of the category of firms operating at overcapacity at time t is high in the years before, during and after the overcapacity event, which—coupled with the lasting growth of the wage bill of these firms—suggests that firms that reach a state of overcapacity are more likely to be HGEs that create many jobs of good quality (Fig. 4).

Second, our observations that growth of sales and profits precede entry into overcapacity is consistent with suggestions in the previous literature that firms enter into a state of operating above maximum capacity because of surging demand (Nikiforos, 2013; Foster et al., 2016; Pozzi & Schivardi, 2016). This is also in tune with evidence (in Appendix OSM-6) that firms at overcapacity are less likely to report major barriers related to demand or uncertainty about the future.

Third, our results connect to our novel "fork in the road" hypothesis, which posits that firms that find themselves in a state of critically high capacity utilisation can respond in one of two ways: either they build upon the current momentum to make the broad-based investment in various areas or else they wait for the wave to subside, remaining within the current capacity limits while the capacity utilisation rate falls. Therefore, the "fork in the road" hypothesis suggests that firms in a state of overcapacity are a heterogeneous group and therefore have a high variation in their growth rates.

We observed some support for the "fork in the road" hypothesis. An indicator of variation (the interquartile range, IQR) was high for the category of firms "above maximum capacity", although the highest IQR values were observed for firms "far below maximum capacity" (Appendix OSM-10). Focusing on investment dynamics, we observed a sharp increase in the variation in growth rates for total investment in the year after the overcapacity event (Appendix OSM-10). Clearer support came from our quantile regressions for sales growth (Fig. 6, Table 6): overcapacity status was observed to have a significant positive association with the growth of sales and investment at the upper quantiles of the growth rates distribution (for the fastest-growing firms), while having a significant negative association with the growth of sales and investment at the lower quantiles. Overall, therefore, our evidence on the "fork in the road"

[19] See Appendix OSM-12 for the quantile regression results for the interaction terms terms "d_overcapacity × Exporter" and "d_overcapacity × R&D".

[20] At a country-level, preliminary explorations show that a larger shares of firms operating above maximum capacity are positively related to higher average scores for management capabilities (see Appendix OSM-11).

Table 7 Quantile regression results. Dependent variable: growth of employment in the following period (i.e. in *t*:*t* + *1*), where the predictor variables are measured at time *t*. The main explanatory variable of interest is the overcapacity dummy, which takes value 1 if firms report currently operating at above their maximum normal capacity

	10% quantile	25% quantile	50% quantile	75% quantile	90% quantile
overcapacity dummy	0.004	0.009	0.003**	0.029***	0.031**
	[0.011]	[0.007]	[0.001]	[0.008]	[0.012]
Sales (in logs)	0.076***	0.103***	-0.005***	0.034***	-0.016
	[0.013]	[0.008]	[0.002]	[0.009]	[0.015]
Sales squared (in logs)	0.0020***	- 0.003***	0.0002***	- 0.001***	0.001
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Age: 2 ≤ age < 5 years	-0.075	-0.028	0.000	0.038	0.053
	[0.050]	[0.032]	[0.006]	[0.033]	[0.055]
Age: 5 ≤ age < 10 years	-0.06	-0.028	0.000	0.046	0.034
	[0.049]	[0.031]	[0.006]	[0.033]	[0.054]
Age: 10 ≤ age < 20 years	-0.049	-0.027	0.000	0.019	0.019***
	[0.049]	[0.031]	[0.006]	[0.033]	[0.054]
Age: 20 + years	-0.037	-0.025	0.000	-0.01	-0.043
	[0.049]	[0.031]	[0.005]	[0.033]	[0.054]
Wagebill (in logs)	0.003	-0.005***	0.000	- 0.013***	- 0.024***
	[0.003]	[0.002]	[0.001]	[0.002]	[0.003]
Subsidiary dummy	-0.010*	- 0.009**	0.000	- 0.009**	- 0.011*
	[0.006]	[0.004]	[0.001]	[0.004]	[0.007]
Exporter dummy	0.000	0.002	0.000	0.003	−-0.005
	[0.005]	[0.003]	[0.000]	[0.004]	[0.006]
R&D dummy	0.001	0.002	0.000	0.015***	0.019***
	[0.006]	[0.004]	[0.001]	[0.004]	[0.007]
Constant	-0.882***	-0.883***	0.028**	-0.079	0.598***
	[0.112]	[0.071]	[0.013]	[0.075]	[0.123]
Observations	17,738	17,738	17,738	17,738	17,738

Standard errors are in brackets. These quantile regressions do not control for country, year and macro-sector fixed effects, because of computational issues of non-convergence. Key to significance stars: ***p < 0.01, **p < 0.05, *p < 0.1

hypothesis highlights the relevance for future theorizing and empirical investigations about overcapacity as trigger points for subsequent high-growth episodes.

Fourth, it appears that firms in a state of operating above maximum capacity are not currently investing massively in R&D, but rather firms that are making more incremental investments as they presumably continue riding the wave of surging demand. This was also not clear from the previous literature. Table 4 shows their investment plans are relatively focused on "capacity expansion for existing products" and that firms at overcapacity invest a relatively small share in "development and introduction of new products". Figure 5 shows how overcapacity firms cut back on total investment (although they continue to invest heavily in fixed assets). Similarly, Appendix OSM-7.1 shows that firms above maximum capacity are more likely (than firms far below capacity) to invest in buildings and new equipment. Appendix Table OSM-7.2 shows that firms at or above maximum capacity are actually *less* likely to invest in R&D. Appendix Table OSM-7.2 also shows that firms above maximum capacity are more likely to invest in training and also organisation/business process improvements, which are investments with more immediate and certain payoffs that could help these firms to alleviate the pressures of overcapacity through a more efficient use of their inputs. There is also evidence that firms in a state of overcapacity have a higher share of their equipment that is state-of-the-art, and invest more in digitalisation (Appendix Table OSM-7.3), and have recently introduced process innovations. A tentative explanation

Table 8 Quantile regression results. Dependent variable: growth of investment in the following period (i.e. in *t:t+1*), where the predictor variables are measured at time *t*. The main explanatory variable of interest is the overcapacity dummy, which takes value 1 if firms report currently operating at above their maximum normal capacity

	10% quantile	25% quantile	50% quantile	75% quantile	90% quantile
overcapacity dummy	-0.088***	-0.025	0.003	0.054***	0.092**
	[0.042]	[0.018]	[0.004]	[0.016]	[0.045]
log_sales	0.009	0.021	0.001	0.019	-0.127**
	[0.053]	[0.023]	[0.006]	[0.020]	[0.057]
Sales squared (in logs)	0.000	-0.001	0.000	-0.001	0.004**
	[0.002]	[0.001]	[0.000]	[0.001]	[0.002]
Age: $2 \leq$ age <5 years	0.031	0.178**	0.099***	0.085	0.303
	[0.201]	[0.086]	[0.021]	[0.077]	[0.215]
Age: $5 \leq$ age <10 years	0.105	0.217**	0.113***	0.0871	0.311
	[0.197]	[0.085]	[0.021]	[0.076]	[0.211]
Age: $10 \leq$ age <20 years	0.341*	0.297***	0.123***	0.068	0.166
	[0.196]	[0.084]	[0.021]	[0.075]	[0.209]
Age: 20 + years	0.443**	0.336***	0.126***	0.0543	0.0815
	[0.195]	[0.084]	[0.020]	[0.075]	[0.209]
Wagebill (in logs)	0.036***	0.019***	0.001	-0.009**	-0.057***
	[0.011]	[0.005]	[0.001]	[0.004]	[0.011]
Subsidiary dummy	-0.047**	-0.018*	-0.003	0.015*	0.074***
	[0.022]	[0.009]	[0.002]	[0.009]	[0.024]
Exporter dummy	-0.003	-0.009	0.000	0.005	-0.019
	[0.019]	[0.008]	[0.002]	[0.007]	[0.020]
R&D dummy	−0.0220	−0.0164*	−0.00171	0.00178	−0.0113
	-0.022	- 0.016*	-0.002	[0.009]	[0.024]
Constant	- 1.475***	- 0.873***	- 0.143***	0.0423	2.129***
	[0.448]	[0.194]	[0.047]	[0.173]	[0.480]
Observations	19,263	19,263	19,263	19,263	19,263

Standard errors in are brackets. These quantile regressions do not control for country, year and macro-sector fixed effects, because of computational issues of non-convergence. Key to significance stars: ***$p<0.01$, **$p<0.05$, *$p<0.1$

could be that firms operating above maximum capacity are the victims of their success since they do not have enough slack to engage in exploratory innovation investments. Instead, they take a relatively short-term investment horizon, coasting along on their current wave of surging demand for successful products, rather than attempting to introduce radical changes or implementing long-term investment plans.

Fifth, we can cautiously propose a sequential co-evolutionary model of firm growth around the time of operating above maximum capacity. Initially, a rise in demand leads to growth of sales, which subsequently pushes up profits, because the firm can sell larger quantities and cover its fixed costs of operations. At the same time, growing firms hire new employees to help release the pressure of overcapacity. This is in line with other research using causal methods (e.g. Moneta et al., 2013; Coad & Grassano, 2019): with sales growth and employment growth at the start, and profits growth as a useful by-product rather than the stimulus, and finally where R&D investment comes later on (if at all) as firms invest available funds into R&D. This model of growth dynamics is also consistent with suggestions of the key role of rising demand on the capacity choices of new firms (Foster et al., 2016; Nikiforos, 2013; Pozzi & Schivardi, 2016).

Sixth, overcapacity can be seen as a firm-specific problem, as firms face their own workflows and bottlenecks, and as such, it may be difficult to generalize an appropriate policy response when the barriers faced by firms are so heterogeneous (Fischer &

Karlan, 2015). We therefore suggest that policy could focus on a narrow set of specific junctures or "trigger points" that may be related to capacity utilisation (as discussed in Section 2.3): hiring a first employee,[21] first steps into internationalisation, introducing a second product, building a second production plant, investing in next-generation capital equipment, overcoming a regulatory threshold for firm size and so on. For example, an initiative to help firms overcome regulatory thresholds for more stringent employment protection requirements could include a temporary (e.g. 2 years) freeze on the costs of growth, until firms become accustomed to operating at a larger size.[22]

6 Conclusions and future research

High-growth enterprises (HGEs) make a disproportionately large contribution to economic dynamism, innovation and productivity growth (Birch, 1979; Henrekson & Johansson, 2010; Coad et al., 2014). It is no surprise, therefore, that they receive considerable policy interest (Grover Goswami et al., 2019; Flachenecker et al., 2020). However, previous research has shown that HGEs are difficult to predict, thus making them a difficult policy target. A major problem is that rapid growth events are episodic, rather than being a stable time-invariant characteristic of firms (Daunfeldt & Halvarsson, 2015). This study takes a different approach. On the one hand, we develop a theory of capacity utilisation constraints as critical junctures in the growth process, showing in particular how operating above maximum capacity could be a springboard for subsequent HGE episodes. We investigate this theory by drawing on a unique and novel data source to provide a multifaceted view on firm growth and capacity utilisation, thereby giving a rich set of new results.

Firms begin to operate above maximum capacity after a period of the rapid growth of sales and profits, consistent with explanations that capacity constraints follow on from rising demand. Firms at overcapacity have rapid employment growth both before and after being at overcapacity. Firms at overcapacity make investments in future growth, but more from the angle of capacity expansion, process improvements and investment in modern machinery, rather than in R&D and new product development. There is evidence that firms take two routes out of overcapacity: either overcapacity is linked to the subsequent rapid growth of sales (as firms launch into subsequent growth) or overcapacity is linked to declining sales (as firms shrink back to "normal" production levels)—in line with our "fork in the road" hypothesis.

We contribute to the literature by demonstrating a "fork in the road" effect: i.e. showing that operating above maximum capacity corresponds to a "trigger point" or decision point, whereby firms can either respond by investing massively in further growth or by shrinking back to stay within current capacity utilisation limits. Future research on new datasets could build on this finding to further investigate the heterogeneity between firms that reach this decision point and grow and those that reach this decision point and shrink. Future research could also use fine-grained industry classification codes to explore how the relationship between capacity utilisation and growth is moderated by sector.

The approach taken in this study is descriptive, and the results are presented in the form of (conditional) associations rather than causal effects. While our results can be useful for making predictions, nevertheless they cannot per se conclusively identify the causal mechanisms in place. Future work could potentially find clever ways to investigate whether being above maximum capacity is entirely demand led (i.e. if firms struggle to satisfy their hungry customers) or a proactive business decision (i.e. if firms produce as much as possible while hoping to find buyers or while making proactive marketing efforts to find new buyers).

[21] Please bear in mind, however, that our analysis cannot provide any direct evidence on the hiring of the first employee since firms in our sample have 5 + employees.

[22] In France, a critical threshold is reached when a firm has 50 employees, because this is when many restrictive labour regulations come into force. As a result, many firms stay just below this threshold, with just 49 employees (see Garicano et al., 2016, their Fig. 2). Policy could help firms overcome this size threshold in such a way that the costs of growth are delayed—e.g. by fixing that firms only have to apply these regulations 2 or 3 years after they cross the threshold, as long as firm size remains above this threshold. If firms shrink back below this threshold before 2–3 years, then they will not be affected by these labour regulations. This way, risk-averse firms could taste the benefits of larger size before having to face the full costs. This strategy of promoting growth could help firms overcome their short termism and could be politically feasible in the corporate world, a cynic might suggest, e.g. if top managers serve fixed terms and seek to be rewarded for growth, while "dumping" the costs of growth on their successors.

Appendix

Capacity utilisation categories for different countries

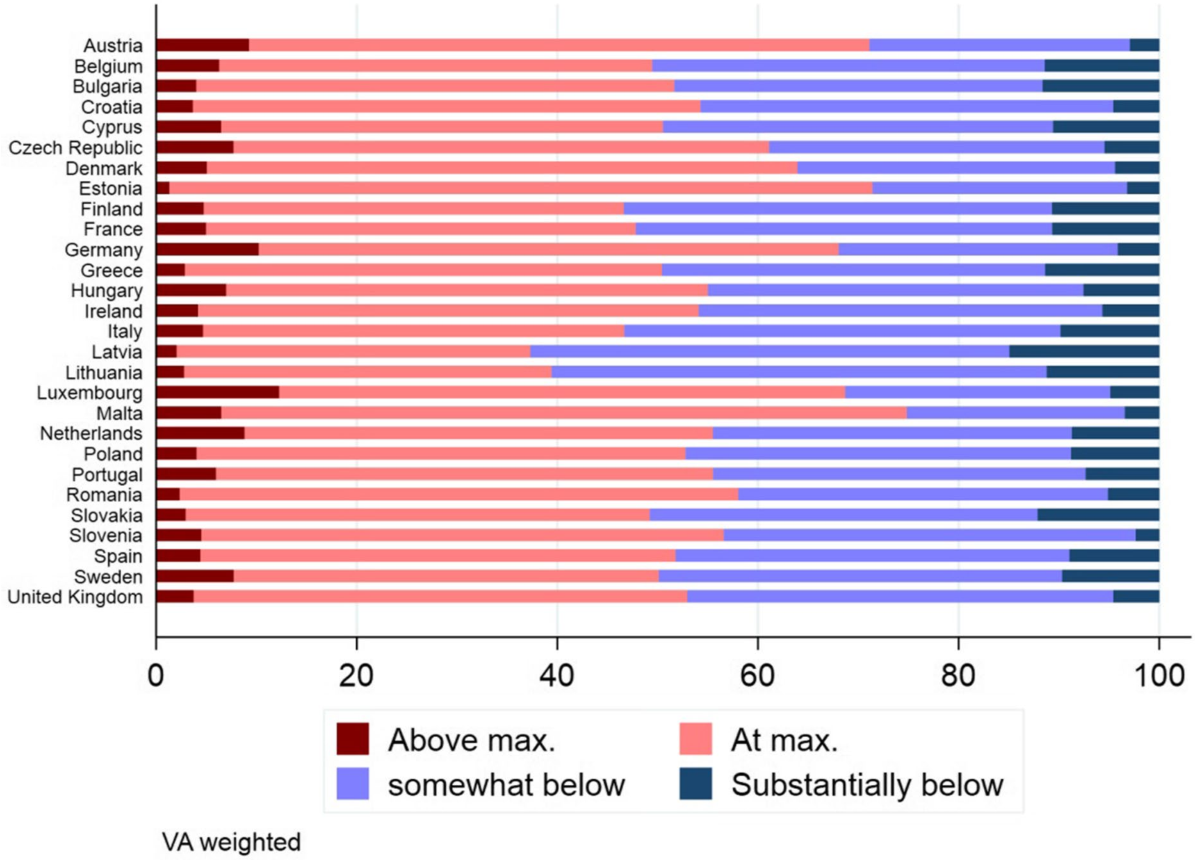


Fig. 7 Capacity utilisation categories for different countries, weighted by value added

Acknowledgements We are grateful to Maja Andjelkovic, Xavier Cirera, Sven-Olov Daunfeldt, Pedro de Lima, James Gavigan, Arti Grover Goswami, Po-Hsuan Hsu, Hyun Ju Jung, Denis Medvedev, Santiago Reyes Ortega, Stjepan Srhoj, Alexander Stepanov, Jonathan Timmis and participants at the European Commission workshop on "Entrepreneurial Ecosystems for high-growth enterprises", the 8th Taiwan Symposium on Innovation Economics and Entrepreneurship (National Tsing Hua University, Taipei), and seminars at HSE Graduate School of Business (Moscow) and the World Bank (Washington, DC) for many helpful suggestions and comments. Aroa Ortiga Rodríguez provided excellent research assistance. The opinions expressed herein are those of the authors and do not necessarily reflect those of the European Investment Bank or the European Commission. The usual disclaimers apply.

Funding Open access funding provided by Universitat Rovira i Virgili. This work was supported by CT-EX2017D318324-101 and CT-EX2014D180880-103 by the European Commission. Mercedes Teruel received also support from Universitat Rovira i Virgili (2019PFR-URV-B2-80) and the Consolidated Group of Research (2014-SGR-1395). Alex Coad received support from the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2018S1A3A2075175). Open access funding provided by Universitat Rovira i Virgili.

References

Arata, Y. (2019). Firm growth and Laplace distribution: The importance of large jumps. *Journal of Economic Dynamics and Control, 103*, 63–82.

Ashton, T. S. (1926). The growth of textile businesses in the Oldham district, 1884–1924. *Journal of the Royal Statistical Society, 89*(3), 567–583.

Bak, P., Chen, K., Scheinkman, J., & Woodford, M. (1993). Aggregate fluctuations from independent sectoral shocks: Self-organized criticality in a model of production and inventory dynamics. *Ricerche Economiche, 47*(1), 3–30.

Baldwin, J. R., Gu, W., & Yan, B. (2013). Export growth, capacity utilization, and productivity growth: Evidence from the Canadian manufacturing plants. *Review of Income and Wealth, 59*(4), 665–688.

Baltagi, B. H., Griffin, J. M., & Vadali, S. R. (1998). Excess capacity: A permanent characteristic of US airlines? *Journal of Applied Econometrics, 13*(6), 645–657.

Basu, S. (1996). Procyclical productivity: Increasing returns or cyclical utilization? *Quarterly Journal of Economics, 111*(3), 719–751.

Berndt, E. R., & Morrison, C. J. (1981). Capacity utilization measures: Underlying economic theory and an alternative approach. *American Economic Review Papers and Proceedings, 71*(2), 48–52.

Birch, D. L. (1979), The Job Generation Process. MIT Program on Neighborhood and Regional Change. Massachusetts Institute of Technology: Cambridge, MA.

Bloom, N., Lemos, R., Sadun, R., Scur, D., & Van Reenen, J. (2014). JEEA-FBBVA Lecture 2013: The new empirical economics of management. *Journal of the European Economic Association, 12*(4), 835–876.

Bornhäll, A., Daunfeldt, S. O., & Rudholm, N. (2017). Employment protection legislation and firm growth: Evidence from a natural experiment. *Industrial and Corporate Change, 26*(1), 169–185.

Bottazzi, G., & Secchi, A. (2006). Explaining the distribution of firm growth rates. *Rand Journal of Economics, 37*(2), 235–256.

Brown, R., & Mawson, S. (2013). Trigger points and high-growth firms. *Journal of Small Business and Enterprise Development, 20*(2), 279–295.

Brutscher P.-B., Coali A., Delanote J., Harasztosi P., (2020). EIB Group Survey on Investment and Investment Finance: A technical note on data quality. European Investment Bank, working paper 2020/08. https://doi.org/10.2867/772584

Butters, R. A. (2020). Demand volatility, adjustment costs, and productivity: An examination of capacity utilization in hotels and airlines. *American Economic Journal: Microeconomics, 12*(4), 1–44.

Caliendo, L., Monte, F., & Rossi-Hansberg, E. (2015). The anatomy of French production hierarchies. *Journal of Political Economy, 123*(4), 809–852.

Coad, A. (2009). *The growth of firms: A survey of theories and empirical evidence*. Edward Elgar.

Coad, A., Daunfeldt, S. O., Hölzl, W., Johansson, D., & Nightingale, P. (2014). High-growth firms: Introduction to the special section. *Industrial and Corporate Change, 23*(1), 91–112.

Coad, A., Frankish, J. S., & Storey, D. J. (2020). Too fast to live? Effects of growth on survival across the growth distribution. *Journal of Small Business Management, 58*(3), 544–571. https://doi.org/10.1080/00472778.2019.1662265

Coad, A., Grassano N. (2019). Firm growth and R&D investment: SVAR evidence from the world's top R&D investors. *Industry & Innovation, 26*(5), 508–533. https://doi.org/10.1080/13662716.2018.1459295

Coad, A., Nielsen, K., & Timmermans, B. (2017). My first employee: An empirical investigation. *Small Business Economics, 48*(1), 25–45.

Coad, A., & Planck, M. (2012). Firms as bundles of discrete resources – Towards an explanation of the exponential distribution of firm growth rates. *Eastern Economic Journal, 38*, 189–209.

Coad, A., & Srhoj, S. (2020). Catching gazelles with a lasso: Big data techniques for the prediction of high-growth firms. *Small Business Economics, 55*(3), 541–565. https://doi.org/10.1007/s11187-019-00203-3

Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly, 17*(1), 1–25.

Cruz, M., Bussolo, M., & Iacovone, L. (2018). Organizing knowledge to compete: Impacts of capacity building programs on firm organization. *Journal of International Economics, 111*, 1–20.

Dana, J. D., Jr., & Orlov, E. (2014). Internet penetration and capacity utilization in the US airline industry. *American Economic Journal: Microeconomics, 6*(4), 106–137.

Daunfeldt, S.-O., & Halvarsson, D. (2015). Are high-growth firms one-hit wonders? Evidence from Sweden. *Small Business Economics, 44*, 361–383.

Delmar, F., Davidsson, P., & Gartner, W. B. (2003). Arriving at the high-growth firm. *Journal of Business Venturing, 18*, 189–216.

Doms, M., & Dunne, T. (1998). Capital Adjustment Patterns in Manufacturing Plants. *Review of Economic Dynamics 1*(2), 409–429

ECB .(2007). Monthly Bulletin: October 2007. European Central Bank. Frankfurt, Germany.

Fairlie, R. W., & Miranda, J. (2017). Taking the leap: The determinants of entrepreneurs hiring their first employee. *Journal of Economics & Management Strategy, 26*(1), 3–34.

Ferrando, A., Pal, R., & Durante, E. (2019). Financing and obstacles for high growth enterprises: the European case (No. 2019/03). EIB Working Papers.

Fischer, G., & Karlan, D. (2015). The catch-22 of external validity in the context of constraints to firm growth. *American Economic Review, 105*(5), 295–299.

Flachenecker, F., Gavigan, J., P., Goenaga, X., Pasi, G., Preziosi, N., Stamenov, B., & Testa, G. (2020). High growth enterprises: Demographics, financing & policy measures. JRC Technical Report. Joint Research Centre. Brussels, Belgium. https://ec.europa.eu/jrc/en/publication/high-growth-enterprises-demographics-finance-policy-measures

Foster, L., Haltiwanger, J., & Syverson, C. (2016). The slow growth of new plants: Learning about demand? *Econometrica, 83*(329), 91–129.

Garicano, L., & Steinwender, C. (2016). Survive another day: Using changes in the composition of investments to measure the cost of credit constraints. *Review of Economics and Statistics, 98*(5), 913–924.

Garicano, L., Lelarge, C., & Van Reenen, J. (2016). Firm size distortions and the productivity distribution: Evidence from France. *American Economic Review, 106*(11), 3439–3479.

George, G. (2005). Slack resources and the performance of privately held firms. *Academy of Management Journal, 48*(4), 661–676.

Geroski, P., & Gugler, K. (2004). Corporate Growth Convergence in Europe. *Oxford Economic Papers, 56*, 597–620.

Grover Goswami, A., Medvedev, D., & Olafsen, E. (2019). *High-growth firms: Facts, fiction, and policy options for emerging economies*. World Bank

Henderson, J., & Cool, K. (2003). Corporate governance, investment bandwagons and overcapacity: An analysis of the worldwide petrochemical industry, 1975–95. *Strategic Management Journal, 24*(4), 349–373.

Henrekson, M., & Johansson, D. (2010). Gazelles as job creators: A survey and interpretation of the evidence. *Small Business Economics, 35*, 227–244.

Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica, 46*(1), 33–50.

Liu, G., Zhang, X., Zhang, W., & Wang, D. (2019). The impact of government subsidies on the capacity utilization of zombie firms. *Economic Modelling, 83*, 51–64.

McKenzie, D. (2017). Identifying and spurring high-growth entrepreneurship: Experimental evidence from a business plan competition. *American Economic Review, 107*(8), 2278–2307.

Moneta, A., Entner, D., Hoyer, P., & Coad, A. (2013). Causal inference by independent component analysis: theory and applications. *Oxford Bulletin of Economics and Statistics, 75*(5), 705–730.

Nikiforos, M. (2013). The (normal) rate of capacity utilization at the firm level. *Metroeconomica, 64*(3), 513–538.

Nohria, N., & Gulati, R. (1996). Is slack good or bad for innovation? *Academy of Management Journal, 39*(5), 1245–1264.

Ocasio, W. (1997). Towards an attention-based view of the firm. *Strategic Management Journal, 18*(S1), 187–206.

Penrose, E. T. (1959). *The theory of the growth of the firm*. Basil Blackwell.

Pozzi, A., Schivardi, F. (2016). Demand or productivity: What determines firm growth? *Rand Journal of Economics, 47*(3), 608–630

Reypens, C., Delanote, J., & Rückert, D. (2020). From starting to scaling.

Rostamkalaei, A., & Freel, M. (2016). The cost of growth: Small firms and the pricing of bank loans. *Small Business Economics, 46*(2), 255–272.

Schivardi, F., & Torrini, R. (2008). Identifying the effects of firing restrictions through size-contingent differences in regulation. *Labour Economics, 15*(3), 482–511.

Shen, G., & Chen, B. (2017). Zombie firms and over-capacity in Chinese manufacturing. *China Economic Review, 44*, 327–342.

Stanley, M. H. R., Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Leschhorn, H., Maass, P., Salinger, M. A., & Stanley, H. E. (1996). Scaling behavior in the growth of companies. *Nature, 379*, 804–806.

Storey, D. J. (1994). *Understanding the small business sector*. Routledge.

Storey, D. J. (2011). Optimism and chance: The elephants in the entrepreneurship room. *International Small Business Journal, 29*(4), 303–321.

Sutton, J. (1997). Gibrat's legacy. *Journal of Economic Literature, 35*(1), 40–59.

Tornqvist, L., Vartia, P., & Vartia, Y. O. (1985). How should relative changes be measured? *American Statistician, 39*(1), 43–46.

Wiklund, J., & Shepherd, D. (2003). Aspiring for, and achieving growth: The moderating role of resources and opportunities. *Journal of Management Studies, 40*(8), 1919–1941.

Zhang, H., Zheng, Y., Ozturk, U. A., & Li, S. (2016). The impact of subsidies on overcapacity: A comparison of wind and solar energy companies in China. *Energy, 94*, 821–827.

Zhou, H., & van der Zwan, P. (2019). Is there a risk of growing fast? The relationship between organic employment growth and firm exit. *Industrial and Corporate Change, 28*(5), 1297–1320.