Small Bus Econ
https://doi.org/10.1007/s11187-022-00614-9

High-growth firms' contribution to aggregate productivity growth

Márta Bisztray · Francesca de Nicola · Balázs Muraközy

Accepted: 9 February 2022
© The Author(s) 2022

Abstract This paper investigates the contribution of high-growth firms (HGFs) to aggregate productivity growth, using Hungarian firm-level data. Three stylized facts emerge. First, output-based HGFs substantially outperform employment-based ones in terms of their productivity contribution: on average, sales-based HGFs contribute 5 times as much as employment-based ones. Further, the contribution of employment-based HGFs is negative in 48-50% of industry-years, compared to 25-31% for sales-based HGFs. Second, HGFs tend to contribute to productivity growth only during their high-growth phase but not afterwards. Third, HGFs' contribution to productivity growth is higher in industries with more effective reallocation and with more young firms, but none of these are strong predictors of the HGFs' contribution. Finally, we present a simple benchmark model to show that these patterns arise naturally under realistic correlation structures.

We thank Jozef Konings and Miklós Koren for their insightful suggestions. We thank the Hungarian Academy of Sciences for funding this research as part of its Momentum Grant "Firms, Strategy and Performance".

M. Bisztray
Centre for Economic and Regional Studies (KRTK), Budapest and Corvinus University Budapest, CERGE-EI Foundation, Teaching Fellow, Budapest, Hungary

F. de Nicola
World Bank, Washington, DC, USA

B. Muraközy (✉)
University of Liverpool Management School, Liverpool and KRTK, Budapest, Chatham St, L69 7ZH Liverpool, UK
e-mail: murakozy@liverpool.ac.uk

Plain English Summary Firms that increase their sales quickly are responsible for a large part of industry-level productivity growth, but only during their high-growth phase. In contrast, firms that increase their employment quickly often experience falling productivity. This paper quantifies the contribution of high-growth firms (HGFs) to industry-level productivity growth, using Hungarian data. We find that i) the contribution depends strongly on the way growth is measured: firms growing in terms of revenue tend to contribute more than firms growing in terms of employment, ii) HGFs contribute to productivity growth mainly through their high-growth period, but not afterwards, iii) these contributions are not strongly associated with industry characteristics, though they tend to be larger in industries with more young firms. Our results are relevant for policymakers who are interested in the productivity effects of HGFs not only job creation, and suggest that expected productivity effects i) depend on the type of high growth, ii) are concentrated to the high-growth period, and iii) might not be enhanced by industry targeting.

Keywords High-growth firms · Productivity growth · Reallocation · Industry dynamic · Productivity decomposition

JEL classification: L25 · O40

Published online: 30 July 2022

1 Introduction

The benefits, costs, and rationale of policies promoting high-growth firms (HGFs) have been strongly debated (see e.g., (Acs et al., 2009; Bos and Stam, 2014; Autio and Rannikko, 2016; Brown et al., 2017; Grover Goswami et al., 2019)). This debate has emphasized that the main benefit of having high-growth firms is their ability to create jobs (e.g., (Henrekson and Johansson, 2010)). At the same time, additional effects of having more HGFs, including their productivity effects, are less clear, even though productivity growth, if not the sole aim of a policy, is likely to be at least one of the policymakers' objectives, and as such, should appear in well-designed cost-benefit analyses of entrepreneurship policies. As Brown et al. (2017) put it: "This myopic focus on HGF growth rates has also largely overlooked crucially important aspects of firm growth, notably firm profitability and sustainability and wider productivity." While some influential papers mention increased productivity growth as an established effect of HGFs,[1] others are very skeptical.[2]

This paper sheds light on these wider productivity effects by quantifying the contributions of HGFs to industry-level TFP and labor productivity growth based on detailed microdata from Hungary. The concept of productivity contribution we use, which follows Foster et al. (2008), captures both within-firm productivity growth and Schumpeterian reallocation effect, and, as a result, it is closer to the "wider productivity effects" than focusing only on the productivity growth of high-growth firms. While a number of studies have investigated either the productivity growth of HGFs or their initial productivity level, the main novelty of our paper is using a concept which captures both these channels with a focus on key questions from the HGF literature.[3][4] Importantly, it is not automatic that correlations between firm size and productivity growth will generalize to this level: as we show, the industry-level contribution combines a number of correlations, and the within and reallocation effects often have opposite signs. Studying which firm-level results persist at this aggregation level is an important contribution of our paper.

Besides deciding on whether to implement HGF promotion, policymakers also have a choice between promoting firms growing along different dimensions. We analyze the contribution of HGFs defined in different ways to identify any potential trade-off between productivity and employment contribution.[5] Indeed a large literature has demonstrated that different definitions capture different sets of firms with different average characteristics (Delmar, 2006; Shepherd and Wiklund, 2009). Our results contribute to this debate by demonstrating the presence of a trade-off between our measure of "wider productivity effects" and job creation. Furthermore, we show that definitions differ in the relative importance of the within-firm and Schumpeterian contribution. These results are mainly driven by the negative or weak relationship between employment growth and productivity growth during the high-growth period, also documented by Coad and Broekel (2012) and Guillamón et al. (2017). In particular we find, the contribution during the high-growth phase depends on the definition used, varies widely across industries and years and it is often negative. Negative contribution mainly comes from negative productivity growth during the high-growth phase, which is more frequent in

[1] As, for example Decker et al. (2016) writes, "the evidence shows that these high-growth young firms were relatively more innovative and productive, so their rapid growth contributed positively to productivity growth as more resources were shifted to these growing firms" or, as Brown and Mawson (2016) write "this work has overwhelmingly found HGFs to be a critically important Schumpeterian stimulus which drives up competition, increases firm entry and exits, generate exports, increases productivity and enhances overall economic competitiveness".

[2] Grover Goswami et al. (2019) write that "The evidence in this book shows that policies that seek out potential HGFs based on outward characteristics or target some desired share of HGFs are likely to be misguided. This is because the link between productivity and high growth is often weak".

[3] Closest to our work, Daunfeldt et al. (2014) calculates the contribution from the within term, but not from the between term. Haltiwanger et al. (2016) uses a similar methodology to us, but does not consider the timing, definition and industry characteristics we focus on in this paper.

[4] Even though, admittedly, our approach captures only the direct contribution of HGFs and not the externalities generated by them (see (de Nicola et al. , 2019; Du and Vanino , 2020), for such results).

[5] This idea is a key motivation for other studies on the heterogeneity of the definition. For example, Daunfeldt et al. (2014) write that "one question of importance is whether policymakers should target firms that experience high growth in terms of employment, sales, value added or productivity. There may for example be large societal costs to targeting HGFs in terms of employment by economic policy if the policy at the same time disfavors HGFs in terms of productivity".

the case of employment-based HGFs. As a result, the average contribution of sales-based HGFs is substantially higher than that of employment-based HGFs for the high-growth period of three years

HGF policies are also often motivated by longer-term effects of high-growth firms, therefore, it is crucial to understand the longer-term effects of HGFs. A number of results confirm that HGF status itself is not persistent (Coad, 2007; Daunfeldt et al., 2015; Hölzl, 2014; Erhardt, 2021), which "seriously challenges the notion that policymakers can target high-growth firms in order to promote future firm growth" (Daunfeldt et al., 2014). Still, it is theoretically possible that large size growth is preceded or followed by longer-term productivity growth if growth was enabled by innovation or if size growth is followed by organizational innovations (Greiner, 1989).[6] We follow cohorts of high growth firms to identify their contribution in the long run and show that HGFs mainly contribute positively to productivity growth during their high growth period.

Our third line of investigation focuses on the role of industry features in HGF productivity contribution. We find that specific characteristics of industry dynamics, captured by two key moments, substantially influence the contribution of HGFs to industry-level productivity growth. Stronger reallocation and a stronger correlation between productivity and size growth imply higher reallocation. Policies promoting industry dynamics in these respects may increase the benefits of entrepreneurship policy. These correlations, however, change quickly and may not be suitable for targeting. We show that more fundamental industry features—concentration, productivity dispersion, the share of young firms, and the technology level of the sector—cannot strongly predict the HGF contribution. We only find evidence for a positive relationship between the share of young firms and HGF contribution. These results are in line both with the low predictability of HGFs (see e.g., (Coad, 2007; Coad et al., 2014; Coad and Srhoj, 2019)) and the role of young firms in the HGF phenomenon (Navaretti et al., 2014; Coad et al., 2016; Coad, 2018).

This article contributes to four strands of the literature. First, the literature has documented a complex relationship between high growth and productivity at the firm level (Acs et al., 2009; Bianchini et al., 2017; Moschella et al., 2019; Arrighetti and Lasagni, 2013; Du and Temouri, 2015; Daunfeldt et al., 2014). As Grover Goswami et al. (2019) concludes, in general, there seems to be no strong connection between productivity and high-growth status. Our results reinforce the conclusion that high growth and productivity are interlinked in complex ways and sheds more light on this relationship by systematically reviewing how different HGF definitions are related to initial productivity level and productivity growth during the HGF phase.

Second, a large literature has documented how HGF definition matters (Delmar, 2006; Shepherd and Wiklund, 2009), as well as suggesting novel definitions (Acs et al., 2009; Moschella et al., 2019; Daunfeldt et al., 2014). In the current paper, we also emphasize the differences between the various HGF measures in the context of their contribution to aggregate productivity growth.

Third, while many papers have shown that HGFs create a large share of new jobs (see the meta-analysis of (Henrekson and Johansson, 2010)), there are only a few papers investigating the contribution of HGFs to aggregate productivity growth. Daunfeldt et al. (2014) shows that the within contribution of HGFs to aggregate economic growth, employment growth, sales growth, and productivity growth varies across the different HGF measures and can be even negative in some cases. Considering a 7-year period they find that the total productivity growth of employment-based HGFs is negative, while that of sales-based HGFs is positive but relatively low, about 7-8% of the aggregate productivity growth. In the current paper, we go deeper, and by extending the Foster et al. (2008) framework we also consider the reallocation contribution and show that it is a substantial part of HGFs' contribution. Additionally, we also consider cross-industry heterogeneity in HGFs contribution to aggregate productivity and investigate the industry characteristics which influence the magnitude of this contribution. The paper of Haltiwanger et al. (2016) is the closest to ours, as beyond looking at HGFs contribution to employment and real output growth they decompose the contribution of HGFs to industry-level productivity growth, focusing on the role of reallocation. They find that both HGFs and rapidly declining firms have a considerable contribution to aggregate productivity growth through reallocation. In addition to these results, we also look at the factors leading to cross-industry differences in the between and within

[6] For example Du and Temouri (2015) shows that sales-based HGFs tend to have a higher productivity growth after their high-growth period.

terms of HGFs' contribution to industry-level productivity. Moreover, we explicitly focus on differences by HGF definition comparing OECD- and Birch-type measures as well. Lastly, we also provide a simple model to show that the observed differences can be explained by a few moments of industry dynamics.

Finally, we contribute to the literature on reallocation (Foster et al., 2008; Hsieh and Klenow, 2009; Bartelsman et al., 2013; Restuccia and Rogerson, 2017; Baqaee and Farhi, 2020). Some papers already looked at how specific firm groups, like foreign-owned (Balsvik and Haller, 2006; Harris and Moffat, 2013) or exporters (Gleeson and Ruane, 2009; Fuss and Theodorakopoulos, 2018) contribute to aggregate productivity via within-firm growth and reallocation. We focus on the role of HGFs in aggregate reallocation and link that contribution to parameters characterizing the overall strength of reallocation, which has already been found important for aggregate productivity growth via different channels (e.g., (Andrews et al., 2015; Andrews et al., 2016)).

In what follows we first introduce our data and the decomposition methodology in Section 2. Section 3 introduces the different HGF definitions we use and presents descriptive statistics to show the differences between them. Section 4 presents our findings on the contribution of HGFs to aggregate productivity growth and the impact of industry characteristics. Section 5 presents a simple model, and Section 6 concludes.

2 Data and methods

2.1 Hungarian firm-level data

Our main source of information is the database of the firm-level corporate income tax statements during the period 2001-2016 from the Hungarian National Tax Authority (NAV). The dataset has almost universal coverage as it includes all firms using double-entry bookkeeping. The sample covers more than 95% of employment and value added of the business sector and about 55% of the full economy in terms of GDP. The dataset includes the most important balance sheet items and information on a wide range of matters such as ownership, employment, industry at the NACE 2-digit code level, and the location of the headquarters. The Centre for Economic and Regional Studies (KRTK) has extensively cleaned and harmonized the data (see Appendix 1). Nominal variables are deflated by the appropriate 2-digit industry-level deflators from OECD STAN.[7]

Given the scope of our analysis, we restrict the data in several ways. First, we exclude non-profit organizations. Second, we drop firms that operate either in agriculture or in the non-market service sectors of the economy. Third, we drop all firms that never had more than 10 employees, because standard HGF definitions require firms to have at least 10 employees.

When quantifying productivity, we mainly rely on TFP, estimated with the method proposed by Ackerberg et al. (2015). We present robustness checks by using labor productivity, calculated as log value added per employee. Given our focus on productivity, we drop firm-year observations where TFP (and labor productivity) is missing because the firm did not report material cost, or its employees or capital stock was reported to be zero. We use the same productivity measures as in Muraközy et al. (2018), and we describe the procedure in more detail in Appendix 2.

The number of observations is reported in Table 1. Column (1) shows the number of firms in our main sample, falling between 26,000-33,600 per year, and including firms that ever had at least 10 employees in our sample period, i.e., which have the (theoretic) potential to be defined as a HGF.[8] Column (2) shows that around 60-70% of these firms have at least 10 employees in the specific year. According to Column (3), 75-85% of the firms in Column (2) is observed 3 years prior to the focal year, showing that 15-25% of the firms in our sample has entered the sample in the preceding 3 years.[9] Similarly, Columns (4) and (5) include the number of firms that can be still observed 3 and 6 years after the focal year, respectively, showing that 60-66% of the firms having at least 10 employees in the specific year are observable even 6 years later.

[7] The same data and definitions have been used in previous papers on Hungarian HGFs, including the cross country study coordinated by the World Bank (Grover Goswami et al. , 2019) and de Nicola et al. (2019).

[8] We include all the observations of these firms in our sample because we would like to follow HGFs even if their number of employees falls below the threshold, rather than categorize them spuriously as entering or exiting firms in such cases.

[9] As we discussed, we also drop observations from the sample for which productivity cannot be measured. Therefore, the time of entry to the sample can differ from the firm's actual entry, e.g., if the firm operates with zero employees for an extended period.

Table 1 Number of firms by year

Year	(1)	(2)	(3)	(4)	(5)
		Firms which have more than 10 employees in year t,			
Year	Firms	observed in t	observed in t-3	observed in t+3	observed in t+6
2001	26,599	18,175	-	14,326	12,118
2002	28,402	18,906	-	14,836	12,043
2003	29,708	19,632	14,779	15,467	12,113
2004	31,838	20,437	15,185	16,304	12,826
2005	32,433	20,846	15,893	15,897	13,154
2006	33,134	21,364	16,584	15,916	13,118
2007	33,603	21,690	17,578	16,429	13,556
2008	33,506	21,850	17,128	17,093	14,344
2009	32,323	19,880	16,121	15,400	13,664
2010	32,585	20,324	16,496	15,943	13,498
2011	32,483	20,358	17,014	16,231	-
2012	31,381	19,672	16,120	16,391	-
2013	31,727	19,994	16,489	16,096	-
2014	31,990	21,099	17,228	-	-
2015	31,647	22,086	17,993	-	-
2016	29,421	21,603	18,107	-	-

This table shows the number of firms in different years. Column (1) includes the number of firms in our main sample, i.e., firms that had at least 10 employees any time in our sample period and which are present in year t. Column (2) shows how many of these firms had at least 10 employees in year t. Columns (3)-(5) include the subset of these firms that could also be observed in $t-3$, $t+3$ and $t+6$, respectively

Table 2 provides some basic descriptive statistics by year from our main sample. The average number of employees is between 40-45, suggesting that the overwhelming majority of the firms in the sample are SMEs. 13-17% of the firms are foreign-owned and 30-40% of them export. In line with the lower entry rate after 2000 compared to the 1990s, which was the decade after the transition, the share of young firms (at most 5 years old) decreases gradually over time, from 42% in 2000 to only 11% in 2016. The share of OECD employment HGFs is around 3-6%, with low values around the Great Recession in 2008/2009. The share of OECD sales HGFs is about twice as large as that of employment HGFs, also having lower values around the Great Recession. Table 7 in the Appendix shows similar statistics by 2-digit industry.

2.2 Decomposition

Our decomposition is based on Foster et al. (2008), who split aggregate productivity growth into within, between, cross and net entry terms.

The original decomposition starts with the change between t_0 and t (in our empirical exercise $t = t_0 + 3$) in aggregate productivity ($\Delta PROD_t$):

$$\Delta PROD_t = \underbrace{\sum_{i\in C}\theta_{i,t_0}\Delta prod_{i,t}}_{within} + \underbrace{\sum_{i\in C}(prod_{i,t_0} - PROD_{t_0})\Delta\theta_{i,t}}_{\text{between}} + \underbrace{\sum_{i\in C}\Delta prod_{i,t}\Delta\theta_{i,t}}_{\text{cross}} + \underbrace{\sum_{i\in N}\theta_{i,t}(prod_{i,t} - PROD_{t_0}) - \sum_{i\in X}\theta_{i,t_0}(prod_{i,t_0} - PROD_{t_0})}_{\text{net entry}}$$

where $\theta_{i,t}$ is the employment share of firm i in year t, $prod_{i,t}$ and $PROD_t$ are productivity measures at the firm and aggregate level, respectively. Δ always denotes the change between t_0 and t. C stands for continuing firms, N for new entrants and X for exiting firms.

The *within* term captures the sum of firm-level productivity changes for continuing firms, weighted by their initial employment share. This term is large if firms, especially large ones, increased their

Table 2 Descriptives by year

year	(1) av. emp	(2) sh. foreign	(3) sh. exporter	(4) sh. young	(5) sh. OECD (emp)	(6) sh. OECD (sales)
2001	48.3	16.3%	32.4%	36.0%	5.3%	10.8%
2002	45.2	15.5%	31.4%	31.5%	5.3%	9.1%
2003	44.0	14.9%	30.4%	28.9%	6.0%	10.3%
2004	41.6	14.0%	29.8%	27.3%	5.7%	9.8%
2005	40.7	13.7%	30.4%	25.0%	5.6%	9.2%
2006	40.6	13.6%	30.6%	23.6%	3.9%	5.5%
2007	40.5	13.8%	32.5%	22.3%	3.3%	5.3%
2008	40.5	14.0%	33.1%	21.6%	3.2%	5.4%
2009	39.0	14.0%	33.6%	20.6%	4.2%	8.1%
2010	39.2	14.0%	35.0%	20.0%	4.2%	8.1%
2011	40.0	14.1%	36.2%	19.8%	4.4%	9.0%
2012	40.5	14.1%	37.1%	19.3%	4.5%	10.9%
2013	40.5	13.8%	37.9%	18.2%	4.9%	11.3%
2014	41.4	13.9%	38.0%	16.6%		
2015	43.4	13.5%	38.7%	14.1%		
2016	45.6	13.1%	40.0%	11.1%		

This table shows the key descriptive statistics in our main sample. Column (1) shows the average number of employees, column (2) the share of foreign-owned firms, column (3) the share of exporters, column (4) the share of young firms (at most 5 years old), while columns (5) and (6) show the share of employment- and sales-based HGFs (OECD definition), respectively

productivity quickly. The *between* term captures the main channel of reallocation by quantifying the extent to which initially more productive firms grew faster in terms of employment. The *cross* term captures whether firms increasing their employment share were also able to improve their productivity. The *net entry* term is positive if new entrants were more productive relative to exiting firms.

Importantly, the reallocation is additive, and all these terms are sums of firm-level moments. Therefore, we can further distinguish between the contribution of HGFs and other continuing firms (similarly to (Haltiwanger et al., 2016)).[10] The decomposition becomes:

$$\Delta PROD_t = \underbrace{\sum_{i\in HGF} \theta_{i,t_0} \Delta prod_{i,t} + \sum_{i\in otherC} \theta_{i,t_0} \Delta prod_{i,t}}_{\text{within}} + \underbrace{\sum_{i\in HGF} (prod_{i,t_0} - PROD_{t_0})\Delta\theta_{i,t} + \sum_{i\in otherC} (prod_{i,t_0} - PROD_{t_0})\Delta\theta_{i,t}}_{\text{between}} + \underbrace{\sum_{i\in HGF} \Delta prod_{i,t}\Delta\theta_{i,t} + \sum_{i\in otherC} \Delta prod_{i,t}\Delta\theta_{i,t}}_{cross} + \underbrace{\sum_{i\in N} \theta_{i,t}(prod_{i,t} - PROD_{t_0}) - \sum_{i\in X} \theta_{i,t_0}(prod_{i,t_0} - PROD_{t_0})}_{\text{net entry}}$$

The total contribution of HGFs will be the sum of the three HGF terms, i.e., $\sum_{i\in HGF} \theta_{i,t_0} \Delta prod_{i,t} + \sum_{i\in HGF} (prod_{i,t_0} - PROD_{t_0})\Delta\theta_{i,t} + \sum_{i\in HGF} \Delta prod_{i,t}\Delta\theta_{i,t}$.

The terms in this formula suggest that HGF contribution is likely to be large in two cases. First, if high-growth firms *increase their productivity* during their high-growth phase, both the within and cross terms tend to be positive. The HGF *within* term captures whether HGFs increase their productivity. In the HGF *cross* term $\Delta\theta_{i,t}$ is positive by definition for all HGFs, hence the sign of the *cross* term is primarily determined by the sign of productivity changes. Therefore,

[10] In the decomposition we consider firms as HGFs if they are a HGF between t_0 and $t_0 + 3$, i.e., they grow fast in the three years following t_0 independently of what happens with them after $t_0 + 3$, i.e., grow fast, slow or exit. Note that, as the definition of HGFs requires these firms to be present before the HGF phase and by definition HGFs should still operate in t, there is no entry and exit of HGFs between t_0 and t.

Table 3 HGF definitions

	Input-based	Output-based
Relative (OECD) definition	Average annualized employment growth greater than 20% per annum, over a three-year period	Average annualized turnover or sales growth greater than 20% per annum, over a three-year period
Absolute + Relative (Birch) definition	Top 5 percentiles of the three-year average growth distribution, where growth in each period is measured by: $(emp_t - emp_{t-3})\frac{emp_t}{emp_{t-3}}$	Top 5 percentiles of the three-year average growth distribution, where growth in each period is measured by $(sales_t - sales_{t-3})\frac{sales_t}{sales_{t-3}}$

both terms are mainly driven by productivity growth during the high-growth phase, but the *within* term weights firm-level productivity changes with their initial size while the *cross* term weights them by their size growth. The second way HGFs can contribute positively is via reallocation. If high-growth firms are *more productive initially*, reallocation of resources to them will improve aggregate productivity. This channel is captured by the *between* term.

In our main exercise, we decompose productivity growth for 3-year periods starting in each year between 2001 and 2016 for each 2-digit industry.[11] Productivity decomposition is usually less noisy in such 'medium-term' periods, which may better reflect the timeline of economic processes like reallocation. 3-year periods also correspond to the time span of the standard OECD HGF definition. We present these decompositions separately for different *cohorts* of HGFs. We define the cohort of year t as the group of firms that were in their high-growth phase between t and $t + 3$.[12]

3 Definitions

This section introduces the different HGF definitions we use and presents a few patterns. This helps us (i) to understand how the different types of HGFs contribute to aggregate productivity growth and (ii) to investigate the relevant differences between the firms captured by the different definitions. To present the patterns in a transparent way, the figures and tables in this section mainly focus on firms that were HGFs between 2013 and 2016 (the 2013 cohort), which is our last cohort. The patterns are similar for other cohorts with few exceptions which we always note.

3.1 HGF definitions

The literature provides multiple definitions for HGFs (OECD, 2010), differing across two key dimensions. First, any type of size change can be measured in absolute or relative terms. One class of definitions relies solely on relative growth, while definitions in a second class use a combination of relative and absolute growth. For simplicity, we refer to the former as the *OECD* (based on (OECD, 2010)), and the latter as the *Birch* (Birch, 1981) method. Second, firms' performance can be assessed based on employment (more generally, input) or sales (output) dynamics.

Table 3 presents the typical definitions used in the literature. Within the relative definitions, one can distinguish between employment- (input) and sales- (output) based OECD definitions. The OECD definition requires a firm to grow by 20% on average per annum for three years. The Birch definition captures firms that are in the top 5 percentile based on an average of absolute and relative growth. Again, we will distinguish between labor and sales growth-based definitions. To make the results comparable, we will use the 3-year time frame in all cases.

Appendix 3 compares HGFs defined in different ways in several dimensions and documents many results which are consistent with the literature. Different definitions cover a partly overlapping but different set of firms, with significant differences in initial size and growth along the different dimensions (

[11] Consequently, these periods are overlapping: 2001-2004, 2002-2005 etc., resulting in 13 such periods, 2013 being the last year when we can observe 3 years of subsequent growth.

[12] For example, if a firm was growing fast (i.e., its employment increased yearly by more than 20% on average) between 2001-2004 and 2002-2005, but not afterward, it will be considered as a HGF in the 3-year periods between 2001–2004 and 2002–2005, but not in 2003-2006 or in any of the subsequent three-year periods.

e.g., (Delmar et al., 2003; Delmar, 2006; Shepherd and Wiklund, 2009; Acs et al., 2009; Arrighetti and Lasagni, 2013; Coad et al., 2014; Daunfeldt et al., 2014; Moreno and Coad, 2015; Du and Bonner, 2017; Grover Goswami et al., 2019)). Also consistent with earlier results, we find that HGF status is often a transitory phase in the firm's history, even though there is some persistence. (see (Delmar et al., 2003; Coad, 2007; Acs et al., 2009; Daunfeldt et al., 2015; Lopez-Garcia and Puente, 2012; Hölzl, 2014; Daunfeldt and Halvarsson, 2015; Coad and Srhoj, 2019)). Finally, we find that firms are less likely to exit following a HGF phase compared to other firms (e.g., (Acs et al., 2009; Choi et al., 2017; Mohr et al., 2014)), suggesting that exits following the HGF phase are unlikely to affect our results about longer-term dynamics.

3.2 Size growth and productivity growth at the firm level

Table 4 provides an initial view of size and TFP of differently defined HGFs at the firm level, and Table 8 repeats the same exercise with labor productivity. The tables show HGF performance four years before ($t-7$), one year before ($t-1$), four years after ($t+4$), and seven years after ($t+7$) the beginning of the high-growth phase. The premia are expressed as the (unweighted) average number of employees and total factor productivity level of HGFs relative to that of the average firm (scaled to be 100%).[13]

Starting with the relative/absolute dichotomy, we find that the two OECD definitions identify firms that are initially ($t-1$) considerably smaller than the average firm. On average, these HGFs employ 25-27% fewer employees before their high-growth phase than the average firm, and 55-156% more right after that (in $t+4$). In contrast, using the Birch definition, HGFs are 122-160 percent larger than the average firm even before the high-growth phase, increasing their premium to 372-472 percent right after the high-growth phase.

Irrespective of the definition used, HGFs create a substantial number of jobs. As expected, average employment growth during the HGF period is somewhat higher under the employment-based than under the sales-based definition. Using employment-based definitions, relative to the average firm, the average HGF's employment grows from $t-1$ to $t+4$ by 184 percentage points based on the OECD definition, and by 350 percentage points based on the Birch definition. Using sales-based definitions yields smaller results, but the magnitudes are still noteworthy: 80 and 212 percentage points according to the OECD and the Birch definition, respectively. The large growth of initially larger Birch HGFs hints at a larger potential contribution of these firms.

Regarding productivity, apart from the OECD sales definition, HGFs are initially (i.e., in $t-1$) 21-35% more productive than the average firm.[14] This suggests a potentially positive reallocation effect for these three groups of HGFs.

Importantly, during the high-growth phase (approximated by the change between $t-1$ and $t+4$), sales-based HGFs experience a productivity increase (by 13-18 pp), while the productivity of input-based HGFs falls (by 20-23 pp). This suggests that the within contribution of output-based HGFs is likely to be positive, while that of input-based HGFs may be negative.

A key message of this table is that there seems to be some trade-off between job creation and productivity growth, but this is not very strong: employment-based HGFs create more jobs while sales-based HGFs increase their productivity more. However, the difference between the two groups in terms of productivity growth is much more characteristic than their difference in job creation. Sales-based HGFs increase their productivity substantially and create many jobs at the same time, while employment-based HGFs face declining productivity.

Table 4 also allows us to follow the firms before and after their high-growth phase. Compared to the developments during the high-growth period, the

[13] For brevity, here we only provide these statistics for selected years, rather than for all of them. For example, the row 2000 represents firms that grew fast between 2000 and 2003. We focus on $t-1$ and $t+4$ rather than t and $t+3$ to clean the results from potential regression to the mean bias. The results are very similar when considering t and $t+3$. Note that the table includes employment and TFP levels, rather than growth rates.

[14] This is in line with Moschella et al. (2019), who find for Chinese firms that firms with higher productivity are more likely to become HGFs.

Table 4 Employment and TFP before and after the high-growth phase by HGF definition

Panel A: OECD3 (emp)								
	employment				TFP			
year	t-4	t-1	t+4	t+7	t-4	t-1	t+4	t+7
2000		68.2%	238.1%	256.7%		116.4%	96.7%	96.7%
2003	74.1%	77.3%	260.2%	251.5%	123.6%	123.6%	97.4%	106.1%
2006	67.0%	72.4%	268.2%	235.0%	99.2%	118.6%	103.3%	102.0%
2009	64.7%	76.5%	258.8%		112.8%	125.0%	101.1%	
2011	66.9%	69.3%			128.9%	131.7%		
average	68.1%	72.7%	256.3%	247.8%	116.1%	123.0%	99.6%	101.6%
Panel B: OECD3 (sales)								
	employment				TFP			
year	t-4	t-1	t+4	t+7	t-4	t-1	t+4	t+7
2000		74.7%	152.0%	161.6%		91.7%	114.5%	109.3%
2003	73.7%	75.1%	156.5%	158.1%	95.5%	101.6%	113.2%	120.7%
2006	67.7%	73.6%	161.9%	167.9%	89.5%	96.8%	114.7%	107.0%
2009	69.3%	76.5%	148.7%		91.5%	98.4%	120.1%	
2011	69.3%	73.4%			103.2%	100.8%		
average	70.0%	74.7%	154.8%	162.6%	94.9%	97.8%	115.6%	112.3%
Panel C: Birch3 (emp)								
	employment				TFP			
year	t-4	t-1	t+4	t+7	t-4	t-1	t+4	t+7
2000		194.1%	559.9%	617.6%		113.6%	97.5%	94.5%
2003	221.2%	212.7%	609.7%	619.4%	117.2%	129.7%	96.8%	106.3%
2006	185.2%	189.7%	525.9%	505.6%	111.7%	121.1%	106.9%	105.4%
2009	233.5%	246.9%	592.9%		112.4%	116.5%	104.1%	
2011	303.9%	268.6%			110.2%	123.0%		
average	235.9%	222.4%	572.1%	580.9%	112.9%	120.8%	101.3%	102.1%
Panel D: Birch3 (sales)								
	employment				TFP			
year	t-4	t-1	t+4	t+7	t-4	t-1	t+4	t+7
2000		242.4%	488.3%	575.4%		135.6%	157.0%	133.7%
2003	236.9%	233.4%	499.6%	540.7%	134.6%	142.7%	139.6%	141.4%
2006	169.3%	188.8%	396.1%	450.9%	119.7%	131.7%	139.9%	131.8%
2009	271.2%	293.0%	502.2%		135.8%	124.3%	153.6%	
2011	400.4%	342.7%			129.5%	140.1%		
average	269.5%	260.1%	471.6%	522.4%	129.9%	134.9%	147.5%	135.6%

This table shows HGF performance one year before (t-1) and four years after (t+4) the beginning of the high-growth phase, as well as an additional 3 years before (t-4) and after that (t+7).
The premia are expressed as the (unweighted) average number of employees and total factor productivity of HGFs relative to that of the average firm (scaled to be 100%)

productivity and employment changes before and after that are relatively small. Still, the high-growth phase seems to be typically preceded by productivity growth (growing TFP between $t-4$ and $t-1$) for all the four definitions, although to a different extent. On the contrary, the average productivity levels are very similar right after the end of the growth period ($t+4$) and three years afterward ($t+7$) and in some cases they become even lower. There is no further productivity increase after the end of the high-growth phase.

The fact that these pre- and post-trends in terms of both employment and productivity are relatively modest compared to changes during the high-growth period suggests that high growth tends to be a transitory phase in most firms' life.[15] From an empirical point of view, this finding suggests that by focusing on the high-growth period one is likely to capture the bulk of HGFs' productivity contribution. We use a more formal event study methodology in Section 4.2 to quantify HGFs' contribution before, during, and after the high-growth period.

4 HGFs' contribution to total productivity growth

While the previous section has provided descriptive evidence for HGFs' productivity and employment growth patterns, here we go deeper, focusing explicitly on HGFs' contribution to productivity growth, which is the main interest of this paper.

In the following, we rely on the decomposition methodology presented in Section 2.2. Our unit of observation is a cohort (c) in industry (j). We denote the base year of cohort c by t_0^c and firms are considered as part of the cohort if their annual growth rate between t_0^c and t_0^c+3 was above the threshold prescribed by the relevant definition. The productivity contribution of cohort c in industry j between t_0^c and t_0^c+3 will be denoted by $cont_{jc}$. These objects, which are defined at the industry-cohort (year) level, will be our units of observation.

Table 5 Decomposing the HGF contribution to TFP growth

Panel A: All firms				
	HGF within	HGF between	HGF cross	HGF total
OECD (emp)	-0.09%	2.56%	-1.90%	0.58%
OECD (sales)	2.32%	2.15%	-1.53%	2.94%
Birch (emp)	0.01%	3.24%	-2.19%	1.05%
Birch (sales)	3.06%	2.78%	-1.91%	3.93%
Panel B: Young firms				
	HGF within	HGF between	HGF cross	HGF total
OECD (emp)	-0.23%	1.70%	-1.29%	0.18%
OECD (sales)	1.41%	1.35%	-0.88%	1.88%
Birch (emp)	-0.19%	2.33%	-1.58%	0.57%
Birch (sales)	2.34%	1.96%	-1.34%	2.96%

This table presents the employment weighted average (across year-cohorts) of the different components of the total HGF contribution for the four definitions. All numbers are measured in percentage points of the total industry productivity growth. Panel A considers all the firms and Panel B includes only young (i.e., less than 5 years old) HGFs. Therefore, OECD (emp) HGFs contribute on average 0.58 pp. to industry-level productivity growth, from which 0.18 pp. is contributed by young HGFs

4.1 How much do HGFs contribute during their high-growth phase?

Let us start with the overall average of HGF contributions across industries and cohorts, $cont_{jc}$ during their high-growth phase. Table 5 shows that the total HGF contribution differs strongly across definitions. The table shows the contribution in percentage points of the total industry-level productivity growth, i.e., OECD (emp) HGFs contributed 0.58 pp on average. In general, it is larger for sales-based than for

[15] Previous research already provides evidence for the temporary nature of the high-growth phase (see e.g., Acs et al. (2009) using US data, Daunfeldt and Halvarsson (2015) using Swedish data or Hölzl (2014) using Austrian data). Delmar et al. (2003) show that high employment or sales growth comes from a single year in 22% and 27% of all Swedish HGFs respectively. On the other hand, Coad and Srhoj (2019) find that Croatian and Slovenian firms with higher previous employment growth become HGFs with a higher probability. Additionally, Lopez-Garcia and Puente (2012) demonstrate that more than half of Spanish HGFs were already HGFs in the previous period.

employment-based definitions and it is also larger for Birch definitions than for OECD ones. Birch-sales HGFs contribute the most on average (3.93 percentage points), followed by OECD sales HGFs (2.94 pp.). The contribution of employment HGFs is substantially lower, with 0.58 pp. for the OECD and 1.05 pp. for the Birch definition. The difference between the Birch and OECD definitions is likely to result from the larger size of Birch firms as well as from the fact that it is a stricter HGF definition than the OECD one.

The productivity decomposition exercise reveals the source of these differences. First, the within contribution of the sales-based definitions is strongly positive while that of the employment-based definitions is small and rather negative. This comes from the fact that productivity typically increases (or is stable) for sales-based definitions while it falls for the employment-based definitions (see Table 4). The average *between* effect is positive for all the definitions, showing that HGFs are typically more productive than their peers. There is some variation, mainly resulting from the differences in initial productivity advantage and in the employment growth rate during the high-growth period. Finally, the *cross* contribution is negative for all definitions. This reflects that overall, there is a negative correlation between size and productivity growth, as we will discuss in the next section.

A policy-relevant insight from these patterns is that output HGFs, which contribute more than input HGFs, typically contribute via their *within* term, while their initial productivity level is not especially high (see Table 4). Therefore, HGF policies that only target firms with high initial productivity levels – in the implicit hope of a strong within contribution – will capture potential input HGFs and miss potential output HGFs. Similarly, HGF policies centered on promoting employment growth rather than output growth may capture firms that contribute less to productivity growth.

While these absolute numbers are of some interest, it is probably more relevant to ask whether HGFs contribute more than the average firm. In other words, whether the share of HGFs' contribution is larger than their (initial) share in terms of inputs or outputs. Figure 1 illustrates this relationship at the industry level, with the average initial employment share of HGF cohorts on the horizontal axis and the average ratio of total HGF contribution to industry-level TFP growth on the vertical axis.[16]

Again, there is a quite obvious dichotomy between employment- and sales-based HGF definitions. Sales-based HGFs tend to contribute substantially compared to their initial share. The initial employment share of OECD (Birch) sales HGFs is 7.1% (18.3%) on average while their share in TFP growth is 24.7% (28.3%). In contrast, employment-based HGFs often contribute negatively and tend to be below the 45-degree line. The initial employment share of OECD (Birch) employment HGFs is 3.1% (14.7%) on average while their share in TFP growth is 0.2% (3.1%).[17]

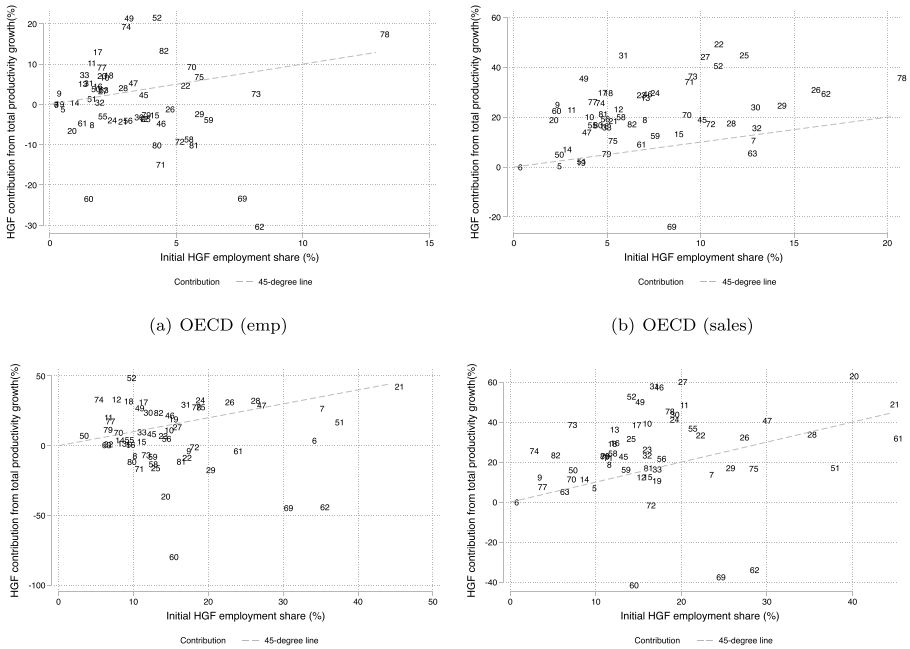
The average effects in Table 5 hide a large degree of heterogeneity. Figures 2 and 9 (the latter for labor productivity) focus on the heterogeneity of $cont_{jc}$. Clearly, the HGF contribution is not necessarily positive for either of the definitions. The 10th percentile is negative for all four definitions. On the other extreme, in some industries a very substantial part of productivity growth results from the activities of HGFs. For example, in 20% of the industries with positive productivity growth, more than half of total productivity growth is contributed by sales-Birch HGFs.

The distributions provide a more complex picture than the means in Table 5. Within the OECD/Birch dimensions, the difference in means clearly transforms into stochastic dominance. The distribution of OECD (sales) firms' contribution dominates that of OECD (emp) firms and the same is true for the Birch definitions. The variance of the Birch definitions is substantially higher than that of the OECD definitions, with both having thicker tails than the two OECD definitions. Therefore, while the mean and median of the OECD (sales) distribution are larger than that of the Birch (emp) distribution, the top percentiles of the Birch (emp) distribution are larger. Clearly, idiosyncratic characteristics of some of the larger firms captured by the Birch definitions can have very substantial effects on industry productivity, with a strong upside, as shown by the very thick upper tail of the Birch

[16] Here we omit observations where industry-level productivity growth was negative, because in such cases the ratio is hard to interpret.

[17] We find similar patterns when we replace the initial employment share with the initial sales share.

Fig. 1 The initial employment share of HGFs and their TFP growth contribution as a share of industry-level TFP growth.



(a) OECD (emp)

(b) OECD (sales)

(c) Birch (emp)

(d) Birch (sales)

This figure shows the relationship between the initial employment share of HGFs and their productivity contribution during their high-growth period. These numbers are averaged across cohorts for all 2-digit industries, and we omit year-industry combinations in which industry productivity growth was negative. The dashed line is the 45-degree line.

(sales) distribution. In contrast, the OECD (sales) distribution is less dispersed, with few negative and some decent-sized contributions.

This section has documented several patterns about HGFs' contribution during their high-growth phase. However, they may also contribute in other periods. The next section investigates this topic with an event study approach.

4.2 The contribution of HGF cohorts over time

While the lack of average pre- and post-trends presented in Table 4 is suggestive of HGFs' productivity and size growth being concentrated in the HGF phase, we use an event study methodology to investigate whether these findings can be generalized to the productivity contribution itself.[18]

For the event study, we extend our previous notation as we need to follow each cohort (c) for a number of periods (p).[19] Here we denote contributions by $cont_{cjp}$. To present the results, we run event study-type regressions at the industry-cohort-period level of the form:

$$cont_{cjp} = \sum_{\theta} \delta_{cp}^{\theta} + \eta_j + \xi_p + \epsilon_{cjp} \tag{1}$$

where c indexes cohorts, j industries and p periods. δ_{cp}^{θ} are event time dummies taking the value of 1 when the $t_p - t_0^c = \theta$, where t_p is the beginning of period p and t_0^c is the initial year for cohort c. The coefficients of these event time dummies show the difference between the actual contribution of the cohort relative to its contribution in the base period, $t_0^c - 6$ (which is the omitted period).[20] η_j and ξ_p are 2-digit industry and period fixed effects,

[18] The results in Table 4 do not guarantee this, because the contribution depends on firm-level correlations which are not equivalent to aggregate averages. Additionally, our analysis here controls for industry heterogeneity, filtering out composition effects, and it also allows us to quantify statistical uncertainty.

[19] In the event study, we use periods $[t_0^c - 6;t_0^c - 3]$, $[t_0^c - 3;t_0^c]$, $[t_0^c;t_0^c + 3]$, $[t_0^c + 3;t_0^c + 6]$, $[t_0^c + 6;t_0^c + 9]$.

[20] We choose $t_0 - 6$ for the omitted period because this is the earliest period which we can use, this ensures that any pre-trends before the high-growth phase are clearly visible.

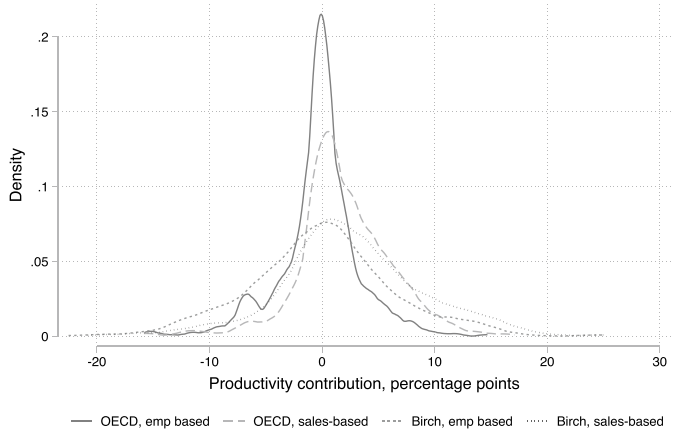


Fig. 2 Productivity contribution of HGFs by industry-cohort.

This figure shows the kernel distribution of the productivity contribution of HGF cohorts at the 2-digit industry-cohort level between t_0 and $t_0 + 3$, winsorized at the 5th and 95th percentiles. The horizontal axis shows the HGF contribution in log percentage points during the 3-year period.

respectively. We weigh the regressions by the number of employees of the industry at the beginning of the period.

Figure 3 shows the results for TFP and Fig. 8 for labor productivity. The horizontal axis represents the event time, and the vertical axis shows the point estimates and the standard errors of the event time dummies. Zero is the contribution in the base period, starting at $t_0 - 6$.[21] We can draw two main conclusions. First, relative to the omitted period, we find positive contributions during the high-growth period $[t_0;t_0 + 3])$ for the sales-based definitions and negative contributions for the employment-based definitions. The additional positive contributions of the sales-based HGFs during the high-growth period are substantial: OECD (sales) HGFs contribute by 2.3 percentage points while Birch (sales) HGFs by 2.8 percentage points.[22] This pattern contrasts strongly with that of the employment HGFs, which report a negative contribution during their high-growth phase compared to their contribution in the base period.

Second, regarding timing, HGFs tend to contribute unusually during their high-growth period, between t_0 and $t_0 + 3$. We find evidence for a positive contribution before the high-growth phase for the Birch (sales) HGF, which is about half as large as the one observed during their high-growth phase.[23] This suggests that high sales growth of larger firms often takes place after a period when both productivity and size are growing. Importantly, in line with our earlier observations in Section 3.2, we do not find evidence for extraordinary contributions after the high-growth phase for any of the HGF definitions. This latter observation implies that the long-run productivity contribution of HGFs is likely to be similar to their short-term contribution.[24]

4.3 Industry characteristics and HGF contribution

Motivated by the substantial heterogeneity across industries and periods, as documented by Fig. 2, this section investigates some of the industry characteristics which may influence the role HGFs play

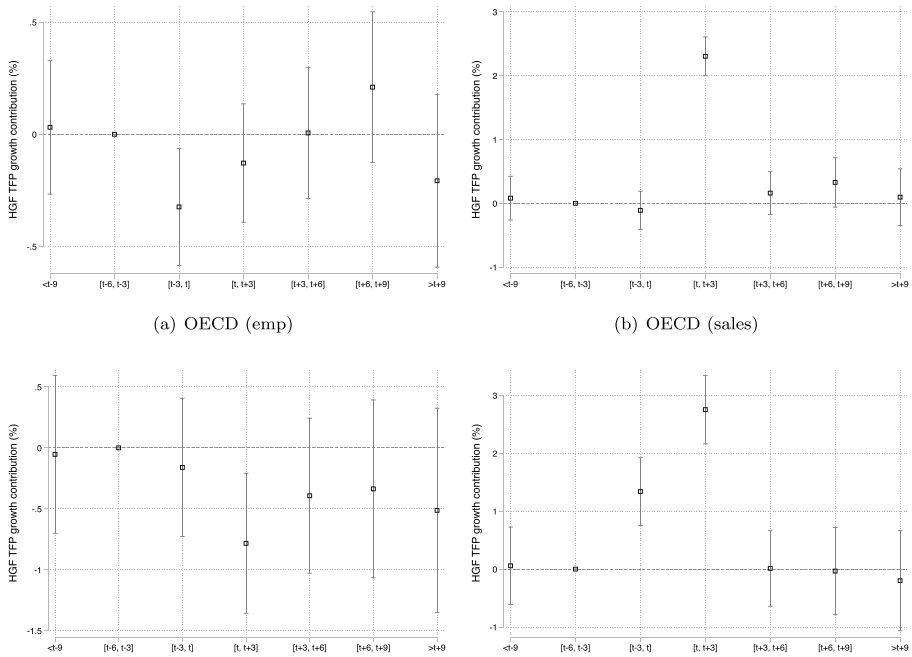
[21] The event study shows whether the contribution is higher relative to this base period. We can only estimate the relative contribution due to the set of fixed effects.

[22] The main difference between these numbers and those presented in Table 5 is that in the table we show the "raw" contribution, while the event study shows the extra contribution in the high-growth period of the cohort relative to the base period. Considering that the OECD (Birch) HGFs base contribution is 0.95 (1.4) pp., the two numbers become quite similar. The remaining differences are explained by the presence of other controls in the event study regressions.

[23] We also find evidence for a negative pre-high growth contribution for OECD (emp) HGFs.

[24] The main patterns are similar for labor productivity (see Fig. 8 in the Appendix), but we can also observe a small positive contribution during the high-growth phase for the employment-based definitions, and also a significant positive contribution of employment-based OECD HGFs between $t_0 + 6$ and $t_0 + 9$, suggesting that taking into account changes in the capital stock matters to some extent.

Fig. 3 Total TFP contribution: event study.



(a) OECD (emp)

(b) OECD (sales)

(c) Birch (emp)

(d) Birch (sales)

This figure shows the results of an event study regression in which one observation is the productivity contribution of a cohort of HGFs in a 2-digit industry in different event-times before and after the high-growth period. HGFs are firms that are in their high-growth phase in the 3-year period starting with year t. The vertical axis shows the productivity contribution of HGFs in percentage points, normalized to zero two periods before the high-growth phase, i.e., in $[t-6, t-3]$.

in productivity growth.[25] We use two approaches. First, based on our decomposition, we focus on the dynamic parameters which show up directly in the decomposition: the strength of reallocation and the correlation between size and productivity growth. The second approach, instead, focuses on more fundamental structural parameters, such as the technology of the industry, concentration and the share of young firms within the industry.

Dynamic correlations Our first approach starts from the decomposition in Section 2.2. The first relevant correlation is the *strength or efficiency of reallocation,* ρ^{real}. This captures the extent to which more productive firms expand in terms of their employment or market share faster, represented by $corr(prod_{it_0}, \Delta\theta_{it})$. Clearly, this determines how likely do HGFs come from more productive firms. This 'selection effect' has an impact on the between term, $\sum_{i \in HGF}(prod_{i,t_0} - PROD_{t_0})\Delta\theta_{i,t}$, by showing the extent to which HGFs are initially more (or less) productive than the average firm. We expect a stronger reallocation to be positively associated with the between components and, in turn, with productivity growth.

A second key measure is the extent to which size growth is accompanied by productivity growth $corr(\Delta prod_{it}, \Delta\theta_{it})$, which we call the *cross-correlation,* ρ^{cross}. This correlation captures a key property of industry dynamics: whether firms improving their productivity can acquire sufficient resources rapidly enough to also expand on the market in the same period typically considered in HGF definitions. The cross-correlation is the fundamental behind both the HGF within and cross terms. While the cross-correlation shows up directly in the cross term, it also affects the within term positively, since with a stronger

[25] 2-digit industry fixed effects explain 10-11% of the variance in HGF contributions for the OECD definitions and 18% for the Birch definitions.

cross-correlation HGFs are more likely to improve their productivity. Empirically, we define these correlations for industry j in year t based on levels in year t and changes between t and $t+3$. Therefore, the efficiency of reallocation for firms i in industry j is $\rho_{jt}^{real} = corr(prod_{it}, \theta_{i,t+3} - \theta_{i,t})$. Similarly, the cross-correlation is $\rho_{jt}^{cross} = corr(prod_{it+3} - prod_{it}, \theta_{i,t+3} - \theta_{i,t})$. In our data, the reallocation correlation is positive, with an average value of 0.14. The cross-correlation is typically negative, with an average of -0.1, i.e., 3-year size growth is negatively correlated with productivity growth during the same period.

We run regressions to investigate how these correlation parameters are related to the industry-level HGF contribution. The dependent variable is the HGF contribution of cohort c in industry j during its high-growth period, between t_0^c and $t_0^c + 3$. The main independent variables are dummies showing the quartile (q) of the industry in the overall ρ^{real} and ρ^{cross} distributions while controlling for year fixed effects (δ_t).[26] Therefore, we estimate the following equation:

$$cont_{jc} = \sum_q \beta_q^{real} (\rho^{real})_{jt_0^c}^q + \sum_q \beta_q^{cross} (\rho^{cross})_{jt_0^c}^q + \delta_t + \epsilon_{jc} \tag{2}$$

We weight the regression by industry employment and calculate standard errors clustered at the industry level.[27]

Regression results are presented in Fig. 4 and Panel A of Table 9. We find that HGF contribution increases in these industry dynamics parameters for all four definitions. Clearly, stronger industry dynamics both in terms of reallocation and cross-correlation are associated with higher TFP contributions. The difference in terms of contribution between the industries with the weakest and the strongest reallocation is between 1.7 and 3.4 percentage points, depending on the definition. This is quite large given that the average contribution is between 0.6-3.9 pp. for the various definitions (Table 5). The strength of reallocation matters more for the Birch definitions. The relationship between the cross-correlation and the HGF contribution is similar, with 2.8-7.1 pp. difference between industries with high vs. low cross-correlations.

Panel B of Table 9 presents estimation results of Regression 2 using labor productivity to define the dependent variable. We find patterns similar to TFP, suggesting that these patterns are not sensitive to the definition of productivity.

Figure 10 decomposes the relationship found in Fig. 4 by re-running Regression 2 with HGF within, between and cross contributions as dependent variables. In line with the arguments stated at the beginning of this section, we find that ρ^{real} mainly affects the between component by making more productive firms more likely to become HGFs. In contrast, ρ^{cross} is related to the within and cross components.

Fundamental industry characteristics While the previous results confirm that the reallocation and cross-correlations are significantly related to HGF contribution, these correlations are not easy to observe or interpret. Therefore, our second approach investigates the extent to which one is able to predict HGFs' contribution to productivity growth based on more fundamental industry parameters.

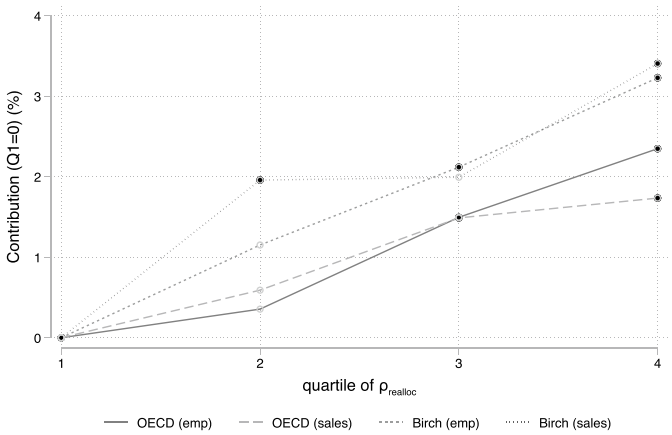
We modify the basic specification in Eq. (2) including several basic industry features measured in t_0 instead of reallocation and cross-correlation quartiles as explanatory variables. We capture market structure by concentration (C5) and by the standard deviation of TFP. We also include the share of young firms (with age of at most 5 years) which, as Panel B of Table 5 shows, tend to contribute substantially to productivity growth, accounting for 31-75% of the total HGF contribution.[28] Finally, we add the technology/knowledge intensity classification of the industry based on Eurostat's categorization. The base category

[26] We calculate ρ^{real} and ρ^{cross} quartiles based on the full distribution of the industry-year level observations to capture the "absolute" value of the correlations rather than their ranking in the given year.

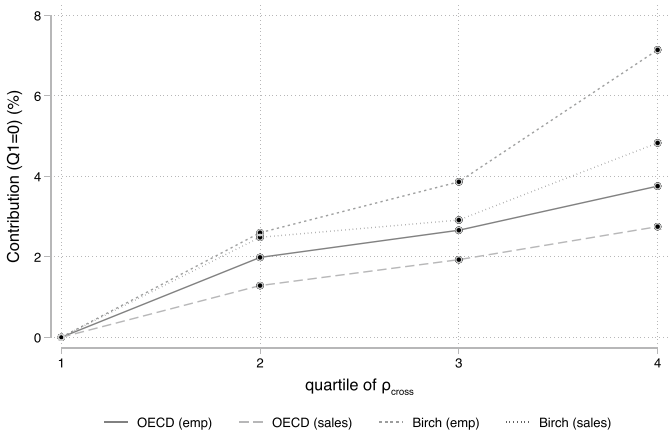
[27] These results are robust to several changes such as (i) including the productivity growth of non-HGFs as an explanatory variable and (ii) adding the means, standard deviations, and correlations of productivity, productivity growth, size and size growth.

[28] Both Panel A and Panel B are measured in percentage points of industry-level productivity growth, therefore the two are directly comparable. For example, all OECD (emp) HGFs contribute 0.58 pp on average, from which 0.18 pp is contributed by young OECD (emp) HGFs.

Fig. 4 The relationship between industry dynamics parameters and HGF contribution



(a) $\rho^{realloc}$



(b) ρ^{cross}

This figure shows point estimates from Equation 2, in which the dependent variable is the HGF contribution to productivity growth of a HGF cohort of an industry in a 3-year period, while the explanatory variables are quartiles of ρ^{real} and ρ^{cross}. The different lines show the point estimates for the four different HGF definitions. Panel A presents the coefficients of the different quartile dummies for $\rho^{realloc}$, relative to the first quartile (base category). Panel B presents similar results for ρ^{cross}. The full circles show that the coefficient is significant at the 10% level. For example, the rightmost point corresponding to Birch (emp) HGFs in panel A shows that such HGFs in an industry which has a high (top quartile) $\rho^{realloc}$ contribute about 2.9 pp. more to industry-level TFP growth compared to an industry with a low (bottom quartile) $\rho^{realloc}$, and this difference is significant at the 5% level. The corresponding regression results are presented in Table A3, Panel A. Year dummies are included, and standard errors are clustered at the industry level.

is primary industries, and we distinguish between high-tech manufacturing, low-tech manufacturing, knowledge-intensive services (KIS), and less knowledge-intensive services (LKIS).[29]

The results of this exercise are reported in Table 6. Column (1) reports the baseline specification. We find that the HGF contribution is larger in industries with more young firms: a 10 pp. increase in the share of young firms is associated with a 1.5 pp. higher HGF contribution. Services and manufacturing industries also seem to differ to some extent, with knowledge-intensive services featuring a lower HGF contribution than other industries. Concentration does not seem to matter. Importantly, the explanatory power of the regression is quite moderate, at 15%, showing that these industry features have limited explanatory power in terms of HGF contribution to productivity growth.[30]

In column (2) we also include the TFP growth contribution of non-HGFs which is likely to capture industry-year level shocks to productivity, a potential omitted variable. The patterns remain similar, though the significance levels are somewhat smaller.

To sum up, we have found that stronger reallocation, measured by industry-level correlations, is positively related to the contribution of high-growth firms. These correlations, however, may be hard to observe for policymakers and may change frequently. Basic industry characteristics predict the HGF contribution only to a limited extent, but we have found evidence for the share of young firms being correlated with the HGF contribution.

5 A benchmark model for HGF contributions

After we have examined the patterns of HGF contributions to aggregate productivity, in this section we ask whether the patterns we find are "surprising" or can be predicted by a simple benchmark model. If the benchmark model of firm dynamics reproduces most findings, the simple framework can help economists or policymakers think about HGF contributions.[31]

Our simple simulation starts from a normal joint distribution of productivity, size (input), productivity growth, and size (input) growth. This distribution is easy to be characterized by the correlations between these four variables, including the strength of reallocation and cross-correlations. HGFs are defined as firms growing strongly either in terms of their input use or in terms of output.

We find this approach useful for several reasons. First, we show that our observations, at least qualitatively, are indeed easily predicted by this benchmark. As an important consequence, the average HGF contribution is determined by overall industry dynamics to a large extent rather than by idiosyncratic dynamics of specific firms. Second, our framework supplies policymakers with a few intuitive and measurable concepts and parameters affecting the productivity contribution of HGFs, such as whether HGFs are defined based on their inputs or outputs, the strength of reallocation, and the cross-correlation in the industry. Taking into consideration these concepts can help in designing effective policies. Finally, as we illustrate through the example of SMEs in Appendix Section 2, the same logic can be applied to thinking about the productivity growth contributions of other groups of firms defined by some easily observable characteristics.

Section 5.1 describes our simple approach and Section 5.2 illustrates the difference between input- and output-based HGFs in this framework. We provide more details in 4, where Appendix Section 1 shows how the simulation reproduces the qualitative patterns found in Section 4.3 and Appendix Section 2 presents how the same simulation framework can be used to model the contribution of other groups of firms via the example of SMEs.

[29] The Eurostat industry classification is available at http://ec.europa.eu/eurostat/cache/metadata/Annexes/htec_esms_an3.pdf. Our high-tech manufacturing category is composed of the high-tech and medium-high tech Eurostat categories, while our low-tech manufacturing category includes the medium-low- and low-tech Eurostat categories.

[30] Including 2-digit industry dummies increases the explanatory power to 0.35.

[31] Whether simple models can reproduce empirical patterns has been investigated in the trade context by Armenter and Koren (2014).

Table 6 HGF contribution to TFP growth and industry characteristics

LHS: Total HGF contribution	(1)	(2)
C5	0.00111	0.00442
	(0.0228)	(0.0180)
TFP (s.d.)	0.0129	0.00474
	(0.0259)	(0.0188)
Share young	0.149**	0.112*
	(0.0612)	(0.0598)
High-tech mfg.	0.0113	0.0148
	(0.0154)	(0.0122)
Low-tech mfg.	-0.00468	0.00434
	(0.00984)	(0.00775)
KIS	-0.0341**	-0.0250*
	(0.0167)	(0.0127)
LKIS	-0.00692	0.00681
	(0.0141)	(0.00999)
non-HGF contr.		0.183***
		(0.0308)
Observations	728	728
R-squared	0.150	0.285
Year FE:	Yes	Yes

These regressions show how industry characteristics are related to the contribution of HGFs to TFP growth. One observation is a 2-digit industry-year combination, where the dependent variable is HGF TFP contribution between t_0 and $t_0 + 3$ and the explanatory variables are measured in t_0. C5 is the revenue share of the 5 largest firms, TFP (s.d.) is the standard deviation of the TFP distribution, share young is the share of firms that were at most 5 years old in t_0 (between 0 and 1). We use the Eurostat categorization to classify industries. The base category is primary industries, 'mfg.' stands for Manufacturing, 'KIS' for knowledge-intensive services and 'LKIS' for less knowledge-intensive services, 'non-HGF TFP cont.' is the TFP growth contribution of non-HGF firms. All regressions include year fixed effects. The regressions are employment-weighted, and the standard errors are clustered at the 2-digit industry-level

5.1 Method: simulation

Based on the decomposition framework, we simulate data in 4 dimensions: (ln) initial input level, (ln) initial productivity level, input growth, and productivity growth. The growth rates are interpreted as the 3-year growth rates used in the empirical exercise. We simulate a 4-dimensional normal distribution[32] for these variables with different correlation structures. Note that these include the two key correlations analyzed in Section 4.3: the strength of reallocation (the correlation between productivity level and subsequent input growth) and the cross-correlation (the correlation between input growth and productivity growth). As a baseline, we calibrate these correlations from the Hungarian data.[33] To keep things as simple as possible, we think of productivity as simply dividing output produced with the amount of inputs used. Therefore, $output = input \times productivity$.[34] As a result, output growth can be written as: $\Delta output = \Delta productivity + \Delta input$, where Δ is interpreted as a (log) percentage change.

We calculate aggregate productivity growth on these simulated data and decompose it with the method described in Section 2.2. We define HGFs based on both input and output growth, as this has proved to be the key distinction in Section 3.1.[35] To make the results comparable across specifications, we define input HGFs as firms with the top 5% input growth. Output HGFs are the top 5% in terms of their $\Delta productivity + \Delta input$ value. We simulate 20,000 firms each time.

[32] In reality, these variables are not normally distributed. Both log productivity and productivity change have fat-tailed distributions (Kang , 2017). Size distribution is highly skewed with a very large number of small firms (De Wit , 2005). Our data reject null-hypothesis in tests for both univariate and joint normality (Doornik and Hansen , 2008). However, we still choose the normal distribution for two reasons. First, its simple correlation structure allows us to have parameters that all have straightforward economic meaning even though we have four underlying variables. Second, the analysis aims to explain qualitative rather than quantitative features of what we observe. These depend on the sign of the correlations rather than more intricate features of the distributions. Therefore, we expect our results to generalize to other distributions.

[33] The correlations are the overall correlations between ln(TFP), market share, and their 3-year growth rates. Using labor productivity instead of TFP or ln(employment) instead of the market share yield similar correlations.

[34] This is clearly a simplification compared to our empirical exercise, where, even in the case of labor productivity, we use value added rather than sales in the calculations. However, introducing more dimensions and correlations would not change the insights from the framework, but would make it substantially less transparent.

[35] We focus on the difference between input and output HGFs and ignore the difference between OECD and Birch definitions because that proved to be most important empirically. We fix the share of HGFs because it allows us to ignore (normalize) parameters affecting average growth.

5.2 Input and output HGFs

In this section we illustrate the differences between input and output HGFs predicted by our simple model in three steps and contrast them with our previous empirical results. First, related to the between term, we investigate how productivity is related to the probability of becoming a HGF. In our empirical exercise, we have seen that input-based HGFs are on average more productive at the beginning of their HGF period. Second, we show the difference between the two types of firms by presenting the joint distribution of size and productivity growth and the within contribution of the two groups. In the empirical part, we found that output-based HGFs are the ones that tend to increase their productivity throughout the HGF period. Third, we simulate the distribution of the two firm groups' total contributions. In the previous empirical exercise, we found that output-based HGFs have a higher contribution on average.

We illustrate the difference between input- and output-based HGFs with the simple simulation described in the previous section. Let us start with reallocation. Figure 5 depicts the share of HGFs within 20 productivity quantiles using parameters partly calibrated from the Hungarian data.[36] The share of *input-based* HGFs is positively associated with productivity, mainly due to the positive reallocation correlation: more productive firms are more likely to increase their inputs. This implies that reallocation to input HGFs indeed contributes positively to industry-level productivity growth, represented by a positive reallocation term.

In contrast, and possibly more surprisingly, the share of *output-based* HGFs is decreasing in productivity. By our definition, output-based productivity growth can either result from rapid input growth and/or strong productivity growth. While more productive firms are more likely to increase their size (as witnessed by the positive reallocation correlation), less productive firms are more likely to experience productivity growth, as evidenced by the *convergence correlation*, $corr(\Delta prod, prod)$, which has an average value of -0.41.[37] Given its larger absolute value, convergence correlation dominates the reallocation correlation in the relationship between productivity and the share of output HGFs, leading to a negative overall relationship.

These findings are clearly in line with the descriptive patterns reported in Table 4 for OECD HGFs[38]: output HGFs tend to be less productive initially than input HGFs – therefore the within contribution of the former is likely to be smaller.

Figure 6 illustrates the within contributions of the simulated input- and output-based HGFs. The cluster of points represents firms in the simulated industry, each point showing a firm's input growth (horizontal axis) and productivity growth (vertical axis) in a (3-year) period. The correlation between these two quantities, the cross-correlation, is slightly negative, calibrated from the Hungarian data. Input HGFs are to the right of a critical value (represented by the vertical dashed line), denoted by squares and triangles.

As we have already mentioned, output growth can be expressed in the model as $\Delta output = \Delta productivity + \Delta input$. Consequently, output HGFs are firms above a line where $\Delta productivity + \Delta input > c$, with c being a critical value. Such firms are located above a specific line with a negative slope and are denoted by diamonds or triangles. Firms that are both input and output HGFs are denoted by triangles.

The figure illustrates some of our earlier points. First, the intersection between input and output HGFs is not especially large (as seen in Table 11). Input HGFs rapidly increase their size but their productivity growth is rarely exceptional. Most output HGFs qualify because of their productivity growth but many of them do not expand too much in terms of their size, as we have shown in Table 4.

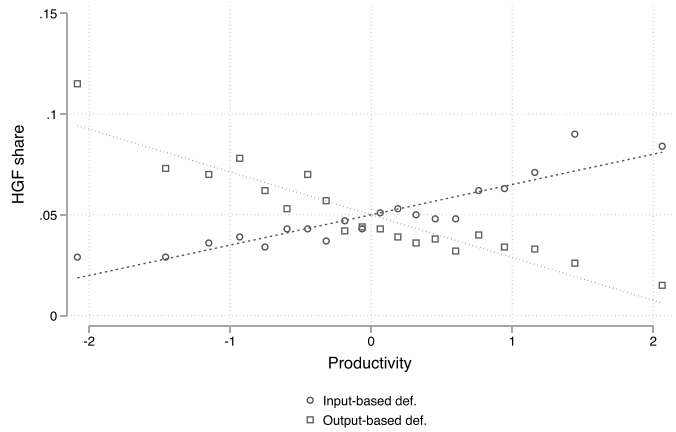
These patterns clearly explain the differences between the within contributions of input and

[36] We normalize each of the variables in such a way that their mean is zero and their standard deviation is 1. The correlations from our data are: $\rho_{real} = 0.1441$; $\rho_{cross} = -0.1$; $corr(\Delta prod, prod) = -0.41$; $corr(prod, inp) = 0.1418$; $corr(\Delta inp, inp) = -0.1852$; $corr(inp, \Delta prod) = 0.0011$.

[37] This correlation shows strong productivity convergence, or catch up of less productive firms, at the 3-year time period.

[38] This mechanism may be less relevant for the Birch HGFs, for which employment and sales HGFs are more similar initially because of the disproportionate share of larger and therefore less dynamic firms captured by this definition.

Fig. 5 Input- and output-based HGFs: reallocation effect



This figure shows the simulated relationship between initial productivity level and the share of input- and output- based HGFs. We normalize each of the variables in such a way that their standard deviation is 1. The correlations are: $\rho_{real} = 0.1441$; $\rho_{cross} = -0.1$; $corr(\Delta prod, prod) = -0.41$; $corr(prod, inp) = 0.1418$; $corr(\Delta inp, inp) = -0.1852$; $corr(inp, \Delta prod) = 0.0011$

output HGFs. The average productivity growth of input HGFs (denoted by HGF_{inp} on the vertical axis) is slightly negative because of the negative cross-correlation, which implies a negative within component for these firms. In contrast, average productivity growth of output HGFs (denoted by HGF_{out} on the vertical axis) is positive, and hence, the within component is strongly positive in this group.

We have presented intuition for the difference between input- and output-based HGFs in terms of their within contributions. Now we continue with simulating the distribution of the total contributions (similar to Fig. 2). In this figure, the distribution represents industry-years with different parameters and with different random components. Therefore, in our simulation, we also vary industry parameters across observations. Figure 7 shows this distribution with industry correlation parameters $(\rho_{real}, \rho_{cross}, corr(\Delta prod, prod), corr(prod, inp), corr(\Delta inp, inp), corr(inp, \Delta prod))$ chosen randomly from a uniform distribution.[39] This figure reproduces the main patterns of Fig. 2: first, the contribution of output-based HGFs clearly stochastically dominates that of input-based HGFs; second, the distribution corresponding to output-based HGFs is more asymmetric with a long right tail but very few negative contributions.

This section has shown that a simple model of firm input and productivity growth generates patterns that are qualitatively very similar to our empirical observations.

6 Conclusion

This paper investigates the role of HGFs in reallocation and industry-level productivity growth. We have shown that productivity contributions differ widely, ranging from a negative contribution in half of industry-period observations for the OECD (emp) definitions to HGFs accounting for more than half of the industry-level productivity growth in 20 percent of cases for the Birch (sales) definition. HGF definition clearly matters. The positive contribution is limited to the high-growth period or the one preceding it, but HGFs do not contribute positively to aggregate productivity afterwards. The size of their contribution is linked to reallocation but cannot be predicted very effectively with fundamental industry characteristics. Finally, a very simple simulation can reproduce most of these findings.

Our results provide a number of insights regarding high-growth firms. Many of the patterns documented at the firm level have fundamental consequences in

[39] Note that this uniform distribution is a meta-distribution of the moments of the 4-dimensional distributions.

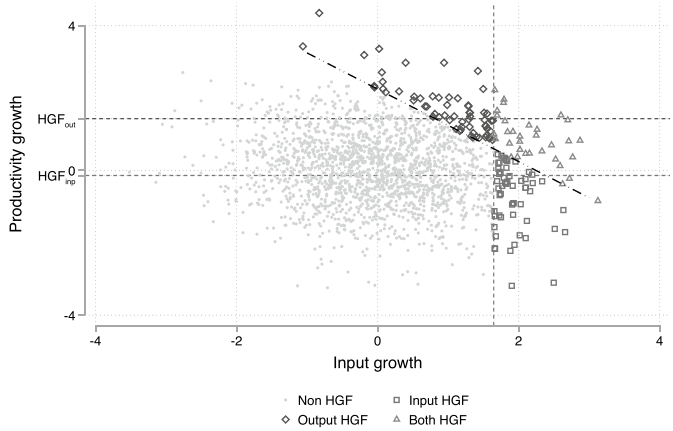


Fig. 6 Input- and output-based HGFs: within effect. This figure shows the scatter plot of input and productivity growth of firms in the simulated data.

This figure shows the scatter plot of input and productivity growth of firms in the simulated data. Input HGFs are those with input growth above a threshold, denoted by squares and triangles. Output HGFs are denoted by diamonds and triangles, thus the triangles are both input and output HGFs. HGF_{out} represents the average productivity growth of output HGFs and HGF_{inp} is the average productivity growth of input HGFs. The correlations are: $\rho_{real} = 0.1441$; $\rho_{cross} = -0.1$; $corr(\Delta prod, prod) = -0.41$; $corr(prod, inp) = 0.1418$; $corr(\Delta inp, inp) = -0.1852$; $corr(inp, \Delta prod) = 0.0011$. The mean of all variables is normalized to zero.

terms of industry-level "wider productivity effects". These links between the firm-level patterns and industry-level contributions are not automatic or trivial. Our measure captures both within-firm and reallocation effects, which often have opposite signs and therefore their net effect is not clear a priori.

One link between the firm-level results and industry-level contribution is the fact that employment growth and productivity growth are weakly or negatively correlated (Coad et al., 2014) and that initial productivity is not a strong predictor of high growth, probably because of inefficient reallocation (Grover Goswami et al., 2019) or weak selection in the framework of evolutionary theories (e.g. (Metcalfe, 1994)) are not too strong. This pattern implies two key trade-offs. First, as a result of the former correlation, the contribution of sales HGFs comes primarily from the within term, while that of the employment HGFs is dominated by the Schumpeterian effect. This fact, combined with the second correlation, which implies that the reallocation effect is generally weak, leads to much stronger productivity contribution by sales-based HGFs.

Another key link between the firm- and industry-level patterns is that the low persistence of firm-level HGF status seems to be associated with low persistence of the productivity contribution. Positive unusual contributions are concentrated at the high-growth period of the cohort. We find that size or productivity growth is typically not followed by further improvements in productivity at the cohort level. This finding is in line with the firm growth model of Penrose (1959), which emphasizes that size growth is often followed by reduced productivity and the cycle models (see e.g. (Greiner, 1989)), which predict a management crisis following growth periods.

A third link between firm- and industry-level patterns is that neither the HGF status nor the industry-level HGF contribution can be predicted with much confidence based on industry-level characteristics. Again, even though these are interlinked, this link is not automatic: even if the HGF status is not linked strongly to industry features, key industry correlations, such as the correlation between employment and productivity growth could be. This, however, is not the case, suggesting that targeting industries has limited rationale when productivity contribution is concerned.

We argue that these findings are of high policy relevance. The productivity evolution of high-growth firms is not only important because of HGFs' direct contribution to productivity growth but also because more spillovers may be expected from more productive firms (e.g., (Stoyanov and Zubanov, 2012)). The insight that sometimes HGFs contribute negatively to productivity growth should

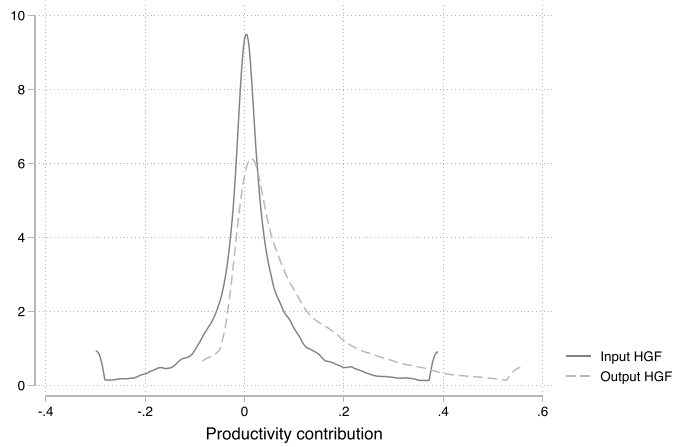


Fig. 7 Simulated productivity contribution of input- and output-based HGFs across industries with different dynamics

This figure shows the kernel density of the simulated contributions of input- and output-based HGFs across industries with different parameter values. The correlations (ρ_{real}, ρ_{cross}, $corr(\Delta prod, prod)$, $corr(prod, inp)$, $corr(\Delta inp, inp)$, $corr(inp, \Delta prod)$) are chosen randomly and independently from a uniform distribution with a range $[-0.5, 0.5]$. The standard deviations of the variables are also drawn randomly and independently from a uniform distribution with a range of $[0, 1]$. the average productivity growth is also uniformly random, $[0, 0.2]$. The contributions are winsorized at the 2nd and 98th percentiles.

warn policymakers that HGF promotion policies can have negative side effects on aggregate productivity, thus these effects should be quantified when designing and evaluating such a policy. The second conclusion on the importance of HGF definitions implies a key trade-off across definitions between productivity growth and job creation. A policy focusing on employment growth (i.e., input-based HGFs) may generate a small or negative productivity effect, while, according to our results, output-based HGFs create only somewhat fewer jobs but contribute to productivity growth substantially. Policies limited to firms that are very productive to start with may miss these firms. Our final insight emphasizes that HGFs are likely to have a higher contribution in environments that are in general conducive to efficient reallocation, while the low predictability of HGF contributions limits the extent to which industry-based targeting can be efficient.

Our research has two key limitations. First, the results are limited to one country, and investigating their external validity on microdata from other countries would be important. Such research would also help in identifying better how country and industry characteristics are related to the HGF contribution. Second, our method only considers the direct contribution of HGFs to productivity growth and ignores external effects. In this respect, it mainly complements recent research on such externalities (de Nicola et al., 2019; Du and Vanino, 2020).

Appendix A. Additional figures and tables

Additional figures

Additional tables

Appendix B. Data and productivity estimation

Data cleaning

We follow the procedure described in Muraközy et al. (2018). Here we describe the main steps.

A baseline cleaning is applied to the values of all the financial variables to correct for possible mistakes of reporting in HUF rather than 1000 HUF or for extremely small or big values in the data. Employment and sales are cleaned of extreme values and outliers. Suspiciously large jumps followed by another jump into the opposite direction are smoothed by the average of the previous and following years. Regarding capital stocks, we use the sum of tangible and intangible assets. Whenever intangible assets are missing, we input a zero. We deflate the different variables with the appropriate price indices.

The value added and the capital input price indices are available from the 2017 EU KLEMS at 2-digit industry level. Revenue is deflated by the producer price index at the 2-digit industry level from Eurostat Short-term Business Statistics. The intermediate inputs price index is calculated from the same database as the producer price index of intermediate goods. Regarding industry codes, the database, in general, includes the 2-digit NACE code of the firm in each year based on the actual NACE classification system. 4-digit industry codes are only available between 2000 and 2005. Because of the changes in NACE classification in 2003 and 2008, industry code harmonization is required. We harmonize to NACE Rev. 2 codes by using concordances from Eurostat. The conversion of industry codes in 2000-2002 to NACE Rev. 1.1 is relatively straightforward and efficient thanks to 4-digit codes before and after. The conversion from NACE Rev. 1.1 to NACE Rev. 2 is less so, as 4-digit codes are only available until 2005. Hence, for each firm, we assume that its 4-digit industry remained the same between 2005-2007 and use this 4-digit industry for the conversion. After these conversions, we clean industry codes when firms switch industries for only one year and fill missing codes from codes in previous or subsequent years. This process leads to a harmonized 2-digit NACE Rev. 2 code for each year.

Productivity estimation

We use the same productivity measures as Muraközy et al. (2018). We describe the approach based on that work.

From many perspectives, the most robust and convenient measure of productivity is labor productivity. We calculate this variable simply as the log of value added per employee. At the same time, the key shortcoming of labor productivity is that it does not reflect the differences in capital intensity across firms. Total Factor Productivity (TFP) aims to control for this issue. We estimate TFP with the method of Ackerberg et al. (2015), which can be regarded the state of the art.

Technically, firm-level TFP estimation involves estimating a production function:

$$\ln VA_{it} = \beta_L \times \ln L_{it} + \beta_K \times \ln K_{it} + \epsilon_{it}$$

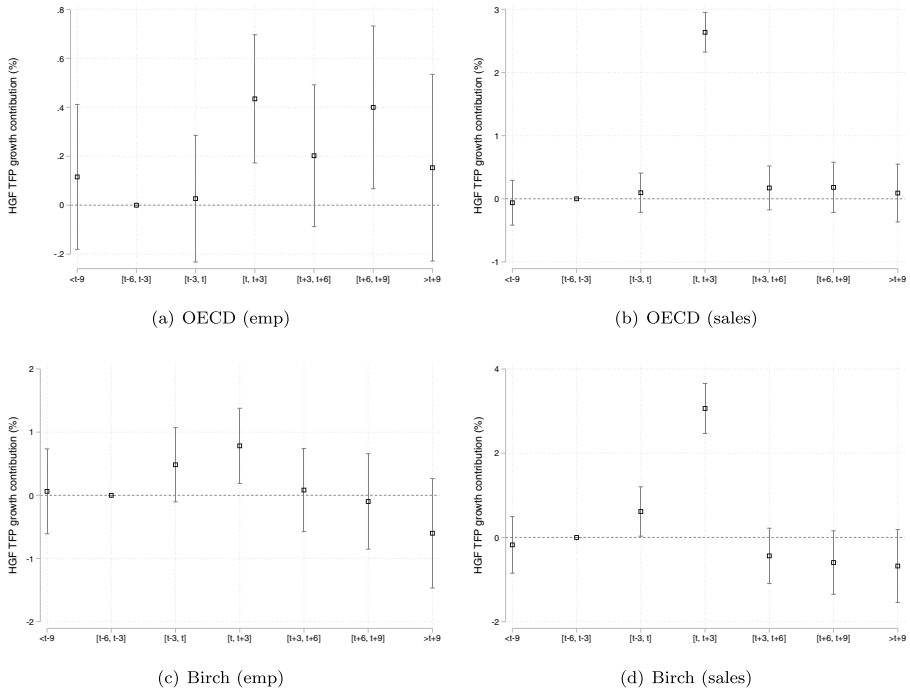
where i indexes firms, t indexes years, L_{it} is the number of employees and K_{it} is capital stock. In this specification, the residual of the equation, ϵ_{it} is the estimated TFP for firm i in year t. β_L and β_K are the output elasticities in the production function, reflecting the marginal product of labor and capital and the optimal capital intensity.

Estimating firm-level production functions involves several choices. First, usually it is important to include year fixed effects to control for macro- or industry-level shocks. Second, industries may differ in their optimal capital intensity, i.e., the coefficients of the two variables. To handle this, we estimate the production function separately for each industry (2-digit NACE within manufacturing and 1-digit NACE outside that).[40] Third, financial data reported by small firms may not be very accurate. Including them into the sample on which the production function is estimated may introduce bias into that regression. Hence, we estimate the production functions only on the sample of firms with at least 5 employees but also predict the TFP for smaller firms.

A key problem with firm-level TFP estimation is that input use (L_{it} and K_{it}) can be correlated with the residual TFP. Consequently, OLS estimation may yield biased coefficients. The bias arises from attributing part of the productivity advantage to the higher input use of more productive firms. A simple and robust solution for this issue is to estimate the production function with a fixed effects estimator. This method controls for endogeneity resulting from unobserved, time-invariant firm characteristics. A second and related problem is that input use can also be correlated with time-variant productivity shocks. This type of endogeneity is not corrected by the fixed effects estimator. More specifically, managers (unlike economists analyzing the balance sheet) may observe productivity shocks at the beginning of the year and adjust the flexible inputs (labor in our case) accordingly. As a result, we may falsely ascribe a productivity improvement to the increase in labor input. The recent best practice of handling this issue is the control function approach, in which one controls for the

[40] We merge a few industries with a very small number of observations to generate more robust estimates (NACE 10-12, 13-15, 19-21, 24-25, 29-30, 31-31). This way we ensure that we have at least 10,000 observations per industry.

Fig. 8 Labor Productivity contribution: event study



(a) OECD (emp)

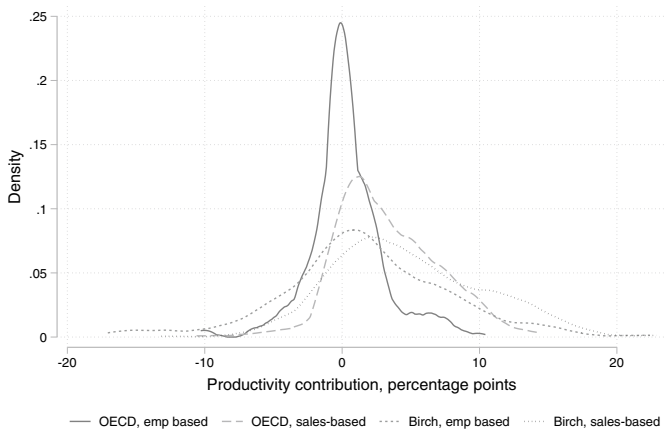
(b) OECD (sales)

(c) Birch (emp)

(d) Birch (sales)

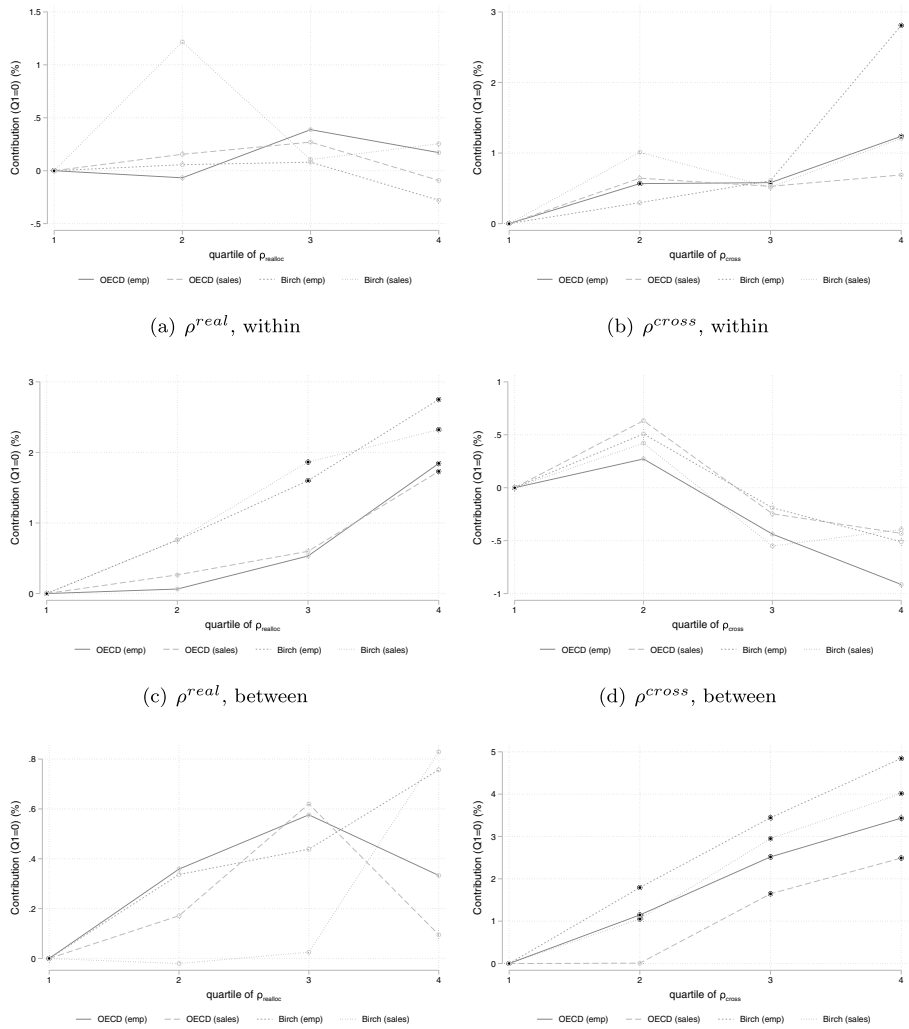
This figure shows the results of an event study regression in which one observation is the labor productivity contribution of a cohort of HGFs in a 2-digit industry in different event-times before and after the high-growth period. HGFs are firms that are in their high-growth phase in the 3-year period starting with year t. The vertical axis shows the labor productivity contribution of HGFs in percentage points, normalized to zero for the period $[t-6, t-3]$.

Fig. 9 Labor productivity contribution of HGFs, by industry-cohort



This figure shows the kernel distribution of the labor productivity contribution of HGF cohorts at the 2-digit industry-cohort level between t_0 and t_0+3, winsorized at the 5th and 95th percentiles. The horizontal axis shows the HGF contribution in log percentage points during the 3-year period.

Fig. 10 The relationship between industry dynamics parameters and the decomposition of the HGF contribution



(a) ρ^{real}, within

(b) ρ^{cross}, within

(c) ρ^{real}, between

(d) ρ^{cross}, between

(e) ρ^{real}, cross

(f) ρ^{cross}, cross

These figures repeat the regressions in Figure 4 with the different components of the HGF contribution as dependent variables: the HGF within contribution in panels a-b, the HGF between contribution in panels c-d and the HGF cross contribution in panels e-f. Year dummies are included, and standard errors are clustered at the industry level.

Table 7 Descriptives by 2-digit industry

nace2	firms	av. emp	sh. foreign	sh. exporter	sh. young	nace2	firms	av. emp	sh. foreign	sh. exporter	sh. young
5	19	113.7	13.9%	42.6%	23.8%	45	3231	21.4	7.3%	29.4%	20.3%
6	5	185.6	66.7%	95.2%	0.0%	46	8852	24.8	22.9%	50.9%	20.3%
7	5	144.6	12.8%	34.0%	4.3%	47	9556	31.2	5.9%	13.0%	24.8%
8	186	30.2	18.7%	34.9%	14.9%	49	2940	61.5	6.2%	38.5%	26.5%
9	25	79.6	30.6%	50.8%	19.8%	50	35	51.1	7.0%	21.3%	17.1%
10	2962	59.1	11.9%	32.3%	19.6%	51	28	175.3	53.5%	78.8%	25.5%
11	479	50.0	16.8%	57.0%	15.3%	52	1099	66.4	24.8%	47.4%	24.1%
12	10	252.8	58.9%	93.3%	4.4%	55	1064	37.4	13.4%	4.6%	21.6%
13	552	64.8	26.5%	60.9%	17.9%	56	4666	19.3	5.9%	3.0%	30.7%
14	1275	56.8	15.6%	51.5%	21.6%	58	576	27.4	22.5%	39.8%	18.8%
15	408	75.5	31.1%	59.8%	19.3%	59	198	20.9	28.3%	42.8%	27.0%
16	1232	27.5	11.3%	40.2%	21.2%	60	117	32.8	17.4%	22.9%	28.2%
17	315	63.0	22.6%	55.6%	18.9%	61	280	137.0	20.4%	26.0%	24.6%
18	785	29.1	7.8%	36.9%	16.4%	62	1498	30.2	17.9%	40.9%	29.6%
19	9	1239.3	59.6%	69.7%	12.1%	63	311	53.3	21.5%	35.1%	33.5%
20	379	77.6	26.1%	72.5%	14.9%	69	1052	21.6	7.9%	19.6%	19.5%
21	77	353.7	43.1%	72.7%	11.8%	70	1338	22.0	21.9%	29.4%	32.1%
22	1202	56.7	24.6%	65.3%	17.2%	71	1688	19.7	10.3%	24.1%	19.3%
23	969	62.8	21.9%	44.5%	19.0%	72	427	29.3	20.6%	57.4%	25.9%
24	228	151.4	29.1%	75.4%	16.5%	73	613	18.7	23.3%	43.1%	26.7%
25	3789	35.5	14.7%	49.2%	21.0%	74	241	17.9	8.8%	24.6%	24.7%
26	842	157.4	32.1%	60.3%	16.3%	75	39	11.7	1.5%	13.0%	14.1%
27	557	113.5	31.4%	57.3%	15.6%	77	426	17.9	18.8%	22.5%	23.9%
28	1510	67.9	20.5%	56.7%	14.5%	78	859	150.3	12.6%	16.3%	51.3%
29	355	287.4	48.6%	78.1%	18.3%	79	280	21.7	25.4%	20.1%	17.9%
30	115	96.9	19.5%	60.5%	17.2%	80	1342	29.8	2.2%	5.8%	34.6%
31	764	34.5	11.6%	41.1%	21.8%	81	1432	41.5	3.5%	4.2%	30.7%
32	340	35.8	17.3%	51.9%	17.0%	82	1550	33.3	11.2%	20.4%	39.3%
33	313	34.0	7.4%	34.8%	17.8%						

This table shows the number of firms and basic descriptives in our main sample by 2-digit industry: the average number of employees, the share of foreign-owned firms, the share of exporters and the share of young firms (at most 5 years old), respectively

Table 8 Employment and labor productivity before and after the high-growth phase by HGF definition

Panel A: OECD3 (emp)								
	employment				labor productivity			
year	t-4	t-1	t+4	t+7	t-4	t-1	t+4	t+7
2000		68.3%	239.2%	255.4%		118.6%	107.4%	106.7%
2003	73.8%	77.1%	258.2%	251.2%	125.0%	129.8%	113.4%	126.2%
2006	67.3%	72.9%	267.9%	235.2%	97.1%	118.3%	112.0%	111.2%
2009	65.1%	76.0%	260.9%		108.8%	130.4%	108.5%	
2011	67.9%	69.9%			123.9%	131.2%		
average	68.5%	72.8%	256.5%	247.3%	113.7%	125.7%	110.3%	114.7%
Panel B: OECD3 (sales)								
	employment				labor productivity			
year	t-4	t-1	t+4	t+7	t-4	t-1	t+4	t+7
2000		161.5%	100.0%	91.0%		118.0%	106.3%	109.7%
2003	154.9%	156.6%	92.1%	100.4%	121.2%	121.1%	104.7%	108.8%
2006	158.5%	165.8%	82.4%	89.3%	118.9%	110.4%	103.0%	104.6%
2009	144.2%	100.0%	88.8%		122.5%	100.0%	103.1%	
2011	100.0%	100.0%			100.0%	100.0%		
average	139.4%	136.8%	90.8%	93.6%	115.6%	109.9%	104.3%	107.7%
Panel C: Birch3 (emp)								
	employment				labor productivity			
year	t-4	t-1	t+4	t+7	t-4	t-1	t+4	t+7
2000		615.0%	100.0%	134.8%		118.0%	111.3%	104.6%
2003	609.1%	622.6%	142.6%	157.1%	124.3%	138.1%	110.5%	104.5%
2006	527.8%	506.8%	126.2%	138.0%	127.0%	125.2%	110.4%	104.4%
2009	598.8%	100.0%	131.5%		127.1%	100.0%	111.6%	
2011	100.0%	100.0%			100.0%	100.0%		
average	235.9%	222.4%	572.1%	580.9%	112.9%	120.8%	101.3%	102.1%
Panel D: Birch3 (sales)								
	employment				labor productivity			
year	t-4	t-1	t+4	t+7	t-4	t-1	t+4	t+7
2000		570.5%	100.0%	178.1%		181.3%	112.0%	104.4%
2003	496.5%	538.1%	174.2%	183.9%	191.5%	193.1%	113.6%	104.5%
2006	389.3%	442.6%	141.2%	160.1%	177.1%	168.1%	109.9%	103.7%
2009	488.5%	100.0%	175.4%		198.7%	100.0%	114.6%	
2011	100.0%	100.0%			100.0%	100.0%		
average	368.6%	350.2%	147.7%	174.1%	166.8%	148.5%	112.5%	104.2%

This table shows HGF performance one year before (t-1) and four years after (t+4) the beginning of the high-growth phase, as well as an additional 3 years before (t-4) and after that (t+7). The premia are expressed as the (unweighted) average number of employees and labor productivity of HGFs relative to that of the average firm (scaled to be 100%)

Table 9 Industry dynamics parameters and HGF contribution

Panel A: Base (TFP)				
	OECD (emp)	OECD (sales)	Birch (emp)	Birch (sales)
$\rho^{real}(Q2)$	1.152*	1.958**	0.356	0.591
	(0.658)	(0.876)	(0.411)	(0.487)
$\rho^{real}(Q3)$	2.119**	1.994*	1.496**	1.489**
	(0.983)	(1.094)	(0.597)	(0.706)
$\rho^{real}(Q4)$	3.229**	3.407***	2.348***	1.735*
	(1.226)	(1.167)	(0.663)	(0.872)
$\rho^{cross}(Q2)$	2.599	2.485	1.985*	1.285
	(1.643)	(1.602)	(0.992)	(0.960)
$\rho^{cross}(Q3)$	3.863**	2.914*	2.660**	1.929*
	(1.699)	(1.564)	(1.074)	(1.109)
$\rho^{cross}(Q4)$	7.141***	4.829***	3.757***	2.749***
	(1.776)	(1.685)	(1.053)	(0.901)
Observations	724	724	724	724
R-squared	0.194	0.144	0.196	0.169
Panel B: Base (Labor productivity)				
	OECD (emp)	OECD (sales)	Birch (emp)	Birch (sales)
$\rho^{real}(Q2)$	0.251	1.731**	-0.011	1.238*
	(0.751)	(0.694)	(0.420)	(0.679)
$\rho^{real}(Q3)$	1.917**	3.225***	0.496	1.838*
	(0.904)	(0.996)	(0.553)	(0.979)
$\rho^{real}(Q4)$	3.947**	4.630***	1.233**	2.671**
	(1.483)	(1.468)	(0.611)	(1.218)
$\rho^{cross}(Q2)$	2.257***	1.457*	1.159**	1.544**
	(0.709)	(0.871)	(0.529)	(0.702)
$\rho^{cross}(Q3)$	3.179***	2.128***	1.480***	1.219*
	(0.974)	(0.668)	(0.538)	(0.673)
$\rho^{cross}(Q4)$	5.505***	3.897***	1.854***	2.612***
	(0.767)	(0.669)	(0.524)	(0.700)
Observations	724	724	724	724
R-squared	0.217	0.203	0.212	0.200

Panel A presents the regression results corresponding to Fig. 4 and Eq. 2. The dependent variable is the HGF contribution to productivity growth of a HGF cohort of an industry in a 3-year period, while the explanatory variables are quartiles of ρ^{real} and ρ^{cross}. The base category is the lowest quartile (Q1). Year dummies are included and standard errors are clustered at the industry level. Panel B re-runs the regression in Panel A with HGF contribution to labor productivity as the dependent variable

productivity shock by using a proxy for it in an initial step. The proxy is another flexible input, usually materials or energy use. As we have reliable data on materials, we will use that variable. We rely on the method of Ackerberg et al. (2015). Importantly, with this method, the production function coefficient estimates are close to what is expected, and the returns to scale are slightly above one (typically between 1 and 1.1, see Figure 11). We also control for attrition of firms from the sample, but this does not affect the estimates significantly. After estimating the coefficients, we simply calculate the estimated TFP for firm i in year t by subtracting the product of inputs and the estimated elasticities:

$$TFP_{it} = \ln VA_{it} - \beta_L \times \ln L_{it} - \beta_K \times \ln K_{it}$$

In this way, we calculate a TFP level (rather than its value relative to year and industry fixed effects), which is important when calculating productivity changes. Note that the calculated productivity changes are very similar to the logic of the Solow-residual.

When estimating the production function, we measure labor by the number of workers, capital with fixed assets and value added as sales minus material expenditures corrected with the change in inventories.

Table 10 presents the average labor productivity and TFP by 1-digit NACE categories in 2003 and 2015.[41]

Appendix C. HGF definitions

Here we provide some additional statistics about the differences and similarities between the four HGF definitions we use, as well as about HGF dynamics.

The economic importance of different HGFs

Starting with the overlap between the definitions, we find that if two definitions differ in only the growth measure, the overlap is between 45-62%, while if they differ in both dimensions, the overlap is lower, 27-41% (see Table 11). In other words, the various HGF measures overlap to some extent but capture a quite different set of firms.[42]

Next, we document how many firms are covered by the different definitions. By definition, the Birch classifications cover 5% of firms, therefore, the question is only relevant for the OECD definitions. Figure 12 shows the share of HGFs in 2013 by industry for the two OECD definitions.[43] With few exceptions, the share of employment OECD HGFs is between 1 and 6% and that of sales HGFs is between 3 and 10%.[44] The key pattern is that the OECD sales definition captures a significantly larger group of firms compared to the employment definition. Therefore, a substantial subset of sales HGFs actually expand their sales faster than their employment during the 3-year period, which suggests that increased productivity is a key source of growth for many sales HGFs.[45]

The empirical relevance of the two dimensions which distinguish between the different HGF definitions can be more clearly seen when we quantify the economic importance of HGFs by their industry share in terms of employment and sales. This is illustrated by Fig. 13, which shows HGFs' employment and sales share at the beginning of the period (horizontal axis) and after 3 years (vertical axis) by 2-digit industry. The 45-degree line represents no change in employment share.

[41] We have estimated all of these with the prodest (Rovigatti and Mollisi , 2018) command in Stata.

[42] As a comparison, Delmar et al. (2003) show that from the Swedish HGFs defined based on absolute growth about 15% have high growth only in sales but not in employment and 20% have high growth only in employment but not in sales. Shepherd and Wiklund (2009) find that the correlation between employment and sales growth is moderate. Daunfeldt et al. (2014) show that within employment HGFs the correlation between Birch-type composite and relative growth-based measures is 47% in Sweden.

[43] The figure omits the outlier industry 7 (Mining of metal ores), in which the number of firms is very low and the share of sales HGFs is 25%. Note that the share of HGFs according to the Birch definitions is fixed for each year. Also, recall that our sample consists of firms with 10 or more employees in at least one year.

[44] While the magnitudes are similar, HGF prevalence depends strongly on the macro cycle. These numbers are similar to what was found in comparison countries, Grover Goswami et al. (2019), Fig. 1.1.

[45] Clearly, instead of productivity growth an increasing use of other inputs (materials or capital) might also explain this pattern. However, as we see in our analysis, TFP growth is the main explanation.

Table 10 Average productivity

		Labor productivity				Total factor productivity			
		2003		2015		2003		2015	
NACE	Description	Mean	St.dev.	Mean	St.dev.	Mean	St.dev.	Mean	St.dev.
B	Mining and quarrying	3.33	0.88	4.29	0.84	1.35	0.7	2.05	0.73
C	Manufacturing	3.11	0.86	3.35	0.82	2.41	0.75	2.58	0.74
D	Electricity, gas	4.6	0.96	4.62	1.44	3.49	0.77	3.27	1.33
E	Water supply; sewerage	3.52	0.76	3.64	0.9	2.89	0.59	2.94	0.79
F	Construction	3.14	0.77	3.39	0.78	2.58	0.66	2.88	0.67
G	Wholesale and retail trade	3.38	0.96	3.44	0.91	2.67	0.85	2.81	0.78
H	Transportation and storage	3.82	0.74	3.69	0.78	2.84	0.64	2.79	0.62
I	Accommodation	2.57	0.75	2.77	0.86	1.84	0.63	2.12	0.74
J	ICT	3.68	1.02	3.91	0.94	2.79	0.87	3.09	0.82
K	Financial and insurance	4.55	1.28	4.19	1.34	3.6	1.16	3.36	1.23
L	Real estate	3.61	1.06	3.51	1.09	2.89	0.91	2.91	0.91
M	Professional, scientific	3.6	0.88	3.74	0.93	2.85	0.8	3.06	0.83
N	Administrative and support	3.07	0.98	3.19	1.02	2.76	0.84	2.96	0.87
S	Other service	3.29	0.66	3.13	0.84	2.81	0.64	2.72	0.75
	Total	3.3	0.94	3.43	0.93	2.6	0.82	2.77	0.81

This figure shows the average labor productivity and TFP and their standard deviation by industry. Firms with at least 5 employees, based on Muraközy et al. (2018) Table 2.1.

HGFs captured by the OECD and Birch definitions have a very different economic importance. For OECD HGFs, the median initial employment and sales shares are around 4% and 8%, respectively, while the typical initial shares are around 20 and 30% for Birch HGFs. By definition, HGFs' industry share increases during their high-growth phase, represented by the cluster of industries being above the 45-degree line.

Regarding employment and sales HGFs, there are three clear patterns in the figures. First, both types of firms increase their share both in terms of employment and sales. Second, as expected, HGFs increase their share more in the dimension corresponding to the definition. Combining these two observations implies that sales-based HGFs are more likely to increase their productivity and still create jobs, though to a smaller extent than employment-based HGFs. Finally, the differences between the employment and sales-based HGFs are more pronounced for the OECD definitions compared to the Birch definitions. This is likely to be driven by the larger fluctuations that characterize smaller firms.

HGF dynamics

In this subsection we consider three questions that may lead to differences in the long-term productivity contribution of HGFs: the role of young firms, transitions between different types of high growth and exit.

First, Fig. 14 shows the share of young firms (at most 5 year old at the start of the high-growth period) among HGFs. In line with their smaller size, the share of young firms is smaller (around 30% on average) among Birch HGFs and larger among OECD HGFs (around 40% on average). Employment and sales HGFs do not differ in this respect. There is clearly a downward time trend, in line with the overall declining entry rate (see Table 2).

Second, we consider the possibility that one type of high growth may be typically followed by another type (for example, employment-based by sales-based). A straightforward way to test for these possibilities is a transition matrix approach. Table 12, shows the probability that a firm which was a HGF between t-3 and t is also a HGF (according to the different definitions) between t and t+3. A key pattern

Fig. 11 ACF production function coefficients. This figure shows the production function coefficients estimated by industry. Based on Muraközy et al. (2018) Fig. 2.1.

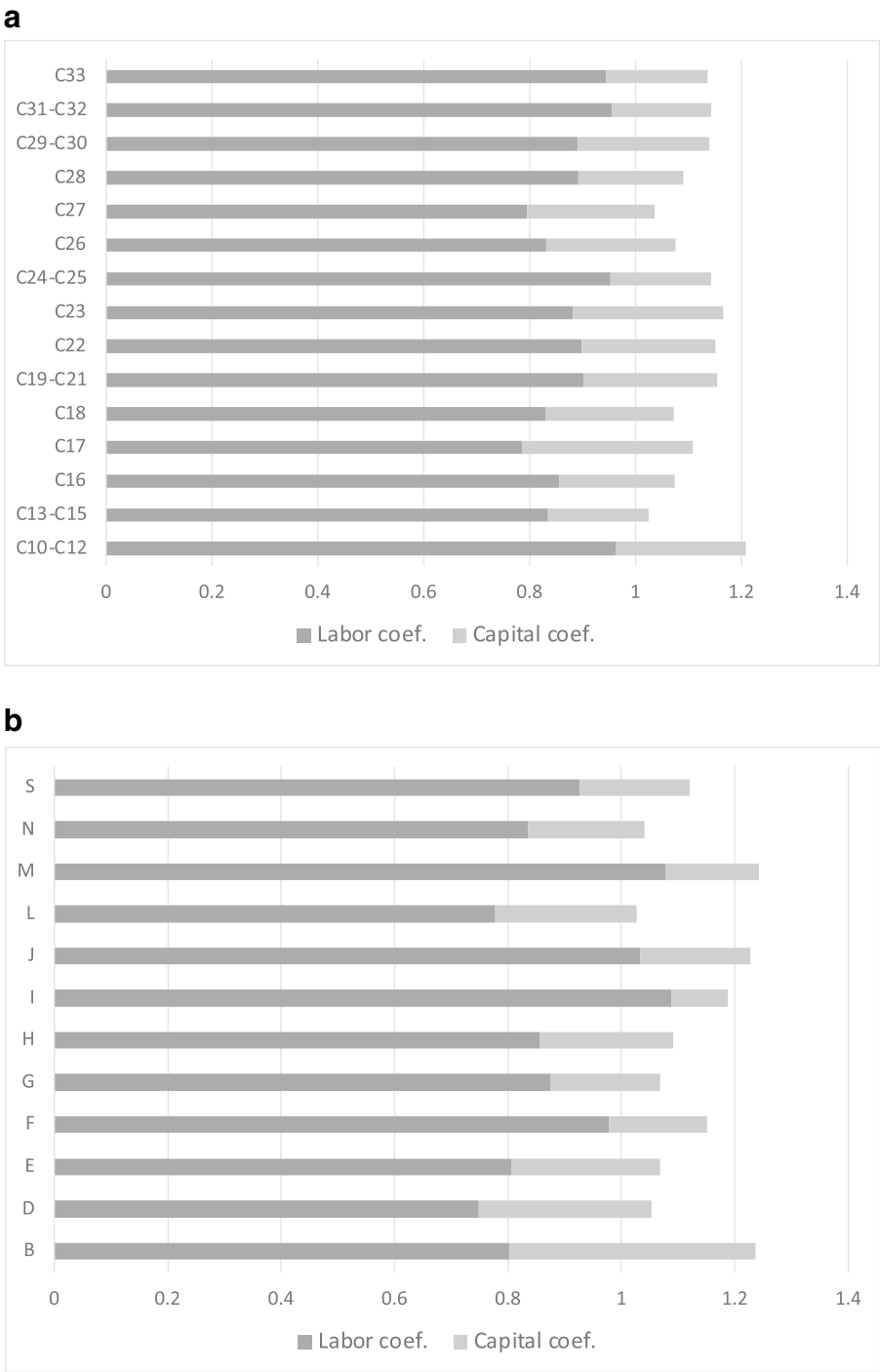


Table 11 Overlap between different HGF definitions, 2013

	OECD (emp)	OECD (sales)	Birch (emp)	Birch (sales)
OECD (emp)	100.0%	61.9%	52.6%	27.0%
OECD (sales)	61.9%	100.0%	40.5%	58.6%
Birch (emp)	52.6%	40.5%	100.0%	44.5%
Birch (sales)	27.0%	58.6%	44.5%	100.0%

The overlap between any definition pairs i and j is calculated as: $overlap_{i,j} = \frac{N_{i \cap j}}{min(N_i;N_j)}$, where N is the number of firms in a set

Table 12 Transition matrix between different types of HGFs, t=2013

	Type of HGF in t			
HGF in t-3:	OECD3 (emp)	OECD (sales)	Birch (emp)	Birch (sales)
all firms	3.2%	8.2%	5.0%	5.1%
OECD3 (emp)	5.1%	12.8%	27.5%	18.8%
OECD (sales)	6.5%	9.7%	24.8%	30.9%
Birch (emp)	8.3%	17.4%	15.6%	9.5%
Birch (sales)	7.7%	14.1%	11.0%	9.6%

This table shows the probability that a firm which was a HGF between t-3 and t is also a HGF (according to the different definitions) between t and t+3

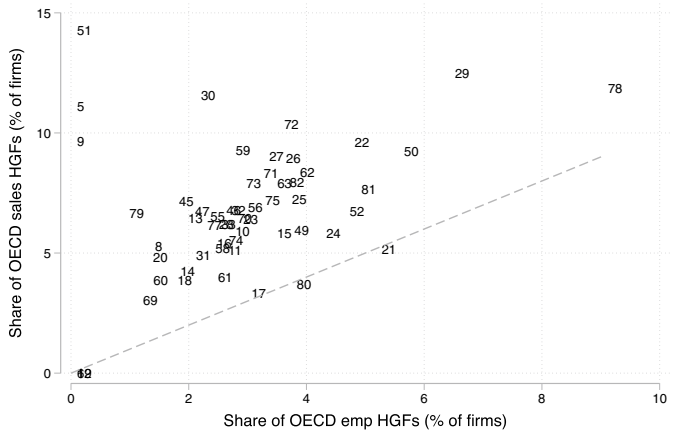
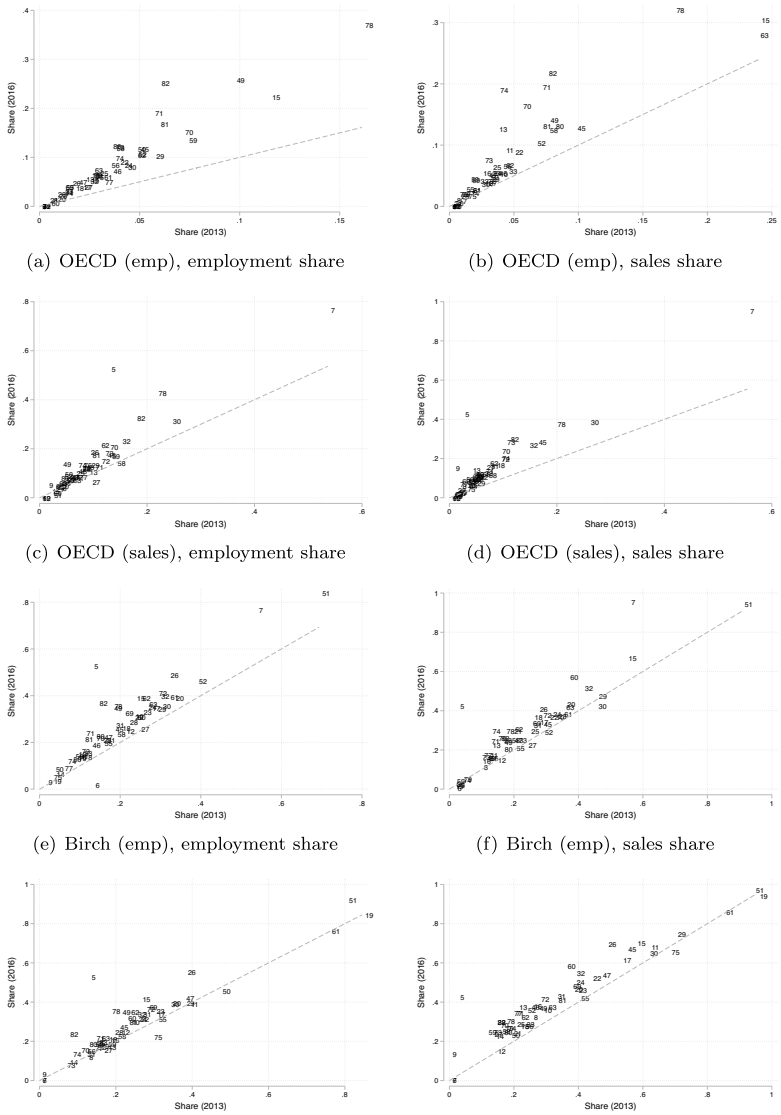


Fig. 12 Share of HGFs in the total number of firms, 2013

The figure shows the share of firms that were HGFs between 2013 and 2016 according to the OECD (employment) and OECD (sales) HGF definitions by 2-digit industry. The sample includes firms that had at least 10 employees at any point.

Fig. 13 Employment and sales shares of different HGFs in the beginning and at the end of the high-growth period starting in 2013



(a) OECD (emp), employment share

(b) OECD (emp), sales share

(c) OECD (sales), employment share

(d) OECD (sales), sales share

(e) Birch (emp), employment share

(f) Birch (emp), sales share

(g) Birch (sales), employment share

(h) Birch (sales), sales share

This figure shows the 2013 and 2016 employment and sales share of firms that were HGFs in 2013 according to the different definitions. The numbers indicate the NACE code of 2-digit industries and the line is the 45-degree line. The sample includes firms that had at least 10 employees at any point.

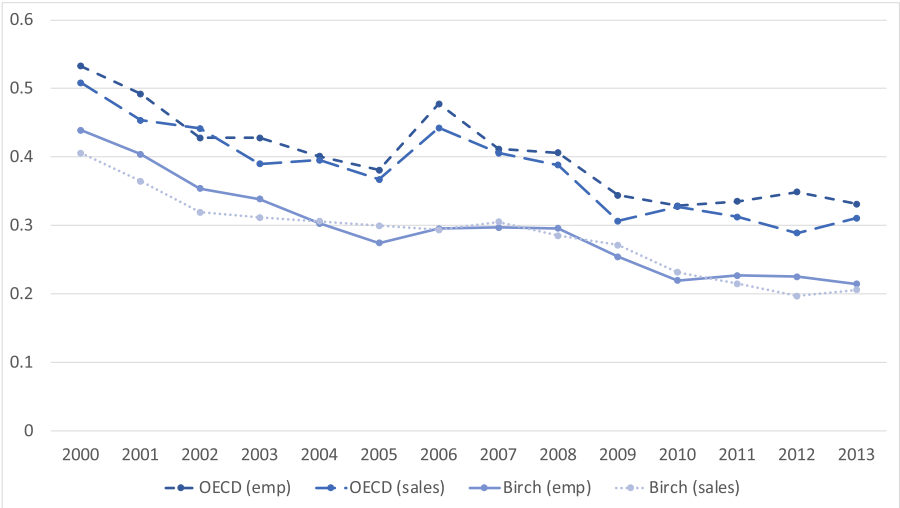


Fig. 14 Share of young firms among HGFs

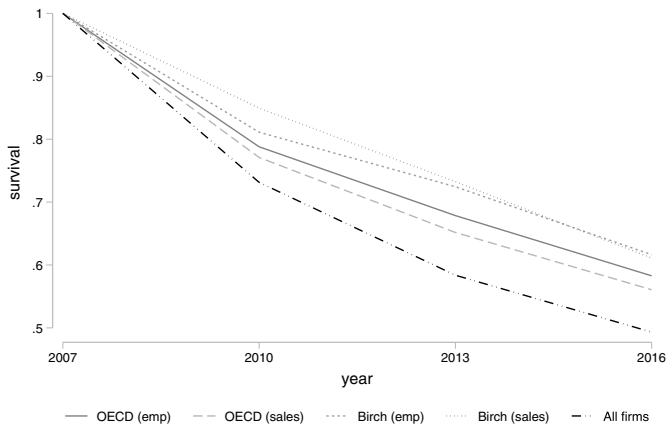


Fig. 15 Survival by HGF definition

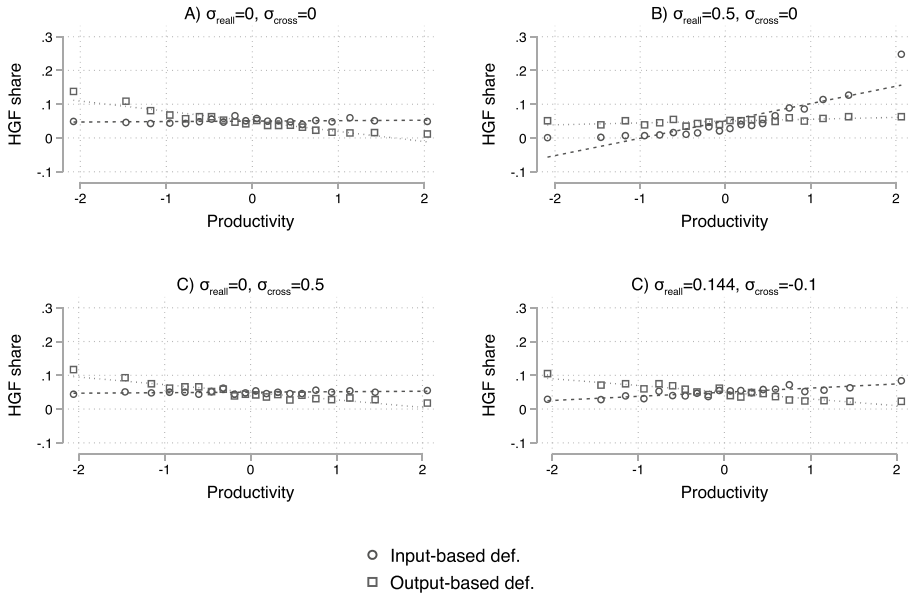
The figure shows the survival rate of firms already existing in 2004 from 2007 on by their HGF status between 2004 and 2007.

is that HGF status is persistent, but only to a small extent: depending on the definition, a firm which was HGF between t-3 and t is about 1.2-3.1 times more likely to become the same type of HGF again between t and t+3 than the average firm. The persistence of a Birch-type HGF status is stronger than that of the OECD-type HGF status. The main driver of this pattern might be the importance of firm size in the Birch-type definition, which is a quite persistent feature. Note, however, that persistence is limited: the overwhelming majority of HGFs will not remain a HGF in the next period, and high-growth status is often only a phase rather than a permanent characteristic of a firm's life.[46]

Table 12 also provides limited evidence for different types of high-growth periods following each

[46] These patterns are similar for the other periods, and they are also in line with the findings of Hölzl (2014) or Daunfeldt et al. (2014), who show for Austrian and Swedish firms that Birch HGFs have more persistent growth than OECD HGFs.

Fig. 16 Parameter values and reallocation



This figure shows the simulated relationship between initial productivity level and the share of input- and output-based HGFs for different values of the key moments of industry dynamics. We normalize each of the variables in such a way that their standard deviation is 1. The remaining correlations are calibrated from the Hungarian data: $corr(\Delta prod, prod) = -0.41$, $corr(prod, inp) = 0.1418$; $corr(\Delta inp, inp) = -0.1852$; $corr(inp, \Delta prod) = 0.0011$

other.[47] For example, employment OECD HGFs are slightly more likely to become sales OECD HGFs in the next period than sales OECD HGFs do, but the difference is small. In the case of Birch firms, we find evidence for the opposite phenomenon: sales-based HGFs are more likely to become sales-based HGFs in the next period than employment-based HGFs do.

Finally, to explore the possibility that different HGFs are more likely to exit, probably thanks to a more risky growth strategy, Fig. 15 shows survival rates calculated from 2007 for firms that were HGFs between 2004 and 2007 and for other firms.[48] We find that all types of HGFs are actually less likely to exit than non-HGFs.[49] The least likely to exit are the sales-based Birch-type HGFs followed by the employment-based Birch-type HGFs and then the two types of OECD HGFs. The differences become smaller but still visible when we control for industry, size and age.

Appendix D. Further simulation results

In this Appendix we report some further results of our simulation exercise. First, in Appendix Section 1 we investigate how changing the industry dynamics parameters affect the HGF contribution in this framework. Second, in Appendix Section 2 we describe how this framework may be applied to investigate the productivity growth contribution of other groups of firms.

[47] Davidsson and Wiklund (2006) suggest that change in demand first leads to a change in sales, and it might change the level of employment only afterwards. As opposed to that, Delmar et al. (2003) emphasize the high variability in growth patterns.

[48] We have chosen this early cohort so that we can follow their survival in the long run. By definition, HGFs cannot exit in their high-growth phase, therefore calculating survival from 2004 would be 'unfair' to non-HGFs. That is why the figure starts from 2007 and compares firms that were in their high-growth phase in 2004 with all the firms that were active in 2004 and survived up to 2007. To be conservative, we do not include new entrants in the non-HGF group.

[49] Acs et al. (2009) find that only 4% of US HGFs exit in the 4-year period after their high-growth phase, and Choi et al. (2017) and Mohr et al. (2014) show evidence for the positive impact of early-phase high-growth status on subsequent survival probability. On the other hand, Delmar et al. (2013) find a negative effect of growth on survival. Gjerlov-Juel and Guenther (2019) show that high employment growth of young firms is linked with a higher subsequent survival probability only if there is low employee turnover after the growth phase.

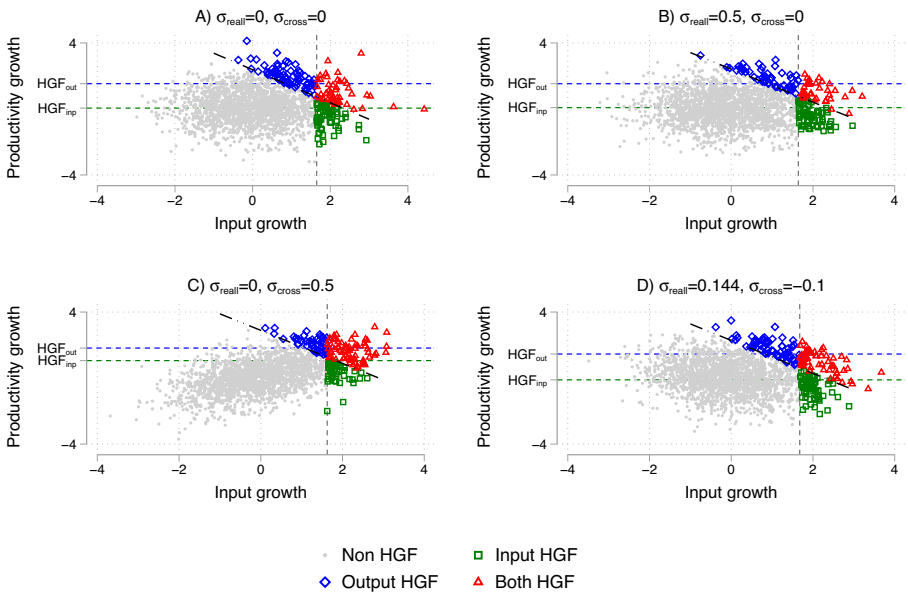


Fig. 17 Parameter values and the within effect

Industry dynamics

In this subsection we focus on the relationship between industry dynamics parameters and productivity contributions. We follow similar steps (and graphical tools) to that of Subsection 5.2, starting with the between component, followed by the within component and presenting regression results about the relationship between the parameters and the overall HGF contribution.

Let us start with how the parameters affect the between term. Figure 16 illustrates how initial productivity is related to the probability of becoming a HGF under different values of the two key parameters while keeping the others constant.[50] Panel A shows that if both key parameters are zero, the share of input HGFs is roughly constant in productivity while that of output HGFs is decreasing with productivity, which, as we have seen, is a consequence of productivity convergence. According to Panel B, when reallocation becomes stronger ($\rho_{realloc} = corr(productivity, \Delta size)$, the slope of the two curves increases. This modification generates a positive relationship between productivity and the share of input HGFs and a constant (rather than negative) relationship between productivity and the share of output HGFs. This result implies that when reallocation becomes stronger and more effective, HGFs are more likely to contribute more via their between term. Panel C shows the effect of increasing the cross-correlation. This change does not seem to affect the between component because it is a correlation between growth rates rather than levels. Finally, Panel D shows the observed average patterns, reproducing Fig. 5 as a reference point.

Figure 17 repeats this exercise for the within contribution, following the structure of Fig. 6. The within contributions (or average productivity growth) of input and output HGFs are represented by HGF_{inp} and HGF_{out} on the vertical axis, respectively. In three out of the four panels we see a similar pattern: the input HGF contribution is much smaller than the output HGF contribution. This is quite different, however, when the cross-correlation becomes strongly positive in Panel C. Now there is a much larger overlap between input and output HGFs, and the within contribution of input HGFs is substantially higher than in the other three cases. These simulations predict that a higher cross-correlation is positively associated with the within contribution of input HGFs (and also of output HGFs, though to a smaller extent). By definition, we expect as well that higher cross-correlation also yields larger cross contribution for both

[50] In the simulations all the other parameters take the value observed in the Hungarian data to make the results more comparable with the empirical exercise. The exact values of the parameters are, however, irrelevant for how the patterns change, as we change the two key parameters.

Table 13 Empirical patterns captured by the model

Panel A: Input based				
	within	between	cross	overall
ρ_{real}	0	+ +	0	+ +
ρ_{cross}	+ +	0	+ +	+ +
Panel B: Output based				
	within	between	cross	overall
ρ_{real}	0	+	0	+
ρ_{cross}	+	0	+	+

This table shows the main predictions from the simulation exercise. The predictions about the between effect come from Fig. 16. Comparing panels A and B of that figure shows that an increase in σ_{real} is positively associated with the reallocation contribution of both types of HGFs, but this increase is larger for input-based HGFs. Comparing Panels A and C shows that the cross parameter has very limited effect on the reallocation contribution. The predictions regarding the within and cross terms come from Fig. 17. Comparing Panels A and C of that figure suggests that the cross-correlations have a positive effect on both types of HGFs, but this effect is larger for input HGFs. Comparing Panels A and C does not indicate a substantial effect of the reallocation correlation

definitions. This simple simulation produces a number of patterns, in line with our empirical findings, presented in Table 13.

Table 14 illustrates that these patterns also work out in a regression analysis. We simulate 10,000 samples of 200 firms for this exercise and draw all correlations randomly and independently from each other for the 4-dimensional correlation matrix between size, productivity, size growth and productivity growth. Then we calculate the different HGF contributions for all these samples.

We run regressions of the form:

$$contribution_i = \beta_0 + \beta_{real}\rho_i^{real} + \beta_{cross}\rho_i^{cross} + \epsilon_i \qquad (3)$$

where $contribution_i$ are the different HGF contributions.

We find that the patterns in these regressions are much in line with the actual findings in Figs. 4 and 10. Stronger reallocation increases the contribution for both types of HGFs via the between component. Larger cross-correlation also improves the

Table 14 Regression results on simulated data

Panel A: Input-based HGFs				
	(1)	(2)	(3)	(4)
Contribution:	total	within	between	cross
$\rho_{realloc}$	0.141***	-0.001	0.145***	-0.003
	(0.005)	(0.001)	(0.003)	(0.003)
ρ_{cross}	0.199***	0.055***	-0.002	0.146***
	(0.005)	(0.001)	(0.003)	(0.004)
Observations	9,853	9,853	9,853	9,853
R-squared	0.234	0.388	0.258	0.173
Panel B: output-based HGFs				
	(5)	(6)	(7)	(8)
Contribution:	total	within	between	cross
$\rho_{realloc}$	0.104***	0.000	0.106***	-0.001
	(0.006)	(0.002)	(0.003)	(0.003)
ρ_{cross}	0.099***	0.010***	-0.000	0.090***
	(0.006)	(0.002)	(0.002)	(0.004)
Observations	9,853	9,853	9,853	9,853
R-squared	0.068	0.003	0.188	0.060

This table shows estimates of Regression 3 on the simulated data, in which one of the components of the total HGF contribution is the dependent variable and ρ^{real} and ρ^{cross} are the explanatory variables. In 10,000 rounds of simulations, the correlations (ρ_{real}, ρ_{cross}, $corr(\Delta prod, prod)$, $corr(prod, inp)$, $corr(\Delta inp, inp)$, $corr(inp, \Delta prod)$ are chosen randomly and independently from a uniform distribution with a range $[-0.5, 0.5]$. The standard deviations of the variables are also drawn randomly and independently from a uniform distribution with a range of [0, 1]. The average productivity growth is also uniformly random, [0, 0.2]. The contributions are winsorized at the 2nd and 98th percentiles

contribution of both types of HGFs both via the within and cross terms.

Applying the model for other groups of firms

This framework, which is able to capture many patterns regarding HGF productivity contributions, can be extended to seize the productivity contributions of other groups of firms with a policy interest. The modification of the model requires two changes. First, the definition of HGFs requires HGFs to survive continuously during the high-growth phase. This allows us to abstract away from exit, which may not be the case for some other groups of firms and requires us to introduce an extra dimension: the probability of exit. Second, different correlations of the joint distributions may be of interest for other groups of firms. Let us illustrate this with the example of SMEs.

SMEs can be defined as firms with a size below a certain threshold of *inp* in year t rather than Δinp, as in the case of input HGFs. When considering the decomposition of the productivity contribution of a cohort of SMEs between t and $t+3$, the SME-within term will depend on the extent to which small firms increase their productivity, which, in turn, depends on $corr(inp, \Delta prod)$. The economic interpretation of this term is whether small firms can increase their productivity more (or less) than large firms. Empirically, this is very close to zero in the Hungarian data, suggesting that the within contribution of SMEs will be around zero.

The SME *between* term can depend on multiple correlations. First, it depends on the general strength of reallocation in the industry (0.144 on average). Moreover, it also depends on what type of firms are "selected" into the SME group. If, for example SMEs are less productive but grow faster than other firms, the correlation within the SME group will be less positive than the reallocation correlation in the full sample. The relationship between SME status and productivity depends on $corr(inp, prod)$. The economic interpretation of this term is close to the *static allocative efficiency*, as defined by Olley and Pakes (1992). Its average value is around zero (0.142) in our data, therefore SMEs are typically less productive than the average firm in our sample. Second, the relationship between SME status and productivity depends on $corr(inp, \Delta prod)$ (the same correlation to the one in the within term), which is close to zero. As a result, the SME *between* term is likely to have the same sign as the reallocation term, therefore it is likely to be positive.[51]

The *cross* term will depend on the cross-correlation and is likely to be negative, as in our HGF application. Similarly to the between term, it also depends indirectly on other correlations, one of which is low ($cov(inp, \Delta prod) = 0.0011$ and $cov(inp, \Delta inp) = -0.19$).

Finally, it would be hard to abstract away from exit for SMEs. The effect of exits depends on whether exiting SMEs are less productive than the average firm, which is mainly driven by how productivity is correlated with the new dimension in our framework, the probability of exit. The exit term will typically be positive for SMEs given the positive correlation between size and productivity.

To sum up, based on the observed moments of the size and productivity distribution, it is likely that SMEs' contribution is positive but similar to other firms' contribution for each 3-year period.[52] This is mainly driven by the positive reallocation and exit terms. We also learned that a crucial correlation which affects the SME contribution is the correlation between size and productivity growth. The strength of reallocation and the efficiency of the exit process are also related to the SME contribution. However, these two latter correlations affect SMEs similarly to other firms. Note that this is not the case for HGFs, for which the cross-correlation does not only affect the *cross* term (similarly to all firms) but it is also

[51] The logic is the following. Assume X and Y variables both have expected value of 0 for all firms. Therefore, on the full sample, $Cov(X, Y) = E(XY)$. Assume that for a subset of firms $E(x) = a$ and $E(Y) = b$. The covariance for this sample of firms is $E(XY) - ab$. Therefore, the correlation for the selected group of firms will only differ if both a and b are different from zero (assuming no difference in the variance). In our example, $X = prod$ and $Y = \Delta inp$ and the overall correlation is the reallocation correlation, 0.14. The correlation for SMEs will differ if SMEs differ from other firms both in terms of *prod* and Δinp, which requires both $corr(inp, prod)$ and $corr(inp, \Delta prod)$ to be substantially different from zero. However, this is not the case, therefore the SME sample reallocation correlation is similar to the full sample. Note that these selection effects for the between term are not an issue for (input-based) HGFs, because they are defined based on one of the variables (Δinp) in this correlation.

[52] All terms, with the possible exception of the exit term, consist of correlations that are very similar to their industry-level counterparts.

important in determining the within term, and the reallocation correlation also affects the type of firms selected, as shown in Fig. 5.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Ackerberg DA, Caves K, Frazer G (2015) Identification properties of recent production function estimators. Econometrica 83(6):2411–2451. https://doi.org/10.3982/ECTA13408

Acs Z, Parsons W, Tracy S (2009) High-impact firms: Gazelles revisited (manuscript). SBA Reports, SOo Advocacy, Editor

Andrews D, Criscuolo C, Gal P, Menon C (2015) Firm dynamics and public policy: evidence from OECD countries. In: RBA annual conference volume. Reserve Bank of Australia

Andrews D, Criscuolo C, Gal PN (2016) The best versus the rest: The global productivity slowdown, divergence across firms and the role of public policy. OECD Productivity Working Papers 2016(5), https://doi.org/10.1787/63629cc9-en

Armenter R, Koren M (2014) A balls-and-bins model of trade. American Economic Review 104(7):2127–51,https://doi.org/10.1257/aer.20151233

Arrighetti A, Lasagni A (2013) Assessing the determinants of high-growth manufacturing firms in Italy. International Journal of the Economics of Business 20(2):245–267, https://doi.org/10.1080/13571516.2013.783456

Autio E, Rannikko H (2016) Retaining winners: Can policy boost high-growth entrepreneurship? Research policy 45(1):42–55, https://doi.org/10.1016/j.respol.2015.06.002

Balsvik R, Haller SA (2006) The contribution of foreign entrants to employment and productivity growth

Baqaee DR, Farhi E (2020) Productivity and misallocation in general equilibrium. The Quarterly Journal of Economics 135(1):105–163. https://doi.org/10.1093/qje/qjz030

Bartelsman E, Haltiwanger J, Scarpetta S (2013) Cross-country differences in productivity: The role of allocation and selection. American Economic Review 103(1):305–34, https://doi.org/10.1257/aer.103.1.305

Bianchini S, Bottazzi G, Tamagni F (2017) What does (not) characterize persistent corporate high-growth? Small Business Economics 48(3):633–656, https://doi.org/10.1007/s11187-016-9790-1

Birch DL (1981) Who creates jobs? The Public Interest 65:3.

Bos JW, Stam E (2014) Gazelles and industry growth: a study of young high-growth firms in the Netherlands. Industrial and Corporate Change 23(1):145–169 https://doi.org/10.1093/icc/dtt050

Brown R, Mawson S (2016) Targeted support for high growth firms: Theoretical constraints, unintended consequences and future policy challenges. Environment and Planning C: Government and Policy 34(5), 816–836.

Brown R, Mawson S, Mason C (2017) Myth-busting and entrepreneurship policy: the case of high growth firms. Entrepreneurship & Regional Development 29(5–6), 414–443, https://doi.org/10.1080/08985626.2017.1291762.

Choi T, Rupasingha A, Robertson JC, Leigh NG (2017) The effects of high growth on new business survival. Review of Regional Studies 47(1):1–23, https://doi.org/10.52324/001c.8025

Coad A (2007) Testing the principle of 'growth of the fitter': the relationship between profits and firm growth. Structural Change and Economic Dynamics 18(3):370–386, https://doi.org/10.1016/j.strueco.2007.05.001

Coad A (2018) Firm age: a survey. Journal of Evolutionary Economics 28(1):13–43, https://doi.org/10.1007/s00191-016-0486-0

Coad A, Broekel T (2012) Firm growth and productivity growth: evidence from a panel var. Applied Economics 44(10), 1251–1269.

Coad A, Srhoj S (2019) Catching gazelles with a lasso: Big data techniques for the prediction of high-growth firms. Small Business Economics 55:541–565, https://doi.org/10.2139/ssrn.3395527.

Coad A, Daunfeldt SO, Hölzl W, Johansson D, Nightingale P (2014) High-growth firms: Introduction to the special section. Industrial and Corporate Change 23(1), 91–112, https://doi.org/10.1093/icc/dtt052.

Coad A, Segarra A, Teruel M (2016) Innovation and firm growth: does firm age play a role? Research policy 45(2):387–400, https://doi.org/10.1016/j.respol.2015.10.015.

Daunfeldt SO, Halvarsson D (2015) Are high-growth firms one-hit wonders? Evidence from Sweden. Small Business Economics 44(2), 361–383, https://doi.org/10.1007/s11187-014-9599-8.

Daunfeldt SO, Elert N, Johansson D (2014) The economic contribution of high-growth firms: Do policy implications depend on the choice of growth indicator? Journal of Industry, Competition and Trade 14(3), 337–365, https://doi.org/10.1007/s10842-013-0168-7.

Daunfeldt SO, Halvarsson D, Johansson D (2015) Using the Eurostat-OECD definition of high-growth firms: a cautionary note. Journal of Entrepreneurship and Public Policy 4(1), 50–56, https://doi.org/10.1108/JEPP-05-2013-0020.

Davidsson P, Wiklund J (2006) Conceptual and empirical challenges in the study of firm growth. Entrepreneurship and the Growth of Firms 1(1), 39–61, https://doi.org/10.4337/9781781009949.00010.

De Wit G (2005) Firm size distributions: An overview of steady-state distributions resulting from firm dynamics models. International Journal of Industrial Organization 23(5–6), 423–450, https://doi.org/10.1016/j.ijindorg.2005.01.012.

Decker RA, Haltiwanger J, Jarmin RS, Miranda J (2016) Where has all the skewness gone? the decline in high-growth (young) firms in the us. European Economic Review 86:4–23.

Delmar F (2006) Measuring growth: methodological considerations and empirical results. Entrepreneurship and the Growth of Firms 1(1), 62–84, https://doi.org/10.4337/9781781009949.00011.

Delmar F, Davidsson P, Gartner WB (2003) Arriving at the high-growth firm. Journal of Business Venturing 18(2), 189–216, https://doi.org/10.1016/s0883-9026(02)00080-0.

Delmar F, McKelvie A, Wennberg K (2013) Untangling the relationships among growth, profitability and survival in new firms. Technovation 33(8–9), 276–291, https://doi.org/10.1016/j.technovation.2013.02.003.

Doornik JA, Hansen H (2008) An omnibus test for univariate and multivariate normality. Oxford Bulletin of Economics and Statistics 70:927–939, https://doi.org/10.1111/j.1468-0084.2008.00537.x.

Du J, Bonner K (2017) Fast-growth firms in the UK: definition and policy implications. Tech. rep., Queen's Management School Working Paper, https://www.enterpriseresearch.ac.uk/wp-content/uploads/2017/12/ERC-ResPap63-DuBonner-Final.pdf

Du J, Temouri Y (2015) High-growth firms and productivity: Evidence from the United Kingdom. Small Business Economics 44(1), 123–143, https://doi.org/10.1007/s11187-014-9584-2.

Du J, Vanino E (2020) Agglomeration externalities of fast-growth firms. Regional Studies 55(2), 167–181, https://doi.org/10.1080/00343404.2020.1760234.

Erhardt EC (2021) Measuring the persistence of high firm growth: choices and consequences. Small business economics 56(1):451–478.

Foster L, Haltiwanger J, Syverson C (2008) Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? American Economic Review 98(1), 394–425, https://doi.org/10.1257/aer.98.1.394.

Fuss C, Theodorakopoulos A (2018) Compositional changes in aggregate productivity in an era of globalisation and financial crisis. Tech. rep., NBB Working Paper

Gjerlov-Juel P, Guenther C (2019) Early employment expansion and long-run survival: Examining employee turnover as a context factor. Journal of Business Venturing 34(1), 80–102. doi: 10.1016/j.jbusvent.2018.05.005.

Gleeson AM, Ruane F (2009) How important is exporting and ownership to productivity growth? Micro evidence from a small open economy. Applied Economics Quarterly 55(3), 197–217, doi: 10.3790/aeq.55.3.197.

Greiner LE (1989) Evolution and revolution as organizations grow. In: Readings in strategic management, Springer, pp 373–387

Grover Goswami A, Medvedev D, Olafsen E (2019) High-growth firms: Facts, fiction, and policy options for emerging economies. The World Bank, URL https://doi.org/10.1596/978-1-4648-1368-9

Guillamón C, Moral-Benito E, Puente S (2017) High growth firms in employment and productivity: dynamic interactions and the role of financial constraints?

Haltiwanger J, Jarmin RS, Kulick RB, Miranda J (2016) High growth young firms: Contribution to job, output and productivity growth. US Census Bureau Center for Economic Studies Paper No CES-WP-16-49 URL https://doi.org/10.2139/ssrn.2866566

Harris R, Moffat J (2013) The direct contribution of FDI to productivity growth in Britain, 1997–2008. The World Economy 36(6), 713–736, https://doi.org/10.1111/twec.12059.

Henrekson M, Johansson D (2010) Gazelles as job creators: a survey and interpretation of the evidence. Small Business Economics 35(2), 227–244, https://doi.org/10.1007/s11187-009-9172-z.

Hölzl W (2014) Persistence, survival, and growth: a closer look at 20 years of fast-growing firms in Austria. Industrial and Corporate Change 23(1), 199–231, https://doi.org/10.1093/icc/dtt054.

Hsieh CT, Klenow PJ (2009) Misallocation and manufacturing TFP in China and India. The Quarterly Journal of Economics 124(4):1403–1448, https://doi.org/10.1162/qjec.2009.124.4.1403

Kang K (2017) Identifying the dispersion of the Pareto productivity distribution in international trade. Economics Letters 161:108–111, https://doi.org/10.1016/j.econlet.2017.09.030.

Lopez-Garcia P, Puente S (2012) What makes a high-growth firm? A dynamic probit analysis using Spanish firm-level data. Small Business Economics 39(4), 1029–1041, https://doi.org/10.1007/s11187-011-9321-z.

Metcalfe JS (1994) Competition, fisher's principle and increasing returns in the selection process. Journal of Evolutionary Economics 4(4), 327–346.

Mohr V, Garnsey E, Theyel G (2014) The role of alliances in the early development of high-growth firms. Industrial and Corporate Change 23(1), 233–259, https://doi.org/10.1093/icc/dtt056.

Moreno F, Coad A (2015) High-growth firms: Stylized facts and conflicting results. In: Entrepreneurial growth: Individual, firm, and region, Emerald Group Publishing Limited, URL https://doi.org/10.1108/s1074-754020150000017016

Moschella D, Tamagni F, Yu X (2019) Persistent high-growth firms in China's manufacturing. Small Business Economics 52(3), 573–594, https://doi.org/10.1007/s11187-017-9973-4.

Muraközy B, Bisztray M, Reizer B (2018) Productivity differences in Hungary and mechanisms of TFP growth slowdown. European Commission https://doi.org/10.2873/33213

Navaretti GB, Castellani D, Pieri F (2014) Age and firm growth: evidence from three European countries. Small Business Economics 43(4), 823–837, https://doi.org/10.1007/s11187-014-9564-6.

de Nicola F, Muraközy B, Tan SW (2019) Spillovers from high growth firms: Evidence from Hungary. Small

Business Economics 57:127–150, https://doi.org/10.1007/s11187-019-00296-w.

OECD (2010) OECD Studies on SMEs and Entrepreneurship, High-Growth Enterprises: What Governments Can Do to Make a Difference. OECD Publishing, https://doi.org/10.1787/20780990

Olley GS, Pakes A (1992) The dynamics of productivity in the telecommunications equipment industry. Tech. rep., National Bureau of Economic Research, https://doi.org/10.3386/w3977

Penrose ET (1959) The theory of the growth of the firm

Restuccia D, Rogerson R (2017) The causes and costs of misallocation. Journal of Economic Perspectives 31(3), 151–174, https://doi.org/10.1257/jep.31.3.151.

Rovigatti G, Mollisi V (2018) Theory and practice of total-factor productivity estimation: The control function approach using Stata. The Stata Journal 18(3):618–662, https://doi.org/10.1177/1536867x1801800307

Shepherd D, Wiklund J (2009) Are we comparing apples with apples or apples with oranges? Appropriateness of knowledge accumulation across growth studies. Entrepreneurship Theory and Practice 33(1):105–123,https://doi.org/10.1111/j.1540-6520.2008.00282.x

Stoyanov A, Zubanov N (2012) Productivity spillovers across firms through worker mobility. American Economic Journal: Applied Economics 4(2):168–98, https://doi.org/10.1257/app.4.2.168

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.