INSTITUTE OF ECONOMICS
Centre for Economic and Regional Studies

Közgazdaság- és Regionális Tudományi Kutatóközpont
KÖZGAZDASÁG-TUDOMÁNYI INTÉZET

CERS-IE WORKING PAPERS | KRTK-KTI MŰHELYTANULMÁNYOK

# Industry dynamics and high-growth firms' contribution to productivity growth

MÁRTA BISZTRAY – FRANCESCA DE NICOLA – BALÁZS MURAKÖZY

**CERS-IE WP – 2020/47**

November 2020

https://www.mtakti.hu/wp-content/uploads/2020/11/CERSIEWP202047.pdf



INSTITUTE OF ECONOMICS, CENTRE FOR ECONOMIC AND REGIONAL STUDIES,
BUDAPEST, 2020

ABSTRACT

This paper investigates the contribution of high-growth firms (HGFs) to aggregate productivity growth. Four stylized facts emerge. First, HGFs mainly contribute to productivity growth during their high-growth phase but not afterwards. Second, their contribution varies substantially across industries and it is not necessarily positive. Third, the impact on productivity depends on how HGFs are defined. Output-based HGFs substantially outperform employment-based ones in terms of their productivity contribution while the difference in terms of job creation is low. Fourth, HGFs' contribution to productivity is higher in industries where industry dynamics favor growing firms, captured by the strength of reallocation and the relationship between productivity growth and size growth. We present a simple model to show that these patterns arise naturally under realistic correlation structures. Our results suggest that policies supporting HGFs may focus on firms increasing their sales, and these can effectively be complemented by framework policies promoting efficient reallocation.



JEL codes: L25, O40

Keywords: high-growth firms, productivity growth, reallocation, industry dynamics



Márta Bisztray
Centre for Economic and Regional Studies, Budapest
and
Corvinus University Budapest, CERGE-EI Foundation Teaching Fellow
e-mail: bisztray.marta@krtk.mta.hu

Francesca de Nicola
World Bank, Washington DC
e-mail: fdenicola@worldbank.org

Balázs Muraközy
University of Liverpool Management School, Liverpool
and
Centre for Economic and Regional Studies, Budapest
e-mail: Balazs.Murakozy@liverpool.ac.uk

# Iparági dinamika és a gyors növekedésű cégek hozzájárulása a termelékenység növekedéshez

BISZTRAY MÁRTA – FRANCESCA DE NICOLA – MURAKÖZY BALÁZS

ÖSSZEFOGLALÓ

Tanulmányunk a gyors növekedésű cégek aggregált termelékenység növekedéshez való hozzájárulását vizsgálja. Négy stilizált tényt állapíthatunk meg. Először, a gyors növekedésű cégek elsősorban a gyors növekedési periódus alatt járulnak hozzá az aggregált termelékenység növekedéséhez, utána már nem. Másodszor, a hozzájárulásuk iparáganként lényegesen eltérő és nem feltétlenül pozitív. Harmadszor, a termelékenységre gyakorolt hatás függ a pontos definíciótól. A gyors árbevétel növekedésű cégek termelékenységhez való hozzájárulása lényeges nagyobb, mint a gyors létszám növekedésűeké, míg a munkahelyteremtés tekintetében sokkal kisebb köztük a különbség. Negyedszer, a gyors növekedésű cégek termelékenység növekedéshez való hozzájárulása magasabb ott, ahol az iparági dinamika kedvez a növekvő cégeknek, amit a reallokáció erősségével, illetve a termelékenység és méretnövekedés közti kapcsolattal mérhetünk. Egy egyszerű modell segítségével bemutatjuk, hogy ezek a mintázatok természetesen adódnak realisztikus korrelációs struktúrák esetén. Mindezek alapján a gyors növekedésű vállalatokat célzó intézkedéseknek érdemes lehet az árbevételüket növelő cégekre összpontosítani, valamint ezeket jól kiegészíthetik a hatékony reallokációt ösztönző általánosabb intézkedések.

JEL: L25, O40

Kulcsszavak: gyors növekedésű vállalatok, termelékenység növekedés, reallokáció, iparági dinamika

# Industry dynamics and high-growth firms' contribution to productivity growth

Márta Bisztray[1] Francesca de Nicola[2] Balázs Muraközy[3]

November 18, 2020

**Abstract**

This paper investigates the contribution of high-growth firms (HGFs) to aggregate productivity growth. Four stylized facts emerge. First, HGFs mainly contribute to productivity growth only during their high-growth phase but not afterwards. Second, the contribution during this phase varies substantially across industries and it is not necessarily positive. Third, the impact on productivity depends on how HGFs are defined. Output-based HGFs substantially outperform employment-based ones in terms of their productivity contribution while the difference between the two firm groups is much lower in terms of job creation. Fourth, HGFs' contribution to productivity is higher in industries where industry dynamics favor growing firms, captured by the strength of reallocation and the relationship between productivity growth and size growth. We present a simple model to show that these patterns arise naturally under realistic correlation structures. Taken together, these results suggest that specific policies supporting HGFs may focus on firms which increase their sales, and HGF policies can effectively be complemented by more general framework policies promoting efficient reallocation.

**Keywords:** high-growth firms, productivity growth, reallocation, industry dynamic, Hungary

[1]Centre for Economic and Regional Studies, Budapest and Corvinus University Budapest, CERGE-EI Foundation Teaching Fellow

[2]World Bank, Washington DC

[3]Corresponding author. University of Liverpool Management School, Liverpool and CERS, Budapest. email murakozy@liverpool.ac.uk, Chatham St, Liverpool L69 7ZH, UK. 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".

# 1 Introduction

High-growth firms (HGFs) have increasingly attracted the attention of policy makers mainly for their job-creating role. However, when evaluating the high-growth phenomenon and the policies that support it, it is not only the number of jobs created that is relevant, but also the "quality" of those jobs. An important measure of quality is the productivity level of these new jobs: do HGFs create exceptionally productive jobs—as often assumed—or are these jobs mostly less productive, contributing little to aggregate productivity growth?

This paper sheds light on this question by quantifying the contributions of HGFs to industry-level TFP and labor productivity growth based on detailed microdata from Hungary. We present results for all of the most common HGF definitions in the literature including the "OECD definitions", which require firms to increase their employment or sales by 20% per year for 3 years, and the "Birch definitions", which rely on an average of absolute and relative growth.

A useful way to think about how a group of firms contribute to productivity growth is to start from the framework of Foster et al. (2008). This framework decomposes industry-level productivity growth of (continuing) firms into three components.[1] First, the *within* term captures the contribution resulting from firms increasing productivity, but disregarding any changes in their market share. Second, the *between* (reallocation) term captures the extent to which the expansion of initially more productive firms contributes to the increase in industry-level productivity. Third, the *cross* term captures the contribution from the correlation between size growth and productivity growth.

This additive decomposition allows distinguishing between HGFs and other firms' contributions to productivity growth within an industry (similarly to Haltiwanger et al. 2016). Quantifying the within, between and cross terms separately for HGFs and other firms allows us to account for two possible channels of productivity growth. First, HGFs may improve their productivity during their high growth phase. This should be reflected both by the HGF *within* term (because high-growth firms also experience improving

[1] We abstract away from entry and exit, because most definitions of HGFs assume that the firm initially operates above a certain size threshold and, by definition, HGFs cannot exit during their high-growth phase.

productivity) and the *cross* term (because there is a positive correlation between size and productivity growth within the HGF group). Second, HGFs may already be exceptionally productive to begin with, hence reallocating resources towards them improves overall productivity, as captured by the HGF *between* term.

We document four stylized facts using data from Hungary. First, on average, HGFs raise productivity only during their high-growth phase, if at all, but not afterwards. Therefore, our analysis—and probably policymakers' attention—should focus on the high-growth phase.

Second, the contribution during the high-growth phase varies widely across industries and years. While HGFs contribute more than 50% to industry-level productivity growth in about 25% of the cases,[2] their contribution is actually negative for 25% of the (industry-period level) observations. Negative contribution mainly comes from negative productivity growth during the high-growth phase.

Third, the definition of HGFs matters when assessing HGFs' contribution to aggregate productivity growth. The most frequently-used HGF definitions differ in two key dimensions. The first dimension is whether they focus on input, i.e. employment, or output, i.e. sales, growth.[3] We find that output-based HGFs tend to experience positive productivity growth (15% on average) while input-based HGFs typically face a fall in their productivity (-24% on average).[4] Even though this difference in within-firm productivity growth is somewhat offset by the lower initial productivity of the sales HGFs, the average (median) contribution of sales-based HGFs is 1.6 (2.6) percentage points higher than that of employment-based HGFs for the 3-year period.[5] The second key dimension of HGF definitions is whether they are based solely on relative growth (OECD definitions) or also take absolute growth into account (Birch definitions).[6] Definitions identifying HGFs solely based on relative growth are likely to capture small firms, therefore, their aggregate

[2] Except for the OECD employment definition discussed later.

[3] While alternative measures also exist, the literature suggests that sales and employment are the preferable ones for measuring HGFs (Delmar et al. (2003)).

[4] This is in line with other findings of the literature. Daunfeldt et al. (2014) show that Swedish firms with high sales growth are also likely to have high productivity growth, but firms with high employment growth do not. Du & Temouri (2015) find a higher post-high-growth period productivity growth for UK HGFs defined based on sales. Haltiwanger et al. (2016) emphasize that sales-based HGFs make a disproportionate contribution to productivity growth in the US. Mason & Brown (2013) show that in spite of their higher than average productivity, employment-based HGFs in Scotland have a limited contribution to aggregate productivity growth. For Italian firms, Arrighetti & Lasagni (2013) find that only sales-based HGFs have typically high productivity.

[5] All numbers are based on the OECD definitions.

[6] See the review paper of Coad et al. (2014) for further details on HGF-measurement in the literature.

contribution is typically more limited.

Our fourth finding is that some basic characteristics of industry dynamics, captured by two key moments, substantially influence the contribution of HGFs to industry-level productivity growth. The first key moment is the correlation between the initial productivity level and subsequent size growth, which captures the strength and efficiency of reallocation within the industry. Under stronger reallocation, HGFs tend to be more productive initially, which leads to a more positive HGF *between* component. The second moment is the correlation between size growth and productivity growth. This correlation captures the magnitude of financial and other constraints firms face when growing. We show that a stronger correlation between productivity growth and size growth leads to a higher productivity contribution of HGFs via higher *within* and *cross* terms.

After presenting the results, we build a simple model and simulation to explain the key patterns in the data. This model starts from a joint distribution of the variables relevant for HGF productivity contribution: initial size, initial productivity, productivity growth and size growth. Correlations between these variables capture the key empirical moments of industry dynamics. Our aim with this exercise is threefold. First, we aim at explaining qualitatively the empirical patterns we find based on simple assumptions. By finding that even this highly simplified framework reproduces all our results, we conclude that they mostly come from industry-level correlations rather than from some peculiarities of HGFs. Second, the framework may help policymakers in designing effective policies by emphasizing the role of some easily interpretable and quantifiable industry and firm characteristics. Finally, the framework can be generalized to model and understand the productivity contributions of other groups of firms, for example SMEs, as we discuss. The simple model we build can reproduce the patterns we find in the empirical exercise.

We argue that these findings are of high policy interest. The productivity evolution of high-growth firms is not relevant only because of HGFs' direct contribution to productivity growth but also because more spillovers may be expected from more productive firms (e.g. Stoyanov & Zubanov 2012). The insight that sometimes HGFs do actually contribute negatively to productivity growth should warn policy makers 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 which are very productive to start with may miss out these firms. Our final insight emphasizes that HGFs are likely to contribute more in environments which are in general conducive to efficient reallocation. This implies that general framework policies improving reallocation can strongly complement specific policies promoting HGFs.

We contribute to four strands of the literature. First, there is some evidence from the literature on the links between the productivity of firms and their HGF status. HGFs tend to have a higher productivity level (Acs et al. 2008, Bianchini et al. 2017, Moschella et al. 2019), but not necessarily when high growth is measured in terms of employment (Arrighetti & Lasagni 2013). Firms with high TFP growth are more likely to become sales-based HGFs, and sales-based HGFs tend to have a higher productivity growth after their high-growth period (Du & Temouri 2015). At the same time, Daunfeldt et al. (2014) suggest a trade-off between growth in terms of employment and in terms of productivity. 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, we know from previous studies that there exists a large heterogeneity in the growth patterns of high-growth firms, thus the way one define HGFs is crucial. Among others, Delmar (2006) and Shepherd & Wiklund (2009) compare the various definitions used in the previous literature. HGF definitions differ in terms of the indicator used: the definitions are most frequently based on sales and employment, but it can also be a combination of these two (Acs et al. 2008, Moschella et al. 2019), productivity or value added (Daunfeldt et al. 2014), more rarely firm performance, market share or assets (Delmar 2006)), in terms of the formula for measuring growth (relative or absolute growth (Delmar 2006), or a combination of the two

(Birch 1981). Definitions also differ in terms of the the time span considered (1, 3 or 5 years most frequently Delmar 2006), and in the type of growth (organic or the result of acquisition). The literature has studied many dimensions along which HGFs defined in various ways can differ. This includes the persistence of HGF status[7], different characteristics of HGFs[8], and the factors predicting HGF status[9]. 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 & Johansson 2010), there are only 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 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. Compared to these results, we also look at the factors leading to cross-industry differences in the between and within terms of HGFs' contribution to industry-level productivity. Moreover,

[7] We know from the literature that high growth is rather a temporary feature than a persistent firm characteristic (Daunfeldt & Halvarsson 2015), but Hölzl (2014) show that Birch-type HGFs are more persistent than the OECD-type HGFs, and Daunfeldt et al. (2014) has the same result for HGFs defined using absolute growth rather than relative one.

[8] As Coad et al. (2014) emphasizes, different HGF definitions capture a different set of firms. HGFs are generally younger but not only startups, not necessarily small and present in all sectors (Acs et al. 2008, Coad et al. 2014, Grover Goswami et al. 2019, e.g.), but only more productive and having more concentrated ownership if they are defined based on sales growth (Arrighetti & Lasagni 2013).

[9] There are some characteristics with which firms are more likely to become HGFs, like higher previous employment growth, lower inventories, higher short-term liabilities (Coad & Srhoj 2019), certain human resource practices, newness or access to credit (Lopez-Garcia & Puente 2012). Coad & Srhoj (2019) emphasize that certain factors like exports or growth in assets, are sensitive to the growth indicator, and they conclude that it is difficult to predict HGF status.

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 & Klenow 2009, Bartelsman et al. 2013, Restuccia & Rogerson 2017, Baqaee & Farhi 2020). Some papers already looked at how specific firm groups, like foreign-owned (Balsvik & Haller 2006, Harris & Moffat 2013) or exporters (Gleeson & Ruane 2009, Fuss & Theodorakopoulos 2018) contribute to aggregate productivity via within-firm growth and reallocation. We focus on the role of HGFs in aggregate reallocation and by linking 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, 2016). We show that framework conditions are important for the contribution of a firm group to industry-level productivity, and by mapping general framework conditions to simple correlations of growth dynamics parameters, we provide an intuitive mechanism for that.

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 investigate differences between them. Section 4 presents our findings on the contribution of HGFs to aggregate productivity and the impact of industry characteristics. Section 5 introduces our 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 that require 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 (CERS-HAS) has extensively cleaned and harmonized the data. Nominal variables are deflated by the appropriate 2-digit industry level deflators from OECD STAN.[10]

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 which never had more than 4 employees, because standard HGF definitions require firms to have at least 5 or 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.

## 2.2 Decomposition

Our decomposition is based on Foster et al. (2008), who distinguish aggregate productivity growth in 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. $\Delta$ always denotes the change between $t_0$ and $t$. $C$ stands for

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

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 firms, increased their 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).[11] The decomposition becomes:

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

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. The HGF

[11] 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 $t$.

*cross* term $\Delta\theta_{i,t}$ is positive by definition for all HGFs and, therefore, the sign of the *cross* term is primarily determined by the sign of productivity changes. Therefore, 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 between 2001 and 2016 for each 2-digit industry. Productivity decomposition is usually less noisy in such 'medium-term' periods and they may better reflect the timeline of such economic processes as 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. The cohort of year $t$ consists of firms which were in their high-growth phase between $t$ and $t + 3$.

# 3 Definitions

The focus of this section is to introduce the different HGF definitions we use and to present a number of patterns. These help us to understand how the different types of HGFs contribute to aggregate productivity growth and 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 which were HGFs between 2013 and 2016 (the 2013 cohort), our last cohort. The patterns are similar for other cohorts, and we note any exception.

## 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 1 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 which 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.

Table 1: 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 $(sales_t - sales_{t-3})\frac{sales_t}{sales_{t-3}}$ |

## 3.2 How dissimilar are the different HGFs?

We now show to what extent and how firms covered by the different definitions are dissimilar from each other. For its policy implications, it is important to assess whether differences in productivity growth contributions stem from the use of different definitions. We proceed in three steps. First, we document the overlap between the different definitions. Second, we show how the economic footprint and its change differ across groups. Finally, and most relevantly for the productivity decomposition, we quantify productivity and employment growth during and around the high-growth phase.

We find that if two definitions differ in only the growth measure, the overlap is between 45-63%, while if they differ in both dimensions, the overlap is closer to 26-46% (Table 2). In other words, the various

HGF measures overlap to some extent but capture a quite different set of firms.[12] Based on these differences it is plausible that differently defined HGFs contribute very differently to aggregate productivity.

Table 2: Overlap between different HGF definitions, 2013

| | OECD (emp) | OECD (sales) | Birch (emp) | Birch (sales) |
|---|---|---|---|---|
| OECD (emp) | 100.0% | 57.9% | 57.1% | 26.0% |
| OECD (sales) | 57.9% | 100.0% | 46.1% | 62.5% |
| Birch (emp) | 57.1% | 46.1% | 100.0% | 45.3% |
| Birch (sales) | 26.0% | 62.5% | 45.3% | 100.0% |

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

Next, we document the footprint, in terms of the number of firms, of the different types of HGFs. By definition, the Birch definitions cover 5% of firms, therefore, the question is only relevant for the OECD definitions. Figure 1 shows the share of HGFs in 2013 by industry for the two OECD definitions.[13] With few exceptions, the share of employment OECD HGFs is between 1 and 6% and that of sales HGFs is between 3 and 10%.[14] 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 then their employment during the 3-year period, which suggests that increased productivity is a key source of growth for many sales HGFs.[15] This is the first indication that sales HGFs are more likely to increase their productivity and, therefore, to contribute more to industry productivity growth.

The empirical relevance of the two dimensions which distinguish between the different HGF definitions can be more clearly seen when we quantify the economic footprint of HGFs by their industry share in terms of employment and sales. This is illustrated by Figure 2, 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.

HGFs captured by the OECD and Birch definitions have a very different economic footprint. For

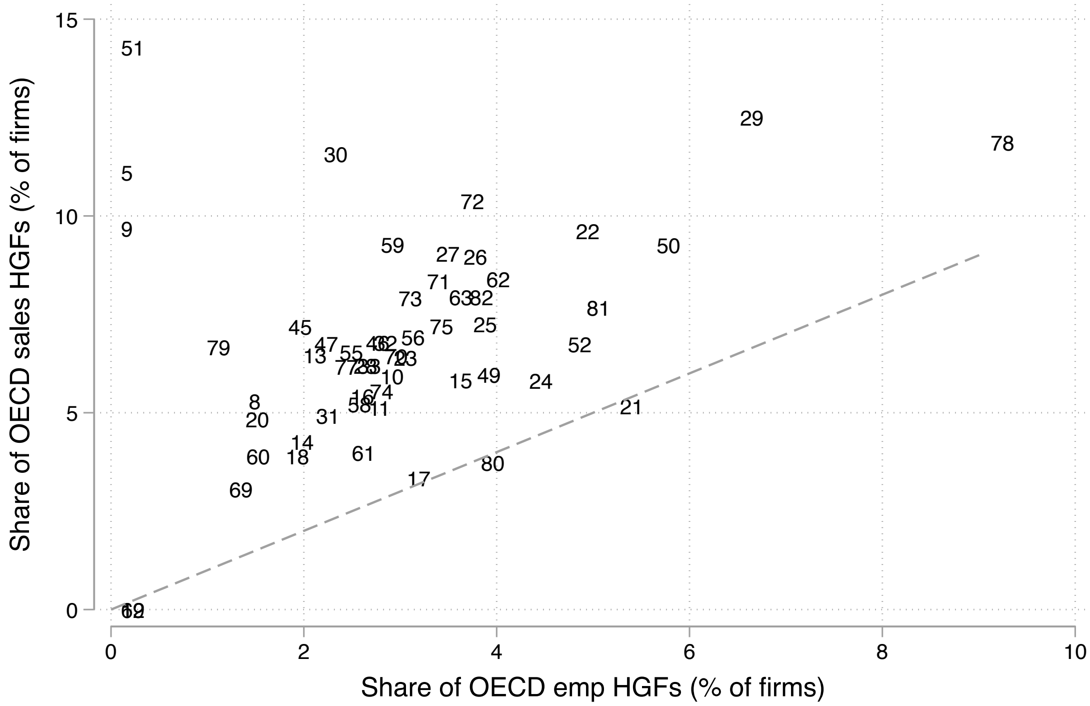
[12]As a comparison, Delmar et al. (2003) show that from the Swedish HGFs defined based on absolute growth about 15% has high growth only in sales but not in employment and 20% has high growth only in employment but not in sales. Shepherd & Wiklund (2009) find that the correlation between employment 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.

[13]The figure omits the outlier industry 7 (Mining of metal ores), where 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 5 or more employees in at least one year.

[14]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), Figure 1.1.

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

Figure 1: Share of HGFs in the total number of firms, 2013



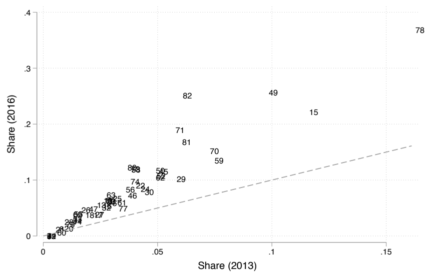

The figure shows the share of firms which were HGFs between 2013 and 2016 according to the OECD (employment) and OECD (sales) HGF definitions by 2-digit industry. The sample includes firms which had at least 5 employees at any point.

OECD HGFs, the median initial employment and sales shares are around 4% and 8%, respectively, while the typical initial shares are around 20-30% 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.

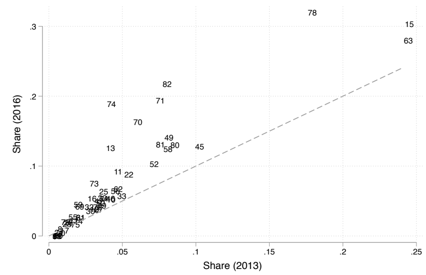
Regarding employment and sales HGFs, there are three clear patterns in the figures. First, both types of firms increase their share both in terms of employment and sales. Second, as expected, HGFs increase their share more in the dimension corresponding to the definition. Combining these two observations 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.

The above patterns suggest that the different definitions cover markedly different firms with a different productivity performance. Table 3 investigates the latter question more explicitly with TFP, while

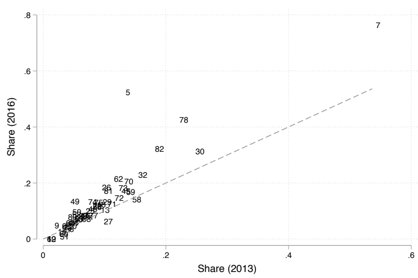
Figure 2: Employment and sales shares of different HGFs in the beginning and at the end of the high-growth period 2013



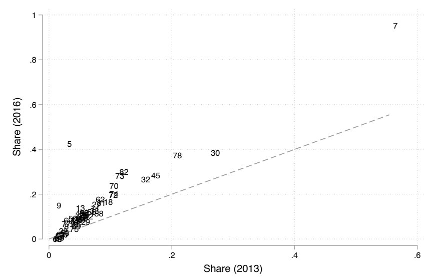

(a) OECD (emp), employment share



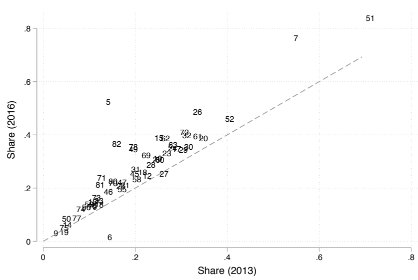

(b) OECD (emp), sales share



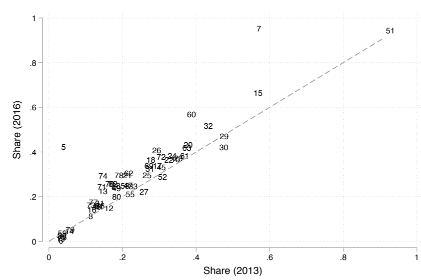

(c) OECD (sales), employment share



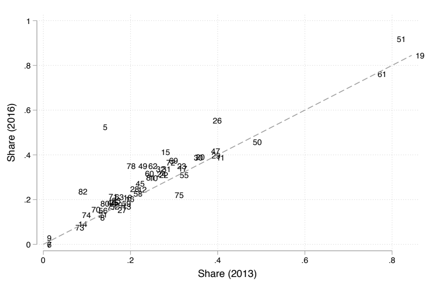

(d) OECD (sales), sales share



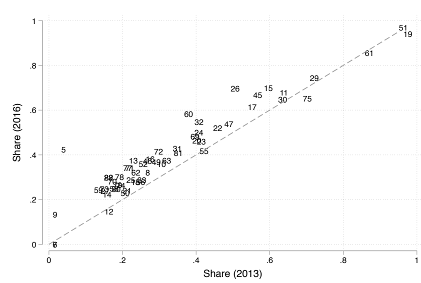

(e) Birch (emp), employment share




(f) Birch (emp), sales share




(g) Birch (sales), employment share




(h) Birch (sales), sales share

This figure shows the employment and sales share in 2013 and 2016 of firms which were HGFs in 2013 according to the different definitions. The numbers show the NACE code of 2-digit industries and the line is the 45-degree line. The sample includes firms which had at least 5 employees at any point.

Table A2 repeats this exercise in terms of labor productivity. The tables show HGF performance one year before (t-1) and four years after (t+4) the beginning of the high-growth phase. 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%).

Let us start with the relative/absolute dichotomy. We find that the two OECD definitions identify firms that are initially significantly smaller than the average firm. On average, these HGFs employ 27-42% fewer employees before their high-growth phase than the average firm, and 18-86% more after that. In contrast, according to the Birch method, HGFs are 104-204 percent larger than the average firm even before the high-growth phase.

Related to their larger size, Birch HGFs are also more productive.[16] The initial productivity premia are substantially larger for the Birch definitions (25-87%) compared to the OECD definitions (-8-37%). These differences in initial productivity premia suggest a larger potential reallocation effect for Birch HGFs.

In line with the patterns in Table 2, irrespective of the definition used, HGFs create a significant number of jobs. As expected, average employment growth is somewhat higher under the employment-based than under the sales-based definition. Using employment-based definitions, relative to the average firm, the average HGFs' employment grows by 118 percentage points based on the OECD definition, and by 290 percentage points based on the Birch definition. Using sales-based definitions yields smaller estimates, but the magnitudes are still noteworthy: 58 and 166 percentage points according to the OECD and the Birch definition, respectively.

Regarding productivity, we find that input-based OECD-type HGFs are more productive initially: their initial productivity premium is 25% on average while it is -3% for output-based HGFs. The pattern is just the opposite for the Birch definition. Additionally, while output-based HGFs have stable (for Birch) or growing (around 15 percentage points for OECD) productivity during the high-growth phase, HGFs that are defined based on input growth experience a productivity decline of a similar magnitude (-24 percentage

[16] As an example, Medrano-Adán et al. (2019) present a model which explains the pattern of positive correlation between size and productivity. Leung et al. (2008) also refer to theoretical and empirical evidence.

points for the OECD and -31 percentage points for the Birch definition).

An important message of this table — in line with the previous figures — is that there seems to be some, but not very strong trade-off between job creation and productivity growth: 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 a large number of jobs at the same time, while employment-based HGFs face declining productivity.

## 3.3 HGF dynamics

When we want to evaluate the effect of HGFs on the economy, the appropriate time frame is far from evident. It is possible that, for example, the high growth in employment is followed by a productivity increase many years later. Alternatively, HGFs may follow quite risky strategies, and any gains generated during the high-growth phase may be undone by regression to the mean or increased probability of exit. The aim of this subsection is to provide descriptive evidence for these eventualities. We do so in three steps. First, we investigate the productivity trends after the high-growth phase by further analyzing Table 3. Second, we use a transition matrix approach to check the persistence of the HGF status itself and whether the same firm is likely to undergo different types of growth subsequently. Finally, we check whether HGFs are more likely to exit compared to other firms.

Table 3 allows us to follow the firms before and after their high-growth phase. Compared to the developments during the high-growth period, the productivity and employment changes before and after are relatively small. Still, the high-growth phase seems to be typically preceded by productivity growth for all the different definitions, although to a different extent.[17] The gains in productivity seem to largely persist until 3 years after the high-growth phase, while the employment level increases further.

The fact that these pre and post trends are relatively modest suggest that high growth seems to be

[17]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, defined as a combination of sales and employment growth.

Table 3: Employment and TFP before and after the high-growth phase by definition

| | oecd3 (emp) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | employment | | | | TFP | | | |
| year | t-4 | t-1 | t+4 | t+7 | t-4 | t-1 | t+4 | t+7 |
| 1999 | | 60.2% | 171.1% | 182.9% | | 137.4% | 104.1% | 104.7% |
| 2002 | 53.9% | 58.3% | 174.3% | 171.7% | 116.5% | 134.3% | 103.4% | 106.7% |
| 2005 | 63.2% | 67.0% | 186.3% | 193.2% | 111.5% | 123.6% | 102.9% | 104.2% |
| 2008 | 58.8% | 58.1% | 185.0% | | 92.9% | 112.9% | 93.5% | |
| 2011 | 64.9% | 62.6% | | | 111.7% | 115.9% | | |
| average | 60.2% | 61.3% | 179.2% | 182.6% | 108.1% | 124.8% | 100.9% | 105.2% |

| | oecd3 (sales) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | employment | | | | TFP | | | |
| year | t-4 | t-1 | t+4 | t+7 | t-4 | t-1 | t+4 | t+7 |
| 1999 | | 70.9% | 118.4% | 127.6% | | 99.2% | 109.2% | 103.8% |
| 2002 | 70.2% | 73.9% | 134.7% | 145.6% | 95.0% | 102.6% | 113.4% | 111.4% |
| 2005 | 65.7% | 70.3% | 129.5% | 138.8% | 93.2% | 94.2% | 111.0% | 109.9% |
| 2008 | 70.5% | 68.1% | 134.5% | | 86.5% | 97.9% | 114.7% | |
| 2011 | 72.1% | 71.2% | | | 92.9% | 91.9% | | |
| average | 69.6% | 70.9% | 129.3% | 137.3% | 91.9% | 97.2% | 112.1% | 108.4% |

| | birch3 (emp) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | employment | | | | TFP | | | |
| year | t-4 | t-1 | t+4 | t+7 | t-4 | t-1 | t+4 | t+7 |
| 1999 | | 203.8% | 499.5% | 545.7% | | 152.4% | 109.5% | 110.8% |
| 2002 | 210.2% | 210.9% | 498.3% | 531.7% | 132.2% | 142.0% | 110.0% | 111.9% |
| 2005 | 259.4% | 246.3% | 568.7% | 603.0% | 131.4% | 151.4% | 113.9% | 116.2% |
| 2008 | 221.9% | 204.7% | 477.1% | | 113.8% | 124.8% | 103.6% | |
| 2011 | 233.5% | 237.6% | | | 119.8% | 129.5% | | |
| average | 231.2% | 220.7% | 510.9% | 560.1% | 124.3% | 140.0% | 109.2% | 113.0% |

| | birch3 (sales) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | employment | | | | TFP | | | |
| year | t-4 | t-1 | t+4 | t+7 | t-4 | t-1 | t+4 | t+7 |
| 1999 | | 304.5% | 480.8% | 522.5% | | 186.6% | 164.7% | 154.3% |
| 2002 | 257.4% | 246.1% | 432.0% | 496.2% | 163.9% | 172.6% | 173.5% | 156.1% |
| 2005 | 249.3% | 253.9% | 436.5% | 487.1% | 165.1% | 175.2% | 171.6% | 164.5% |
| 2008 | 207.7% | 205.5% | 348.3% | | 151.5% | 159.1% | 175.2% | |
| 2011 | 278.8% | 283.0% | | | 153.9% | 164.8% | | |
| average | 248.3% | 258.6% | 424.4% | 501.9% | 158.6% | 171.7% | 171.3% | 158.3% |

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%).

a transitory phase in most firms' life.[18] From an empirical point of view, this finding suggests that one can safely focus on the changes during the high-growth phase when assessing the contributions of these firms to aggregate productivity and can mostly ignore what happens before and after the high-growth phase.[19]

While Table 3 provides no evidence for extra productivity growth on average after the high-growth period, it is still possible that for many firms the high-growth phase takes substantially longer than 3 years, or that one type of high growth is 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 4, 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 is that HGF status is persistent: depending on the definition, a firm which was HGF between t-3 and t is about 2.5-6 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 much stronger than that of the OECD-type HGF status. The main driver of this pattern might be the importance of firm size, quite a persistent feature, in the Birch-type definition. 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.

Table 4 also provides some limited evidence for different types of high-growth periods following each other.[20] 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.[21]

[18] Previous research already provides evidence for the temporary nature of the high-growth phase (see e.g. Acs et al. (2008) using US data, Daunfeldt & 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 & Srhoj (2019) find that Croatian and Slovenian firms with higher previous employment growth become HGFs with a higher probability. Additionally, Lopez-Garcia & Puente (2012) demonstrate that more than half of Spanish HGFs were already HGFs in the previous period.

[19] As we will see in the next section, the Birch (sales) definition is a partial exception. These firms are highly productive prior to their high-growth phase, and grow both in size and productivity in the pre-period. As a result, they contribute substantially to productivity growth even between $t-3$ and $t$, because both their between and within terms are positive.

[20] Davidsson & 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.

[21] 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.

Table 4: 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 | 3.8% | 9.0% | 4.7% | 4.8% |
| OECD (emp) | 9.4% | 17.4% | 14.4% | 9.1% |
| OECD (sales) | 9.3% | 15.1% | 9.9% | 9.4% |
| Birch (emp) | 6.6% | 13.5% | 25.9% | 18.1% |
| Birch (sales) | 7.9% | 10.7% | 24.8% | 29.2% |

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.

As we have mentioned, if HGFs are more likely to exit because their growth strategies are more risky, their long-term contribution may be small. To explore this possibility, Figure 3 shows survival rates calculated from 2007 for firms that were HGFs between 2004 and 2007 and for other firms.[22] We find that all types of HGFs are actually less likely to exit than non-HGFs.[23] 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.

The evidence presented in this subsection suggests that productivity change is concentrated in the high-growth phase. This observation is reinforced by finding that HGF status is typically transitory. Also, excessive exit is unlikely to be an important factor determining long-term HGF contributions.

# 4 HGFs' contribution to total productivity growth

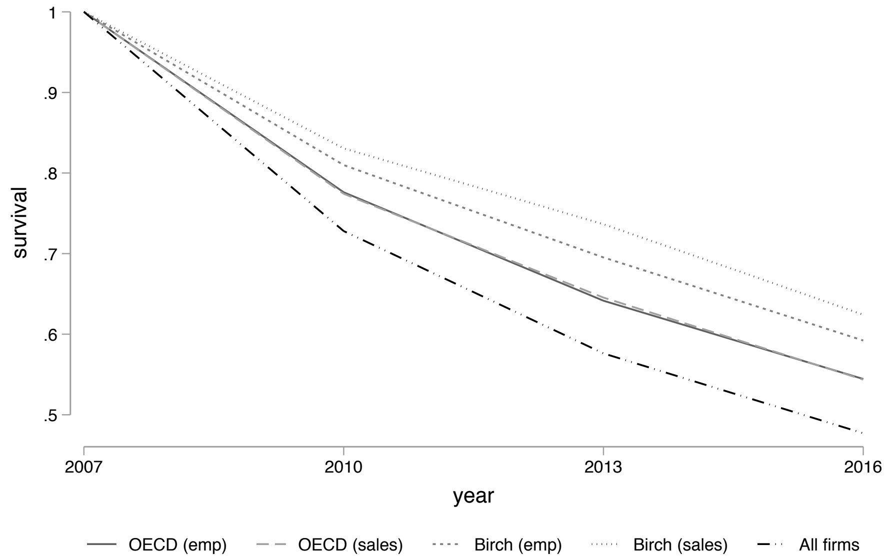
While the previous section has provided descriptive evidence about HGFs' productivity and employment growth, here we focus explicitly on HGFs' contribution to productivity growth, which is the main interest of this paper.

We rely on the decomposition methodology presented in Subsection 2.2. Our unit of observation

[22] 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 which 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.

[23] Acs et al. (2008) 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. Gjerløv-Juel & 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.

Figure 3: 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 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.

## 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, $cont_{jc}$ (Table 5).

The total HGF contribution differs strongly across definitions. In general, it is larger for Birch definitions than for OECD definitions and it is larger for sales-based definitions compared to employment-based definitions. Birch-sales HGFs contribute the most on average (4.12 percentage points), followed by OECD-sales HGFs (2.82 pp.). The contribution of employment HGFs is substantially lower, with 0.24 pp. for the OECD and 0.77 pp. for the Birch definition.

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 negative. This results from the fact that productivity typically increases (or is stable) for sales-based definitions while it falls for the employment-based definitions (Table 3). 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 the employment growth rate during the high-growth period. The relatively low between contribution of OECD (sales) HGFs results from their lower initial productivity level. 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 subsection.

Table 5: Decomposing the HGF contribution to TFP growth

| | HGF within | HGF between | HGF cross | HGF total |
|---|---|---|---|---|
| OECD (emp) | -0.29% | 1.54% | -1.01% | 0.24% |
| OECD (sales) | 2.28% | 0.94% | -0.40% | 2.82% |
| Birch (emp) | -0.19% | 2.17% | -1.21% | 0.77% |
| Birch (sales) | 3.40% | 1.73% | -1.02% | 4.12% |

This table presents the employment weighted average (across year-cohorts) of the different components of the total HGF contribution for the four definitions.

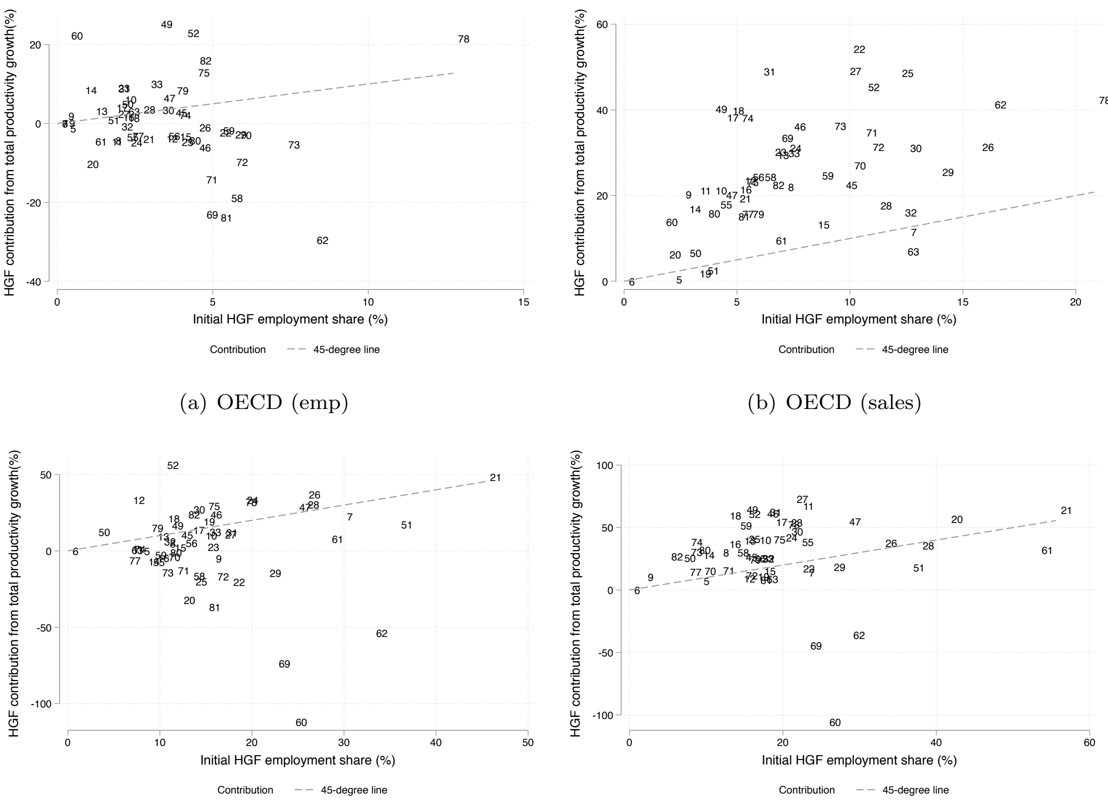
A policy-relevant insight from this pattern is that output HGFs, which contribute more than input HGFs, typically contribute via their *within* term while their initial productivity level is not especially high (Figure 3). Therefore, HGF policies which only target firms with high initial productivity 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 which 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 4 illustrates this relationship at the industry level, with the average initial employment share of HGF cohorts on the horizontal and the average ratio of total HGF contribution and industry-level TFP growth on the vertical axis.[24]

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

There is once more quite an 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%).[25]

Figure 4: 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 when industry productivity growth was negative. The dashed line is the 45-degree line.

The average effects in Table 5 hide a large degree of heterogeneity. Figure 5 and Figure A2 (for labor productivity) focuses on the heterogeneity of $cont_{jc}$.

Clearly, the HGF contribution is not necessarily positive for either of the definitions. The $10^{\text{th}}$

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

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. OECD (sales) firms' contribution distribution dominates that of OECD (emp) firms and the same is true for Birch (sales) and Birch (emp) firms. 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 (sales) distribution. In contrast, the OECD (sales) distribution is less "risky", with few negative and some decent-sized contributions.

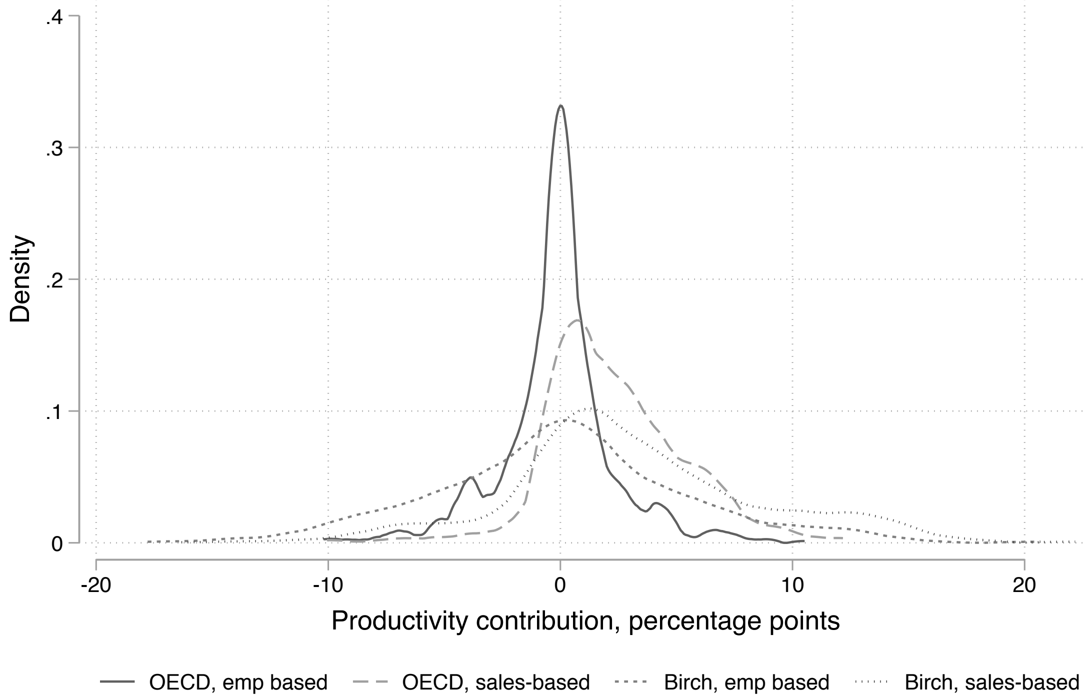
This subsection has shown that different types of HGFs contribute very differently to industry-level productivity growth, with sales-based HGFs contributing consistently above their economic weight and employment HGFs typically contributing below their weight and often negatively. We have also documented that the contributions vary significantly across industries and periods, which may depend on key industry parameters, as we will investigate further in Subsection 4.3.

## 4.2 The contribution of HGF cohorts over time

While the lack of average pre- and post-trends presented in Table 3 is suggestive that HGF productivity and size growth are concentrated in the HGF phase, we use an event study methodology to investigate whether these findings can be generalized to the productivity contribution itself.[26]

[26] The results in Table 3 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 also allows us to quantify statistical uncertainty.

Figure 5: Productivity contribution of HGFs by industry-cohort




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

For the event study, we extend our previous notation as we need to follow each cohort ($c$) for a number periods ($p$). [27] Here we denote contributions by $cont_{jcp}$.

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. $\delta_{cp}^{\theta}$ 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).[28] $\eta_j$ and $\xi_p$ are 2-digit industry and period fixed effects, respectively. We weight the regressions by the number of employees of the industry at

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

[28] We choose this for the omitted period because this is the last period to ensure that any pre-trends before the high-growth period are clearly visible.

the beginning of the period.

Figure 6 shows the results for TFP and Figure A1 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 $t_0 - 6$.[29] 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 the Birch (sales) HGFs by 3.1 percentage points.[30] This contrasts strongly with 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 in an unusual way 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 HGFs, which is about half as large as the one observed during their high-growth phase. This suggests that high sales growth of larger firms often takes place after a period when both productivity and size is growing. [31] Importantly, in line with our earlier observations in 3.3, 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.

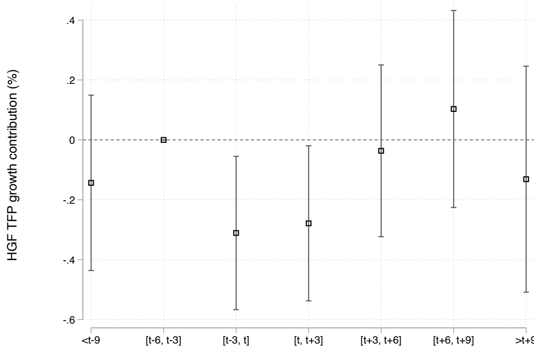
[29]The event study shows whether the contribution is higher relative to this base period. We can only estimate the relative contribution due to the set of fixed effects.

[30]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. Taking into account 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.

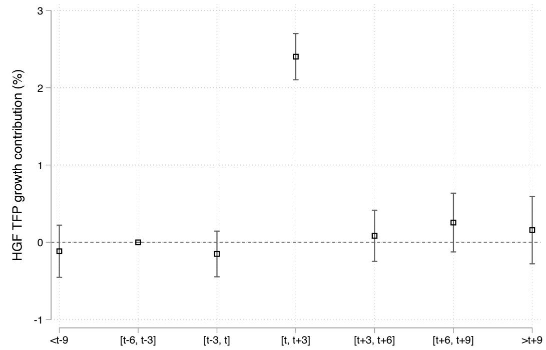
[31]This is documented by Table 3, in which the Birch sales group is the only one in which both productivity and size is increasing before the high-growth period. Given the high productivity of these firms, both their within and between components will be positive in the period between $t_0 - 3$ and $t_0$.

Another factor behind this finding is the unusual persistence of the Birch sales HGFs status, see Figure 4.

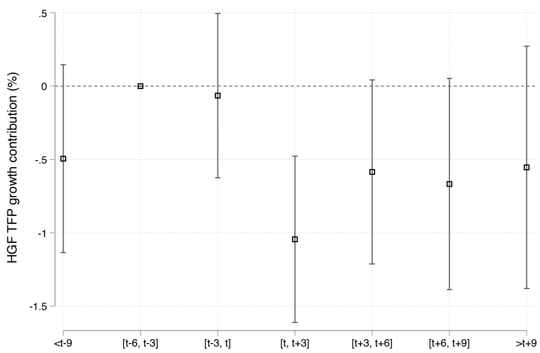
Figure 6: Total TFP contribution: event study



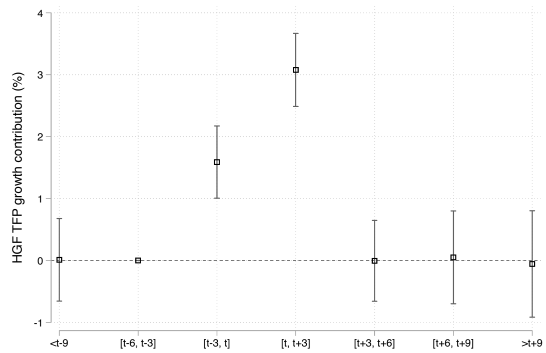

(a) OECD (emp)




(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 which 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 for the period $[t-6, t-3]$.

## 4.3 Industry dynamics and HGF contribution

Motivated by the substantial heterogeneity across industries and time periods documented by Figure 5, this subsection investigates some of the industry characteristics which may influence the role HGFs play in productivity growth.[32] Given the dynamic nature of both HGFs and productivity growth, we focus on basic industry dynamics. The decomposition in Section 2.2 suggests a number of key correlations. The first of these correlations is the *strength or efficiency of reallocation,* $\rho^{real}$. 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 the extent to which HGFs are likely to come from more or less

[32] 2-digit industry fixed effects explain 10-11% of the variance in HGF contributions for the OECD definitions and 18% for the Birch definitions.

productive firms. This 'selection effect' affects 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*, $\rho^{cross}$. This correlation captures a key property of industry dynamics: namely, whether firms improving their productivity are able to acquire sufficient resources rapidly enough to also expand on the market in the same time 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 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 $\rho^{real}_{jt} = corr(prod_{it}, \theta_{i,t+3} - \theta_{i,t})$. Similarly, the cross-correlation is $\rho^{cross}_{jt} = corr(prod_{it+3} - prod_{it}, \theta_{i,t+3} - \theta_{i,t})$. 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: 3-year size growth is negatively correlated with productivity growth during the same period.

These industry-level parameters are associated with other important industry characteristics. Table 6 shows the results from regressions with the two dynamic parameters as dependent variables, and sector, concentration and productivity dispersion as explanatory variables.[33] Starting with the reallocation parameter, we find that it is larger in manufacturing (the omitted category) and in primary industries compared to services. In services, expansion may be less related to productivity relative to industry. Reallocation is also stronger in more concentrated industries, and in industries with a larger productivity dispersion where there is more potential for reallocation. In contrast, the cross correlation is smaller in industries with more dispersion, showing that size growth in such industries is often not accompanied by productivity growth.

[33] The regression equations take the form: $\rho_{jt} = \beta X_{jt} + \eta_t + \epsilon_{it}$, where $j$ indexes 2-digit industries and $t$ years. $\rho_{jt}$ is either the reallocation efficiency or the cross-correlation, and $\eta_t$ are year dummies. $X_{jt}$ includes the sector, the combined market share of the five largest firms (C5) and the standard deviation of the TFP distribution.

Table 6: Determinants of industry dynamics parameters

| VARIABLES | (1) $\rho^{real}$ | (2) $\rho^{cross}$ |
|---|---|---|
| primary | -0.00683 | -0.0404*** |
| | (0.0125) | (0.0120) |
| service | -0.0471*** | -0.00364 |
| | (0.00338) | (0.00287) |
| C5 | 0.0463*** | -0.00803 |
| | (0.00878) | (0.00705) |
| s.d of TFP | 0.0227* | -0.0727*** |
| | (0.0124) | (0.0112) |
| Observations | 3,394 | 3,394 |
| R-squared | 0.149 | 0.051 |
| Year FE: | YES | YES |

These regressions show how the $\rho_{real}$ and $\rho_{cross}$ are related to sector (primary, service and manufacturing, with manufacturing as the base category), concentration (C5 is the combined market share of the 5 largest firms) and the standard deviation of TFP. One observation is a 2-digit industry-year combination, the regression is employment-weighted and the standard errors are robust.

The key question in this section is whether the industry dynamics parameters matter for the 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 $\rho^{real}$ and $\rho^{cross}$ distributions while controlling for year fixed effects ($\delta_t$).[34] Therefore, we estimate the following equation:

$$contribution_{jc} = \sum_q \beta_q^{real} (\rho^{real})^q_{jt_0^c} + \sum_q \beta_q^{cross} (\rho^{cross})^q_{jt_0^c} + \delta_t + \epsilon_{jc} \qquad (2)$$

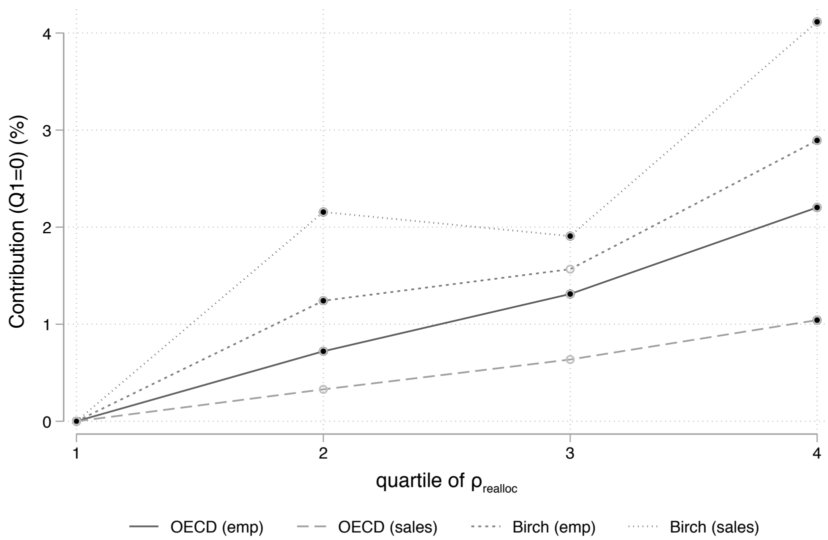
We weight the regression by industry employment and calculate standard errors clustered at the industry level.[35]

The results are presented in Figure 7 and Panel A of Table A2. We find that HGF contribution increases in these industry dynamics parameters for all of the four definitions. Clearly, stronger industry dynamics both in terms of reallocation and cross correlation are associated with higher TFP contributions.

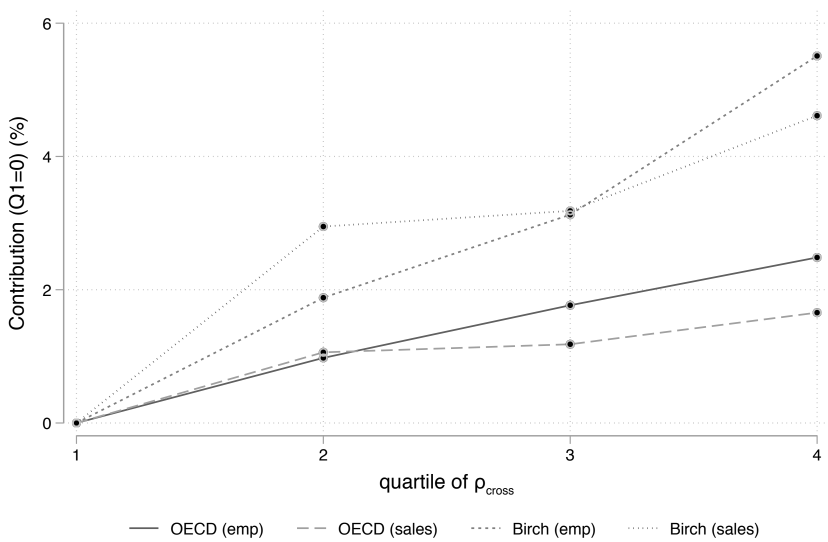
[34] We calculate these based on the full distribution of the industry-year level observations to capture the "absolute" value of the correlations rather than their place in the given year.

[35] These results are robust to a number of 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.

Figure 7: The relationship between industry dynamics parameters and HGF contribution




(a) $\rho^{realloc}$




(b) $\rho^{cross}$

This figure shows point estimates from Equation 3, 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 $\rho^{real}$ and $\rho^{cross}$. 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 $\rho^{cross}$. 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 A2, Panel A. Year dummies are included and standard errors are clustered at the industry level.

The difference in terms of contribution between the industries with the weakest and the strongest reallocation is between 1 and 4 percentage points, depending on the definition. This is quite significant given that the average contribution is between 0.2-4.1 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 a 1-5 pp difference between industries with high vs. low cross correlations.

One possible question related to this exercise is the extent to which the coefficients in Equation 3 are identified from within-industry fluctuations in the $\rho$-s compared to time-invariant industry characteristics, such as the ones presented in Table 6. We approach this question by re-estimating Equation 3 but replacing the $\rho$-s with their industry means:

$$contribution_{jc} = \sum_q \beta_q^{real} (\overline{\rho^{real}})_j^q + \sum_q \beta_q^{cross} (\overline{\rho^{cross}})_j^q + \delta_t + \epsilon_{jc}, \tag{3}$$

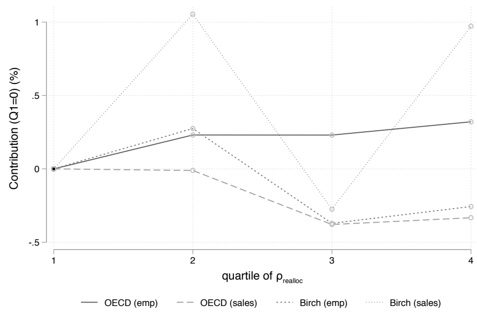
where $\overline{\rho^{real}}$ and $\overline{\rho^{cross}}$ are just the time average of the corresponding $\rho$-s across the years in industry $j$.

Panel B of Table A2 shows the results of this regression. While the coefficients are less significant and somewhat smaller in magnitude, the pattern of the estimates is similar to those in Table 7, especially for the employment-based definitions. The difference between industries with high and low reallocation is 1.5-2 pp while the difference between industries with low and high cross correlation is 2.7-6 pp according to these definitions. The differences are smaller and less significant, but still suggestive for sales-based definitions. The finding that long-term industry characteristics matter for HGF contributions suggests that policies promoting stronger and more efficient industry dynamics can effectively amplify the productivity contribution of HGFs.

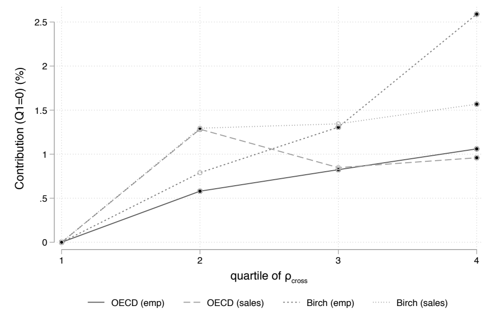
Panel C of Table A2 presents estimation results of Regression 3 using labor productivity to define the dependent variable. We find patterns similar to TFP, suggesting that these patterns are not sensitive to the definition of productivity.

Figure 8 decomposes the relationship found in Figure 7 by re-running Regression 3 with HGF within, between and cross contributions as dependent variables. In line with the arguments stated at the beginning of this subsection, we find that $\rho^{real}$ mainly affects the between component by making more productive firms more likely to become HGFs. In contrast, $\rho^{cross}$ is related to the within and cross components.

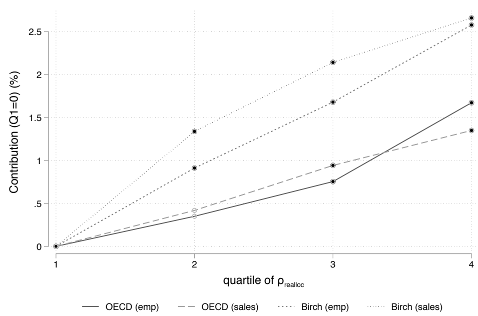
Figure 8: The relationship between industry dynamics parameters and the decomposition of the HGF contribution



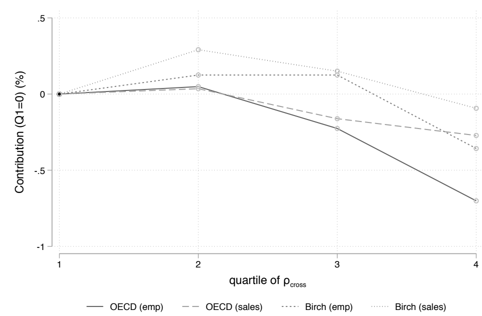

(a) $\rho^{real}$, within



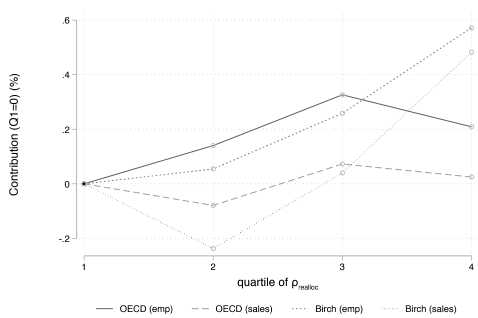

(b) $\rho^{cross}$, within



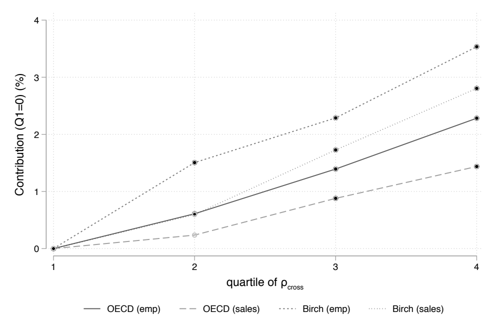

(c) $\rho^{real}$, between




(d) $\rho^{cross}$, between




(e) $\rho^{real}$, cross




(f) $\rho^{cross}$, cross

These figures repeat the regressions in Figure 7 with the different components of the HGF contribution as dependent variables: the HGF within

# 5 Modeling HGF contributions

In this section we present a simple model to explain and illustrate our findings qualitatively. The key idea is to start from a simple (normal) joint distribution of productivity, size (input), productivity growth and size (input) growth. This distribution is easy to characterize 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 easily predicted by this framework. An important consequence of this is that the average HGF contribution is largely determined by overall industry dynamics rather than by idiosyncratic dynamics of specific firms. Second, our framework supplies policy-makers with a few intuitive and measurable concepts and parameters affecting the HGF contribution, 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, this framework can be applied to thinking about the productivity growth contributions of other groups of firms defined by some easily observable characteristics.

## 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[36] 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

[36] In reality, these variables are not normally distributed. Both log productivity and productivity change has fat-tailed distributions (Kang 2017). Size distribution is highly skewed with a very large number of small firms (De Wit 2005). Our data rejects null-hypothesis in tests for both univariate and joint normality (Doornik & Hansen 2008). However, we still choose the normal distribution for two reasons. First, its simple correlation structure allows us to have parameters which all have straightforward economic meaning even though we have four underlying variables. Second, the aim of the analysis is 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.

growth) and the cross correlation (the correlation between input growth and productivity growth). As a baseline, we calibrate these correlations from the Hungarian data.[37] To keep things as simple as possible, we will think of productivity as simply output over input. Therefore, $output = input \times productivity$.[38] As a result, output growth can be written as: $\Delta output = \Delta productivity + \Delta input$, where $\Delta$ is interpreted as a (log) percentage change.

We calculate aggregate productivity growth on these simulated data and decompose it with the method described in Subsection 2.2. We define HGFs based on both input and output growth, as this has proved to be the key distinction in Subsection 3.1.[39] 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% percent in terms of their $\Delta productivity + \Delta input$ value. We simulate 20,000 firms each time.

## 5.2 Input and output HGFs

In this subsection 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 saw 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 which 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

[37] 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.

[38] 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.

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

described in the previous subsection. Let us start with reallocation. Figure 9 depicts the share of HGFs within 20 productivity quantiles using parameters partly calibrated from the Hungarian data.[40] The share of *input-based* HGFs is positively associated with productivity, mostly thanks 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.[41] 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 3 for OECD HGFs[42]: output HGFs tend to be less productive initially than input HGFs – therefore the within contribution of the former is likely to be smaller.

Figure 10 illustrates the within contributions of the simulated input- and output 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 green and red points.

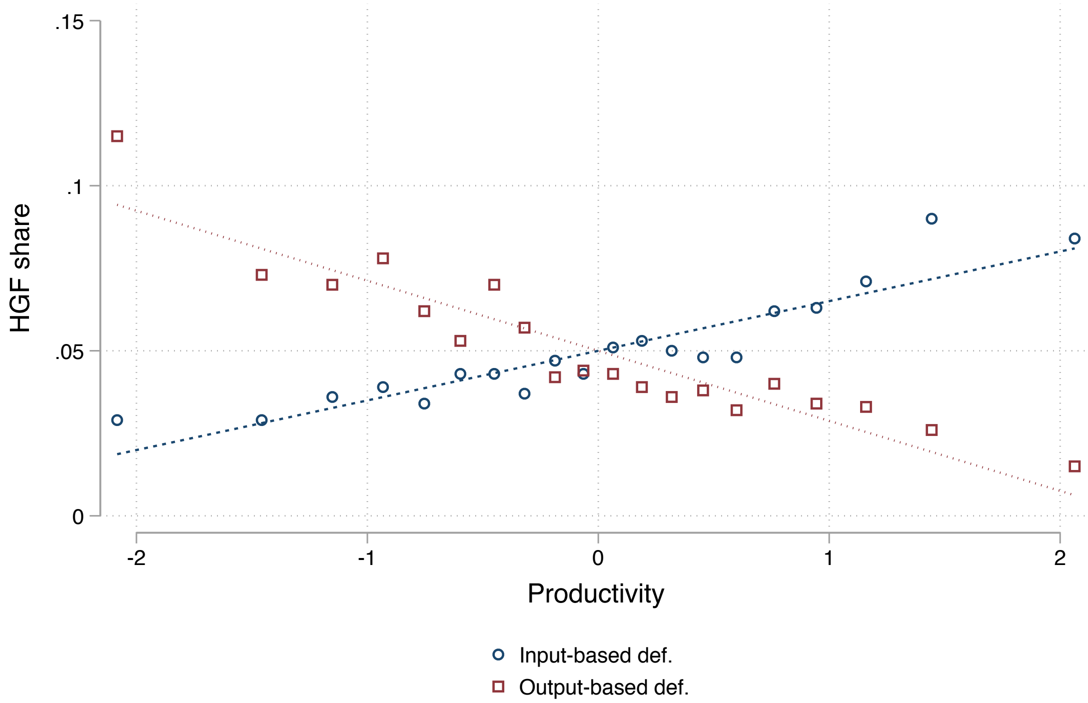
As we have already mentioned, output growth can be expressed in the model as: $\Delta output = \Delta productivity \cdot \Delta input$. Consequently, output HGFs are firms above a line where $\Delta productivity \cdot \Delta input > c$,

[40] 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: $\rho_{real} = 0.1441$; $\rho_{cross} = -0.1$; $corr(\Delta prod, prod) = -0.41$; $corr(prod, inp) = 0.1418$; $corr(\Delta inp, inp) = -0.1852$; $corr(inp, \Delta prod) = 0.0011$.

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

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

Figure 9: 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: $\rho_{real} = 0.1441$; $\rho_{cross} = -0.1$; $\rho(\Delta prod, prod) = -0.41$; $corr(prod, inp) = 0.1418$; $corr(\Delta inp, inp) = -0.1852$; $corr(inp, \Delta prod) = 0.0011$

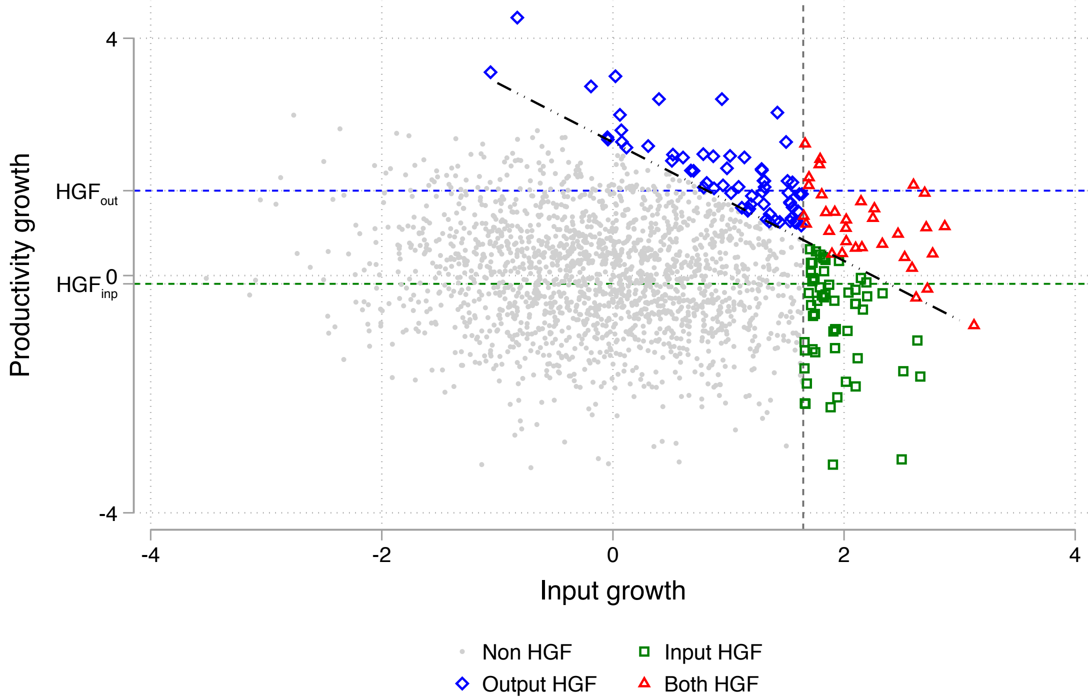
with $c$ being a critical value. Such firms are located above a specific line with a negative slope and are denoted by blue or red. Firms which are both input and output HGFs are denoted by red.

The figure illustrates some of our earlier points. First, the intersection between input and output HGFs is not especially large (as seen in Table 2). 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 3.

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

Figure 10: Input- and output-based HGFs: within effect

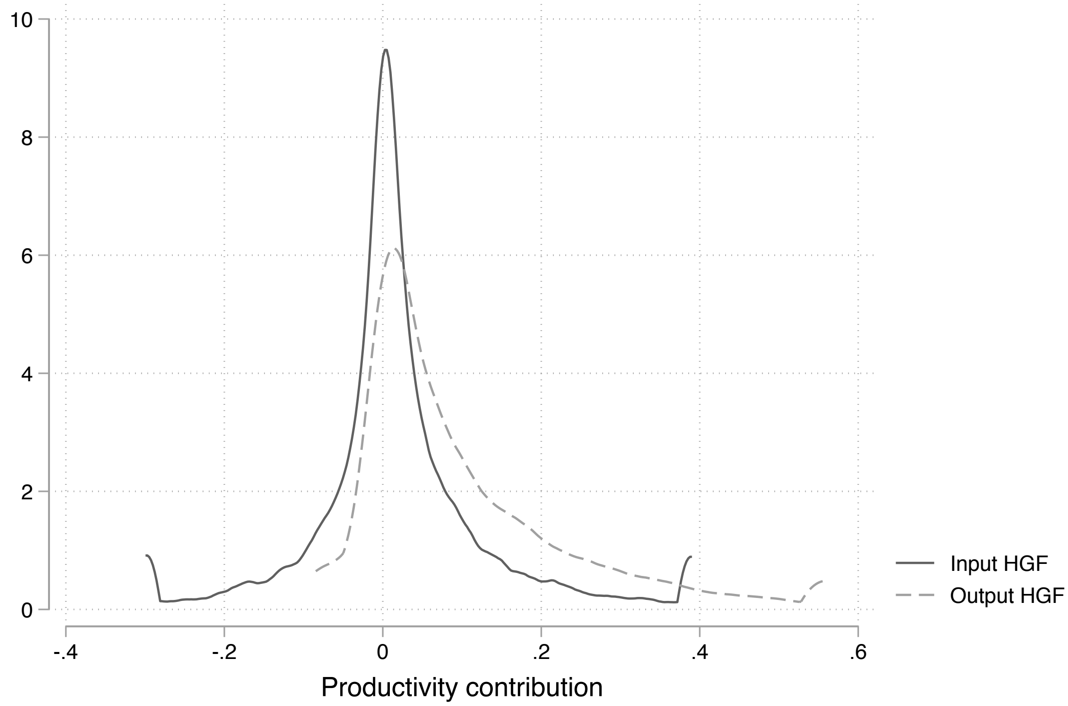



This figure shows the scatter plot of input and productivity growth of firms in the simulated data. Input HGFs are those with input growth above a threshold, denoted by green and red. Output HGFs are the blue and red firms, and the red firms are both input and output HGFs. $HGF_{out}$ represents the 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 are normalized to zero.

of their within contributions. Now we continue with simulating the distribution of the total contributions (similar to Figure 5). In that 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 11 shows this distribution where industry correlation parameters ($\rho_{real}, \rho_{cross}, corr(\Delta prod, prod), corr(prod, inp), corr(\Delta inp, inp), corr(inp, \Delta prod)$) are chosen randomly from a uniform distribution.[43] This figure reproduces the main patterns of Figure 5: first, the contribution of output HGFs clearly stochastically dominates that of input-based HGFs; second, the distribution corresponding to output HGFs is more asymmetric with a long right tail but very few negative contributions.

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

Figure 11: 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 ($\rho_{real}$, $\rho_{cross}$, $corr(\Delta prod, prod)$, $corr(prod, inp)$, $corr(\Delta inp, inp)$, $corr(inp, \Delta prod)$ 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.

## 5.3 Industry Dynamics

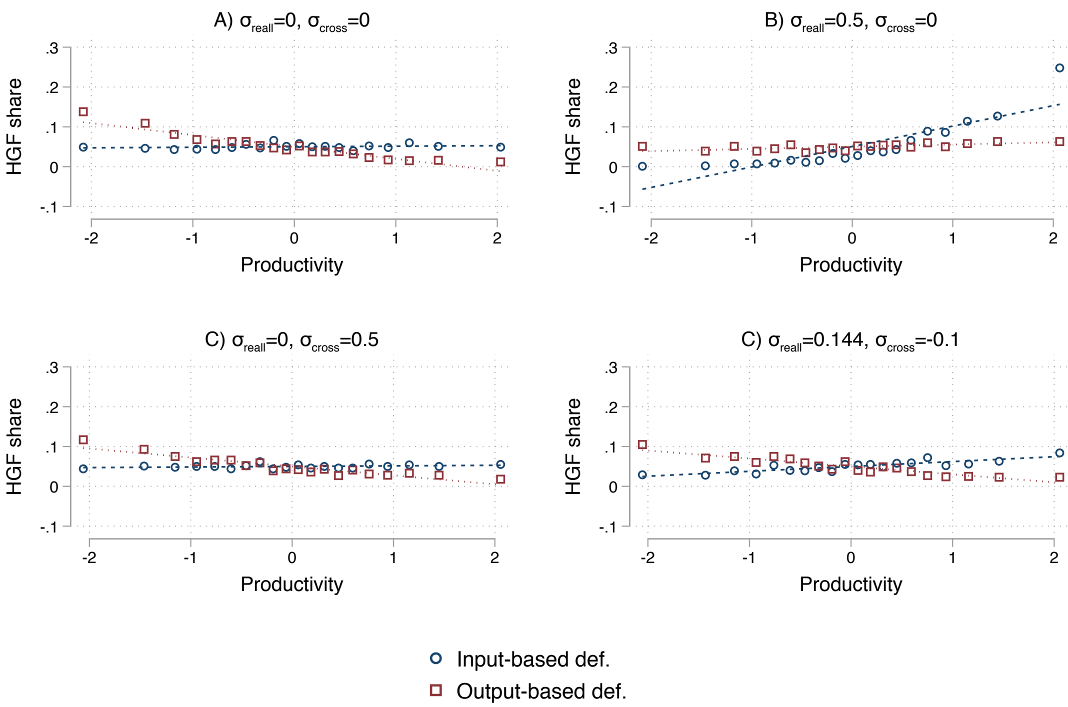
In this subsection we focus on the relationship between industry dynamics parameters and productivity contributions. We follow similar steps (and graphical tools) to that of the previous subsection, starting with the between component, followed by the within component and presenting regressions results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

Let us start with how the parameters affect the between term. Figure 12 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.[44] 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

[44]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.

stronger ($\rho_{realloc} = corr(productivity, \Delta size)$, the slope of the two curves increases. This 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 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 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 Figure 9.

Figure 12: Parameter values and reallocation

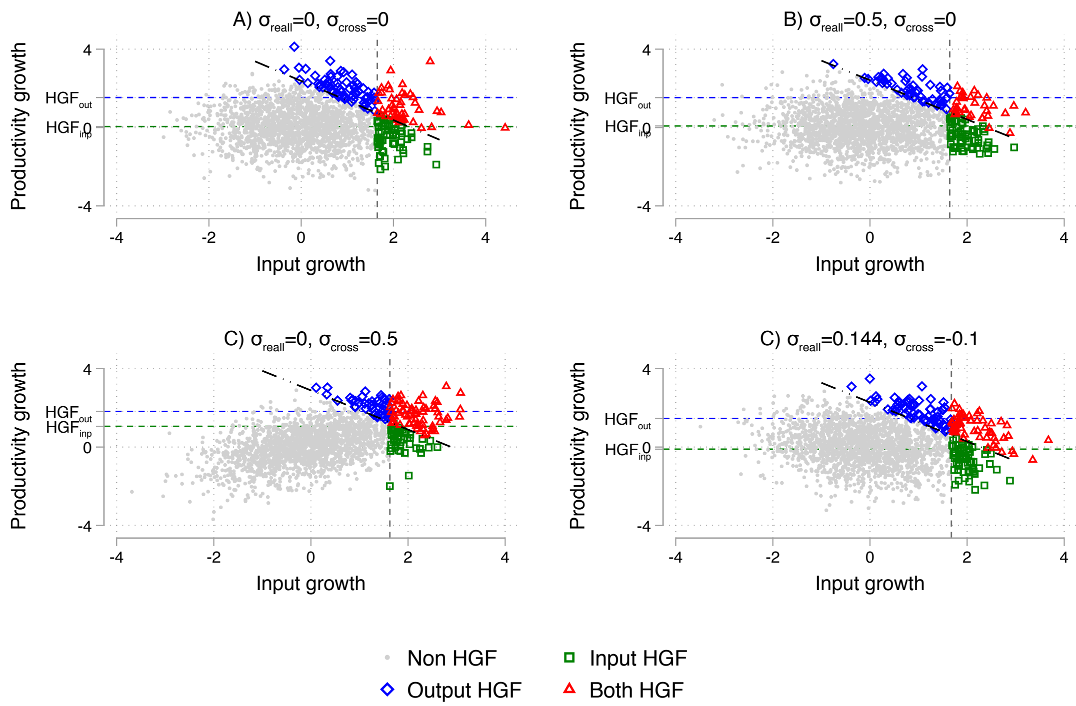



This figure shows the simulated relationship between initial productivity level and the share of input- and output-based HGFs for different values of the key moments of industry dynamics. We normalize each of the variables in such a way that their standard deviation is 1. The remaining correlations are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

Figure 13 repeats this exercise for the within contribution, following the structure of Figure 10. 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 of the four panels we see a similar pattern: the input HGF contribution being 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 before. These simulations predict that a higher cross correlation is positively associated with the within contribution of input HGFs (and of output HGFs, though to a smaller extent).

By definition, we also expect that higher cross correlation also yields larger cross contribution for both definitions.

Figure 13: Parameter values and reallocation




This figure shows the scatter plot of input growth and productivity growth of firms in the simulated data for different key moments of industry dynamics. Input HGFs are those with input growth above a threshold, denoted by green and red. Output HGFs are the blue and red firms, and the red firms 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 remaining correlations are calibrated from the Hungarian data: $corr(\Delta prod, prod) = -0.41$, $corr(prod, inp) = 0.1418$; $corr(\Delta inp, inp) = -0.1852$; $corr(inp, \Delta prod) = 0.0011$

This simple simulation produces a number of patterns, in line with our empirical findings, presented in Table 7.

Table 8 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.

Table 7: Empirical patterns captured by the model

| | Input based | | | |
|---|---|---|---|---|
| | within | between | cross | overall |
| $\rho_{real}$ | 0 | + + | 0 | + + |
| $\rho_{cross}$ | + + | 0 | + + | + + |

| | Output based | | | |
|---|---|---|---|---|
| | within | between | cross | overall |
| $\rho_{real}$ | 0 | + | 0 | + |
| $\rho_{cross}$ | + | 0 | + | + |

This table shows the main predictions from the simulation exercise. The predictions about the within effect come from Figure 12. Comparing panels A and B of that figure shows that an increase in $\sigma_{real}$ 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 Figure 9. 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.

We run regressions of the form:

$$contribution_i = \beta_0 + \beta_{real}\rho_i^{real} + \beta_{cross}\rho_i^{cross} + \epsilon_i \tag{4}$$

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 Figures 7 and 8. Stronger reallocation increases the contribution for both types of HGFs via the between component. Larger cross correlation also improves the contribution of both types of HGFs both via the within and cross terms.

## 5.4 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 capture the productivity contributions of other groups of firms of 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.

Table 8: 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) |
| $\rho_{cross}$ | 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) |
| $\rho_{cross}$ | 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 4 on the simulated data, where the different components of the total HGF contribution is the dependent variable and $\rho^{real}$ and $\rho^{cross}$ are the explanatory variables. In 10,000 rounds of simulations, the correlations ($\rho_{real}$, $\rho_{cross}$, $corr(\Delta prod, prod)$, $corr(prod, inp)$, $corr(\Delta inp, inp)$, $corr(inp, \Delta prod)$ 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.

SMEs can be defined as firms with a size below a certain threshold of $inp$ in year $t$ rather than $\Delta inp$, 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 compared to 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 a number of 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 & 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 data. 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.[45]

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.[46] 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 important

[45] 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 $\Delta inp$, 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 ($\Delta inp$) in this correlation.

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

in determining the within term, and the reallocation correlation also affects the type of firms selected, as shown on Figure 9.

## 6 Conclusion

This paper investigates the role of HGFs in reallocation and industry-level productivity growth. First, HGF contributions vary widely: the contribution is negative in a quarter of the cases and HGFs contribute more than 50% in a substantial number of cases. These results show that the HGFs' effect on industry productivity can be substantial, either positively or negatively. Policy-makers should not ignore this when designing or evaluating policies promoting high-growth firms.

Second, the type of HGF definition matters. We consider two dimensions: whether only relative growth is considered and whether the definition is based on input or output growth. We find that firms with higher output growth are likely to contribute more to aggregate productivity. We identify two trade-offs. First, as long as output HGFs generate fewer jobs than input HGFs, there is a trade-off between job creation and productivity contribution by HGFs: if job creation is maximized, productivity contribution may be small. This trade-off, however, does not seem very stark empirically: while the productivity contribution of sales HGFs is much larger, they also create a large number of jobs. Second, we find that output HGFs, which, on average, contribute much more than input HGFs, are typically less productive initially than input HGFs. Accordingly, output HGFs predominantly contribute through their within growth while input HGFs more via reallocation. HGF policies aiming (implicitly) at a high reallocation contribution may require strong initial productivity from participants. However, such a design may crowd out potential output HGFs, which may contribute substantially via their within productivity growth.

Finally, we find that key parameters of industry dynamics are correlated with HGF productivity contribution. Stronger reallocation, convergence and correlation between size and productivity growth are positively related to productivity contribution. These parameters depend on framework conditions. This suggests that improving industry dynamics can help in maximizing the impact of HGFs. Policies improving framework conditions may complement strongly specific HGF policies.

We use a simple model to explain these patterns. This is based on a simulation of a joint normal distribution of input level, productivity level, input growth and productivity growth. The correlations between these factors are key industry dynamics parameters. Importantly, this simplified framework is able to reproduce our empirical observations qualitatively. This suggests that overall correlations of the industry-level distribution, rather than some specifics of HGFs, are able to explain overall HGF contributions. This exercise may also help policymakers by supplying intuitive and measurable objects (input vs. output HGF, reallocation and cross correlations) when designing their policies. Finally, similar frameworks can be applied when considering the contribution of other groups of firms defined based on similar dimensions, as we demonstrate via the example of SMEs.

# References

Ackerberg, D. A., Caves, K. & Frazer, G. (2015), 'Identification properties of recent production function estimators', Econometrica **83**(6), 2411–2451.

Acs, Z. J., Parsons, W. & Tracy, S. (2008), 'High-impact firms: Gazelles revisited', Washington, DC pp. 1–82.

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

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

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.

Balsvik, R. & Haller, S. A. (2006), 'The contribution of foreign entrants to employment and productivity growth'.

Baqaee, D. R. & Farhi, E. (2020), 'Productivity and misallocation in general equilibrium', The Quarterly Journal of Economics **135**(1), 105–163.

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.

Bianchini, S., Bottazzi, G. & Tamagni, F. (2017), 'What does (not) characterize persistent corporate high-growth?', Small Business Economics **48**(3), 633–656.

Birch, D. L. (1981), 'Who creates jobs?', The Public Interest **65**, 3.

Choi, T., Rupasingha, A., Robertson, J. C. & Leigh, N. G. (2017), 'The effects of high growth on new business survival', Review of Regional Studies **47**(1), 1–23.

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

Coad, A. & Srhoj, S. (2019), ‘Catching gazelles with a lasso: Big data techniques for the prediction of high-growth firms’, Small Business Economics pp. 1–25.

Daunfeldt, S.-O., 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.

Daunfeldt, S.-O. & Halvarsson, D. (2015), ‘Are high-growth firms one-hit wonders? evidence from sweden’, Small Business Economics **44**(2), 361–383.

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.

de Nicola, F., Muraközy, B. & Tan, S. W. (2019), ‘Spillovers from high growth firms: Evidence from hungary’, Small Business Economics pp. 1–24.

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.

Delmar, F. (2006), ‘Measuring growth: methodological considerations and empirical results’, Entrepreneurship and the Growth of Firms **1**(1), 62–84.

Delmar, F., Davidsson, P. & Gartner, W. B. (2003), ‘Arriving at the high-growth firm’, Journal of Business Venturing **18**(2), 189–216.

Delmar, F., McKelvie, A. & Wennberg, K. (2013), ‘Untangling the relationships among growth, profitability and survival in new firms’, Technovation **33**(8-9), 276–291.

Doornik, J. A. & Hansen, H. (2008), ‘An omnibus test for univariate and multivariate normality’, Oxford Bulletin of Economics and Statistics **70**, 927–939.

Du, J. & Temouri, Y. (2015), ‘High-growth firms and productivity: Evidence from the united kingdom’, Small Business Economics **44**(1), 123–143.

Foster, L., Haltiwanger, J. & Syverson, C. (2008), 'Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?', American Economic Review **98**(1), 394–425.

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

Gjerløv-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.

Gleeson, A. M. & 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.

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

Haltiwanger, J., Jarmin, R. S., Kulick, R. B. & 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 .

Harris, R. & Moffat, J. (2013), 'The direct contribution of fdi to productivity growth in britain, 1997–2008', The World Economy **36**(6), 713–736.

Henrekson, M. & Johansson, D. (2010), 'Gazelles as job creators: a survey and interpretation of the evidence', Small business economics **35**(2), 227–244.

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.

Hsieh, C.-T. & Klenow, P. J. (2009), 'Misallocation and manufacturing tfp in china and india', The Quarterly Journal of Economics **124**(4), 1403–1448.

Kang, K. (2017), 'Identifying the dispersion of the pareto productivity distribution in international trade', Economics Letters **161**, 108–111.

Leung, D., Meh, C., Terajima, Y. et al. (2008), 'Productivity in canada: Does firm size matter?', Bank of Canada Review **2008**(Autumn), 7–16.

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.

Mason, C. & Brown, R. (2013), 'Creating good public policy to support high-growth firms', Small Business Economics **40**(2), 211–225.

Medrano-Adán, L., Salas-Fumás, V. & Sanchez-Asin, J. J. (2019), 'Firm size and productivity from occupational choices', Small Business Economics **53**(1), 243–267.

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.

Moschella, D., Tamagni, F. & Yu, X. (2019), 'Persistent high-growth firms in china's manufacturing', Small Business Economics **52**(3), 573–594.

OECD (2010), 'What governments can do to make a difference'.

Olley, G. S. & Pakes, A. (1992), The dynamics of productivity in the telecommunications equipment industry, Technical report, National Bureau of Economic Research.

Restuccia, D. & Rogerson, R. (2017), 'The causes and costs of misallocation', Journal of Economic Perspectives **31**(3), 151–74.

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.

Stoyanov, A. & Zubanov, N. (2012), 'Productivity spillovers across firms through worker mobility', American Economic Journal: Applied Economics **4**(2), 168–98.

# Appendix

Table A1: Employment and labor productivity before and after the high growth phase, by definition

| | oecd3 (emp) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | employment | | | | labor productivity | | | |
| year | t-4 | t-1 | t+4 | t+7 | t-4 | t-1 | t+4 | t+7 |
| 2000 | | 61.1% | 163.8% | 171.6% | | 127.1% | 102.1% | 105.9% |
| 2003 | 69.3% | 69.7% | 194.9% | 194.3% | 116.1% | 121.9% | 105.4% | 112.7% |
| 2006 | 59.3% | 59.8% | 185.2% | 179.0% | 106.5% | 115.8% | 100.9% | 103.9% |
| 2009 | 68.3% | 73.4% | 185.4% | | 98.4% | 110.1% | 102.4% | |
| 2011 | 67.6% | 67.6% | | | 105.3% | 111.3% | | |
| average | 66.1% | 66.3% | 182.3% | 181.6% | 106.6% | 117.2% | 102.7% | 107.5% |
| | oecd3 (sales) | | | | | | | |
| | employment | | | | labor productivity | | | |
| year | t-4 | t-1 | t+4 | t+7 | t-4 | t-1 | t+4 | t+7 |
| 2000 | | 74.0% | 120.1% | 130.4% | | 93.7% | 110.5% | 109.3% |
| 2003 | 75.7% | 79.0% | 138.7% | 146.7% | 96.1% | 99.2% | 118.3% | 120.7% |
| 2006 | 69.5% | 70.7% | 141.5% | 142.3% | 89.6% | 86.9% | 109.6% | 107.0% |
| 2009 | 78.7% | 80.2% | 138.8% | | 100.4% | 102.7% | 122.9% | |
| 2011 | 73.1% | 75.2% | | | 91.1% | 92.1% | | |
| average | 74.2% | 75.8% | 134.8% | 139.8% | 94.3% | 94.9% | 115.3% | 112.3% |
| | birch3 (emp) | | | | | | | |
| | employment | | | | labor productivity | | | |
| year | t-4 | t-1 | t+4 | t+7 | t-4 | t-1 | t+4 | t+7 |
| 2000 | | 237.3% | 581.1% | 631.2% | | 151.6% | 119.9% | 122.7% |
| 2003 | 258.7% | 268.5% | 645.7% | 703.8% | 143.2% | 144.1% | 124.2% | 130.7% |
| 2006 | 253.5% | 238.5% | 582.1% | 576.0% | 133.2% | 146.8% | 124.9% | 126.8% |
| 2009 | 279.1% | 290.3% | 607.1% | | 132.0% | 141.3% | 126.4% | |
| 2011 | 275.9% | 292.8% | | | 126.1% | 136.1% | | |
| average | 266.8% | 265.5% | 604.0% | 637.0% | 133.6% | 144.0% | 123.8% | 126.7% |
| | birch3 (sales) | | | | | | | |
| | employment | | | | labor productivity | | | |
| year | t-4 | t-1 | t+4 | t+7 | t-4 | t-1 | t+4 | t+7 |
| 2000 | | 398.0% | 601.8% | 667.2% | | 185.9% | 200.3% | 183.8% |
| 2003 | 323.3% | 325.8% | 562.2% | 656.6% | 176.5% | 168.3% | 186.0% | 183.5% |
| 2006 | 200.0% | 206.6% | 401.9% | 447.9% | 139.6% | 147.5% | 168.3% | 161.6% |
| 2009 | 225.9% | 247.5% | 405.4% | | 151.4% | 154.8% | 179.0% | |
| 2011 | 304.5% | 343.6% | | | 157.1% | 168.6% | | |
| average | 263.4% | 304.3% | 492.8% | 590.5% | 156.1% | 165.0% | 183.4% | 176.3% |

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

Table A2: Robustness checks

Panel A Base

| HGF type | (1) birch3 (emp) | (2) birch3 (sales) | (3) oecd (emp) | (4) oecd (sales) |
|---|---|---|---|---|
| $\rho^{real}(Q2)$ | 1.242** | 2.155** | 0.721** | 0.329 |
| | (0.616) | (0.839) | (0.353) | (0.514) |
| $\rho^{real}(Q3)$ | 1.567 | 1.908** | 1.311** | 0.636 |
| | (0.961) | (0.911) | (0.634) | (0.469) |
| $\rho^{real}(Q4)$ | 2.894** | 4.115*** | 2.203*** | 1.042** |
| | (1.270) | (1.052) | (0.681) | (0.510) |
| $\rho^{cross}(Q2)$ | 1.881 | 2.948*** | 0.977 | 1.061 |
| | (1.237) | (0.906) | (0.631) | (0.636) |
| $\rho^{cross}(Q3)$ | 3.127** | 3.183*** | 1.767** | 1.182 |
| | (1.312) | (1.023) | (0.742) | (0.756) |
| $\rho^{cross}(Q4)$ | 5.508*** | 4.612*** | 2.484*** | 1.657** |
| | (1.823) | (1.405) | (0.848) | (0.789) |
| Observations | 723 | 723 | 723 | 723 |
| R-squared | 0.176 | 0.151 | 0.157 | 0.122 |

Panel B Mean rhos

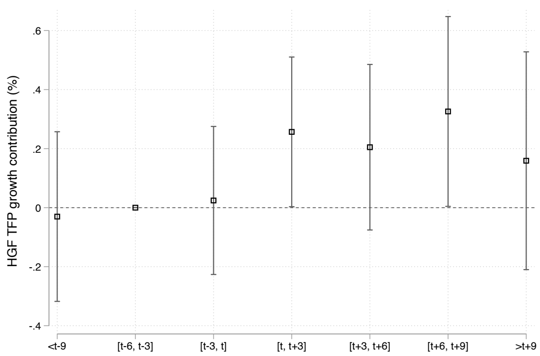
| HGF type | (1) birch3 (emp) | (2) birch3 (sales) | (3) oecd (emp) | (4) oecd (sales) |
|---|---|---|---|---|
| $\rho^{real}(Q2)$ | 0.546 | 0.987 | 0.959** | 0.758 |
| | (0.839) | (0.925) | (0.403) | (0.630) |
| $\rho^{real}(Q3)$ | 1.199 | 0.951 | 1.121** | 1.005 |
| | (1.348) | (1.219) | (0.429) | (0.922) |
| $\rho^{real}(Q4)$ | 1.953* | 1.434 | 1.509*** | 0.974 |
| | (1.034) | (1.038) | (0.504) | (0.693) |
| $\rho^{cross}(Q2)$ | 3.377** | 3.752*** | 1.945** | 1.083* |
| | (1.361) | (1.317) | (0.739) | (0.596) |
| $\rho^{cross}(Q3)$ | 3.180** | 2.782** | 2.279*** | 0.808* |
| | (1.277) | (1.340) | (0.675) | (0.441) |
| $\rho^{cross}(Q4)$ | 6.024*** | 4.916*** | 2.737*** | 1.457 |
| | (1.836) | (1.480) | (0.717) | (1.120) |
| Observations | 726 | 726 | 726 | 726 |
| R-squared | 0.191 | 0.155 | 0.167 | 0.127 |

Panel B Base with Labour productivity

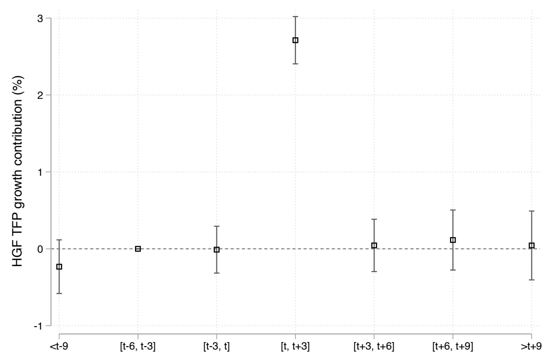
| HGF type | (1) birch3 (emp) | (2) birch3 (sales) | (3) oecd (emp) | (4) oecd (sales) |
|---|---|---|---|---|
| $\rho^{real}(Q2)$ | 1.286 | 1.048 | 0.578* | 0.433 |
| | (0.782) | (0.771) | (0.294) | (0.463) |
| $\rho^{real}(Q3)$ | 3.399*** | 3.843*** | 1.153*** | 1.261* |
| | (0.932) | (0.995) | (0.378) | (0.697) |
| $\rho^{real}(Q4)$ | 5.632*** | 5.343*** | 1.675*** | 2.447*** |
| | (1.220) | (1.223) | (0.479) | (0.751) |
| $\rho^{cross}(Q2)$ | 1.759** | 1.527* | 1.177*** | 1.738*** |
| | (0.804) | (0.845) | (0.438) | (0.615) |
| $\rho^{cross}(Q3)$ | 2.917*** | 2.107** | 1.539*** | 1.357** |
| | (0.952) | (0.887) | (0.461) | (0.563) |
| $\rho^{cross}(Q4)$ | 5.497*** | 4.328*** | 1.987*** | 2.198*** |
| | (0.847) | (0.814) | (0.496) | (0.675) |
| Observations | 725 | 725 | 725 | 725 |
| R-squared | 0.276 | 0.269 | 0.198 | 0.171 |

Panel A presents the regression results corresponding to Figure 7. Panel B re-runs those regressions with replacing the $\rho$s with their industry means. Panel C re-runs the regression in Panel A with HGF contribution to labor productivity as the dependent variable. Year dummies are included and standard errors are clustered at the industry level.

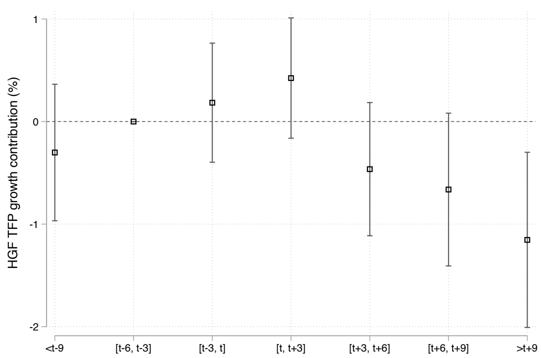
Figure A1: Total labor Productivity contribution: event study



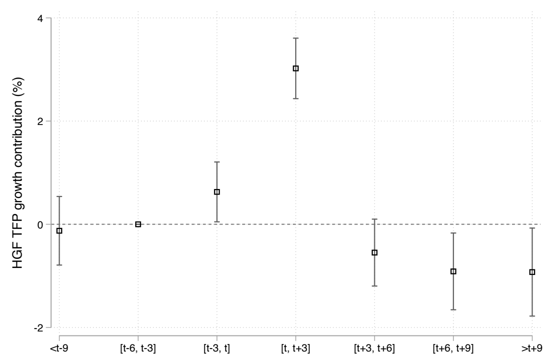

(a) OECD (emp)




(b) OECD (sales)




(c) Birch (emp)




(d) Birch (sales)

Figure A2: Labor productivity contribution of HGFs, by industry-cohort