Regional incidence and persistence of high-growth firms: Testing ideas from the Entrepreneurial Ecosystems literature

Coad, A.

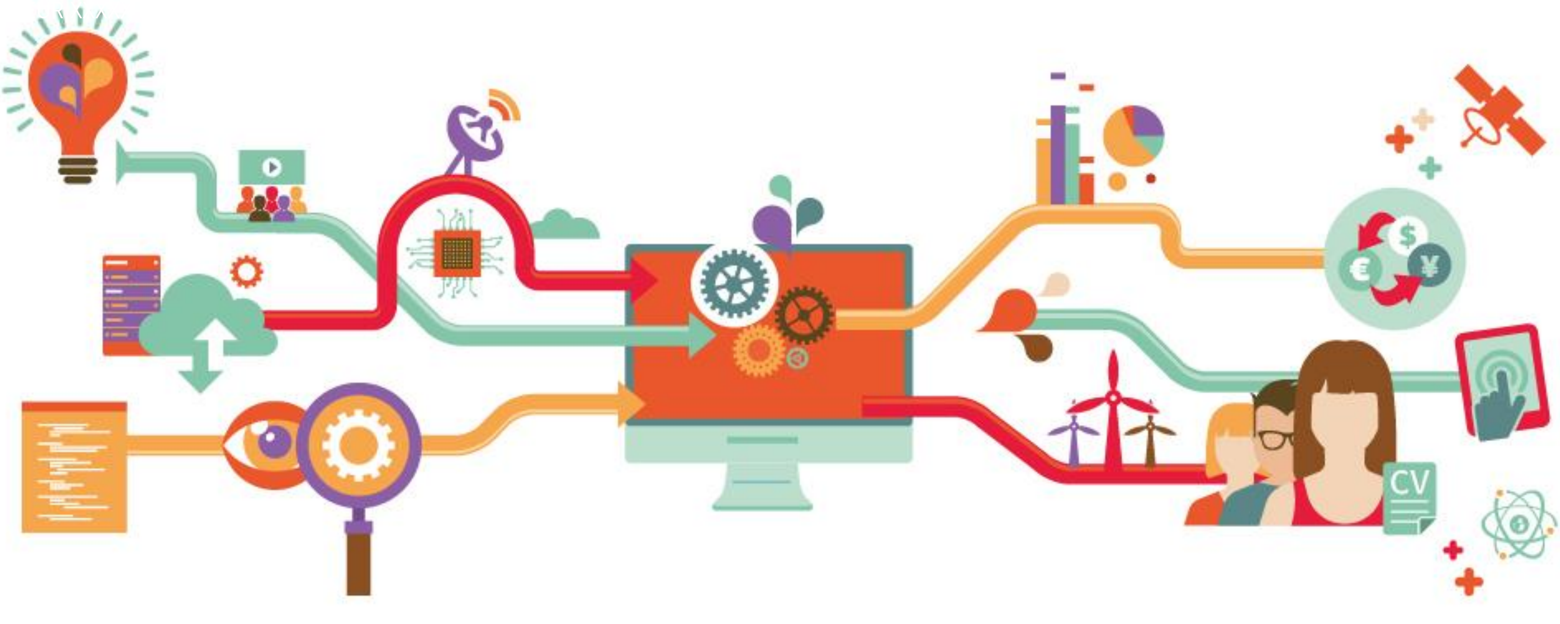
Domnick, C.

Santoleri, P.

Srhoj, S.

JRC Working Papers on Corporate R&D and Innovation

No 02/2023



This document has been produced within the context of the Global Industrial Research & Innovation Analyses (GLORIA) activities that are jointly carried out by the European Commission's Joint Research Centre – Directorate Fair and Sustainable Economy and the Directorate General for Research and Innovation-Directorate F, Prosperity.

Any comments can be sent by email to: JRC-B6-secretariat@ec.europa.eu, More information, including activities and publications, is available at: https://iri.jrc.ec.europa.eu/home/.

Contact information
Pietro Moncada-Paternò-Castello
Address: Edificio Expo. c/ Inca Garcilaso, 3. E-41092 Seville (Spain)
E-mail: JRC-B6-secretariat@ec.europa.eu
Tel.: +34 954488388

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC134469

Seville: European Commission, 2023

How to cite this working paper: Coad, A., Domnick, C., Santoleri, P., Srhoj, S. (2023), *Regional incidence and persistence of high-growth firms: Testing ideas from the Entrepreneurial Ecosystems literature,* JRC Working Papers on Corporate R&D and Innovation series No 02/2023, European Commission, Seville, JRC134469.

The **JRC Working Papers on Corporate R&D and Innovation** are published under the editorial supervision of Pietro Moncada-Paternò-Castello, Alexander Tübke and James Gavigan in collaboration with Sofia Amaral-Garcia, Fernando Hervás, Koen Jonkers, Francesco Rentocchini at the European Commission – Joint Research Centre, and in cooperation with Sara Amoroso (German Institute for Economic Research, DEU), Michele Cincera (Solvay Brussels School of Economics and Management, Université Libre de Bruxelles, BEL), Alex Coad (Waseda University, Tokyo, JAP), Enrico Santarelli (University of Bologna, ITA), Daniel Vertesy (International Telecommunication Union, CHE – and UNU-MERIT, NLD), Antonio Vezzani (Roma Tre University, ITA); Marco Vivarelli (Università Cattolica del Sacro Cuore, Milan, ITA) and Zoltan Csefalvay (Mathias Corvinus Collegium, HUN).

Executive Summary

High growth firms (HGFs) play a disproportionate role in contributing to aggregate economic outcomes. This has led policy-makers to implement initiatives to promote HGFs and such schemes are now embedded within national and regional entrepreneurship policies in many countries. Most academic literature in this field has taken advantage of the increased availability of firm-level data to provide evidence from a national perspective. While this offers a valuable characterization at the national-level, it provides little insight into the incidence of HGFs from a regional point of view, including the relationship of these businesses with local economic development. As a result, this has arguably hindered our understanding and the ability to produce useful evidence to inform the design and implementation of regional policy.

Against this backdrop, this paper provides an in-depth empirical analysis of regional shares of HGFs. To that end, we leverage predictions regarding the dynamics of regional HGF shares stemming from the Entrepreneurial Ecosystems (EEs) literature. In line with what is commonly assumed by policy-makers and practitioners, the EE theory suggests that more developed regions have higher regional HGF shares, and that regional HGF shares are relatively persistent over time. We investigate empirically three broad hypotheses based on various operationalizations of these statements, using Eurostat data at the NUTS-3 level for up to 20 countries over the time span 2008-2020. The representative and cross-country nature of these data is particularly suitable to produce an accurate and rich account of regional dynamics of HGFs.

Concerning the incidence of HGFs, our results suggest that the areas with the highest HGF shares – i.e. the Silicon Valleys of Europe with the highest levels of EE outputs – seem to be peripheral regions such as the Canary Islands (Spain), Sicily (Italy) and Algarve (Portugal). This represents a puzzle for the EE theory. Regression analysis further corroborates this result: regions with high levels of income per capita, innovative activity or EE quality are not those with the highest incidence of regional HGFs.

We then investigate the presence of persistence in regional HGF shares. Results suggest that regional persistence of HGF shares is relatively high, which can potentially be reconciled with EE theoretical notions. However, we do not find evidence that i) persistence is stronger in regions with higher economic development, and that ii) regions at the top of the HGF shares distribution in the past are able to remain at the top of the distribution in the future. In other words, while evidence provides support for persistence, the nature of such persistence does not seem to be caused by path-dependence as put forward by the EE literature. Our results call for a more nuanced understanding of the meaning of the HGF shares both as an economic indicator as well as in terms of its operationalization within the current EEs framework.

Regional incidence and persistence of high-growth firms: Testing ideas from the Entrepreneurial Ecosystems literature

Alex Coad, Clemens Domnick, Pietro Santoleri, Stjepan Srhoj[1]

Abstract

Policy-makers and scholars often assume that a higher incidence of high-growth firms (HGFs) is synonymous with vibrant regional economic dynamics, and that HGF shares are persistent over time as Entrepreneurial Ecosystems (EEs) have slowly-changing features. In this paper we test these hypotheses, which are deeply rooted in the EE literature. We draw upon Eurostat data for up to 20 countries over the period 2008-2020 and study HGF shares in NUTS-3 regions in Europe. Analysis of regional rankings yields the puzzling finding that the leading EEs in Europe, apparently, are in places such as southern Spain and southern Italy. These places would not normally be considered Europe's foremost entrepreneurial hotspots. Additional results do not provide strong support for the hypothesis that more developed regions feature higher HGF shares. We do find evidence consistent with HGF shares displaying persistency over time. However, we show that more developed regions do not have higher persistence in their HGF shares, and that the strength in persistence does not increase across the HGFs distribution, which does not support path-dependency as the main mechanism behind the observed persistence. Overall, we call for a more nuanced interpretation of both regional HGF shares and the EEs literature.

Keywords: Entrepreneurial Ecosystems, High-Growth Firms, Persistence, Firm Growth, Entrepreneurship Policy, Regional Policy

[1] Authors listed alphabetically. Alex Coad (alex.coad@waseda.jp), Waseda Business School, Waseda University, Japan.
Clemens Domnick (clemens.domnick@ec.europa.eu), Joint Research Centre, European Commission, Spain.
Pietro Santoleri (pietro.santoleri@ec.europa.eu), Joint Research Centre, European Commission, Spain.
Stjepan Srhoj (ssrhoj@efst.hr), Department of Economics, Faculty of Economics, Business and Tourism, University of Split, Croatia.
Acknowledgements**:** We are grateful to Sara Amoroso, Werner Hölzl, Mike King-fai Fung, Alexander Lembcke, Carlo Menon, Simone Sasso, and participants at the Tokyo SIBR Conference, Joint Research Centre IID Seminar Series, and Metu-Tekpol for useful comments. We thank Sarmite Visocka for her clarifications on the Eurostat data. The opinions expressed herein are those of the authors and do not necessarily reflect those of the European Commission. The usual caveat applies.

1. Introduction

High growth firms (HGFs) play a disproportionate role in contributing to aggregate economic outcomes (Haltiwanger, et al., 2016; Du and Vanino, 2021). This has led policy-makers to implement initiatives to promote HGFs and such schemes are now embedded within national and regional entrepreneurship policies in many countries (OECD, 2010; Brown et al., 2014; Stam and Bosma, 2015; Flachenecker et al., 2020; Coad et al., 2022). Most academic literature in this field has taken advantage of the increased availability of firm-level data to provide evidence from a national perspective (Friesenbichler and Hölzl, 2020). While this offers a valuable characterization at the national-level, it provides little insight into the incidence of HGFs from a regional point of view, including the relationship of these businesses with local economic development (Brown et al., 2014). As a result, this has arguably hindered our understanding and the ability to produce useful evidence to inform the design and implementation of regional policy.

Against this backdrop, this paper provides an in-depth empirical analysis of regional shares of HGFs. To that end, we leverage predictions regarding the dynamics of regional HGF shares stemming from the Entrepreneurial Ecosystems (EEs) literature (Stam, 2015; Spigel, 2017; Leendertse et al., 2021). In line with what is commonly assumed by policy-makers and practitioners, the EE theory suggests that more developed regions have higher regional HGF shares, and that regional HGF shares are relatively persistent over time. We investigate empirically three broad hypotheses based on various operationalizations of these statements. We make use of Eurostat data at the NUTS-3 level for up to 20 countries over the time span 2008-2020. The representative and cross-country nature of these data is particularly suitable to produce an accurate and rich account of regional dynamics of HGFs.

Concerning the incidence of HGFs, our results suggest that the areas with the highest HGF shares – i.e. the Silicon Valleys of Europe with the highest levels of EE outputs – seem to be peripheral regions such as the Canary Islands (Spain), Sicily (Italy) and Algarve (Portugal). This represents a puzzle for the EE theory. Regression analysis further corroborates this result: regions with high levels of income per capita, innovative activity or EE quality are not those with the highest indidence of regional HGFs.

We then investigate the presence of persistence in regional HGF shares. Results suggest that regional persistence of HGF shares is relatively high, which can potentially be reconciled with EE theoretical notions. However, we do not find evidence that i) persistence is stronger in regions with higher economic development, and that ii) regions at the top of the HGF shares distribution in the past are able to remain at the top of the distribution in the future. In other words, while evidence provides support for persistence, the nature of such persistence does not seem to be caused by path-dependence as put forward by the EE literature. Our results call for a more nuanced understanding of the meaning of the HGF shares both as an economic indicator as well as in terms of its operationalization within the current EEs framework.

Our study contributes to extant literature in several ways. First, it adds to the rich literature on HGF. As already pointed out, most studies in this field adopt a firm-level perspective (see, e.g. Coad et al,. 2014), whereas fewer have addressed this topic from a regional angle (e.g. Sleuwaegen and Ramboer, 2020; Friesenbichler and Hölzl, 2020, Fotopoulos, 2022). We add to this strand by leveraging official regional data encompassing up to 20 European economies. Second, the paper speaks to the literature addressing persistence in regional entrepreneurial dynamics (see, e.g. Fritsch and Mueller, 2007; Fritsch and Wyrwich, 2014). Similar to Friesenbichler and Hölzl (2020) and Coad and Srhoj (2023), we focus on regional HGF. Our results, based on a broader set of countries, suggest that persistence in regional HGF shares is generally stronger than found by these studies. Third, we also contribute to the strand that examines the mechanisms behind persistence in entrepreneurial activities (see, e.g. Andersson and Koster, 2011). We show that persistence in regional HGFs cannot be attributed to path-dependency. Finally, we contribute to the literature on EEs (e.g., Stam, 2015; Spigel, 2017). Our empirical findings are not aligned with EE theoretical predictions and may be useful for its refinement.

The remainder of the paper is organized as follows. Section 2 provides a background based on the theoretical literature and develops some hypotheses. Section 3 presents the methodology. Section 4 describes the data, and Section 5 contains the analysis. Section 6 contains a discussion of our results and their implications, and Section 7 concludes.

2. Background and hypotheses

Economists and policy-makers have continually wondered how to replicate the Silicon Valley model in their own territories. Why is it that some regions seem to consistently outperform others in terms of their ability to produce innovative high-impact entrepreneurship? A number of theories have emerged in the literatures of economic geography and innovation studies to attempt to answer these questions, such as the frameworks of National Systems of Innovation (Lundvall, 1992; Freeman, 1995) and National Systems of Entrepreneurship (Acs et al., 2014), Regional Innovation Systems (Cooke et al., 1997; Fritsch, 2001), the cluster-based theory of competitive advantage (Delgado, Porter and Stern, 2010; Moretti, 2021), the Triple-helix approach (Etzkowitz and Leydesdorff, 2000), National Innovative Capacity (Furman et al. 2002), Competence Blocs (Henrekson et al., 2010), Environments for Entrepreneurship (Malecki, 2018), and more recently the EE approach (Isenberg, 2014; Stam, 2015; Spigel, 2017; Sternberg et al., 2019). Our empirical analysis of regional HGF shares is structured according to the theoretical intuitions of Entrepreneurship Ecosystems theory (Leendertse et al., 2021; Coad and Srhoj, 2023). This section draws on this stream of literature to develop 3 broad groups of hypotheses that can be tested using available data.

2.1 Regional development and incidence of HGF

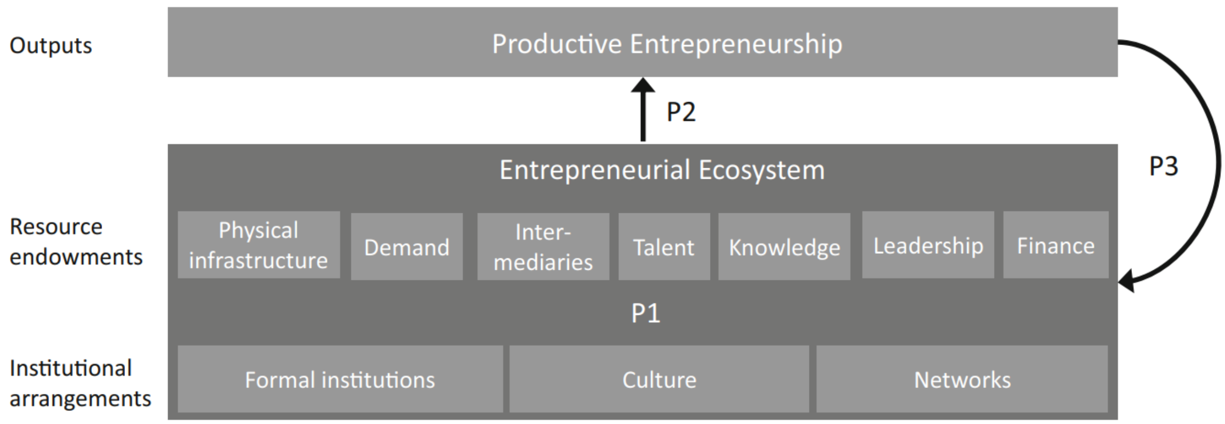
To begin with, we investigate suggestions that developed regions have higher rates of successful entrepreneurship in the form of regional HGF shares.

A distinction between the EE approach and other similar ancestors (e.g. the Clusters approach) is that the EE approach emphasizes entrepreneurs as the main agents (Harima, 2020). Hundt and Sternberg (2016, p. 227) write that "we assume that regions with high levels of GDP per capita demonstrate a higher level of entrepreneurial activity than economically less prosperous regions." For example, stronger regional institutions may enhance firm growth and firm’s integration in supply chains (Cainelli et al., 2023). This idea of “higher level of entrepreneurial activity” is specifically linked to HGFs, which are the preferred indicator of entrepreneurial activity considering that they are put forward as the output of the EE. In this vein, Stam and Van de Ven (2021, p. 809) state that:

> “[…] *the prevalence of high-growth firms in a region is strongly related to the quality of its entrepreneurial ecosystem*.”

Figure 1 visualizes the theoretical predictions of EE theory concerning the relationship between EEs and productive (high-growth) entrepreneurship. The success of EEs relies on the quality of their entrepreneurial inputs, which include elements such as physical infrastructure, demand, intermediate services, talent, new knowledge, leadership, finance, formal institutions, culture, and networks (Stam, 2015, Stam and Van de Ven, 2021, Leendertse et al., 2021). Presumably the quality of these inputs increases with a region's economic development. Given that more developed regions have better Ecosystem inputs, we would expect that regions that are more economically developed have higher HGF shares as output.

Figure 1: the 10 inputs and the output of an Entrepreneurial Ecosystem.



Source: Stam and Van de Ven (2021).

In order to illustrate the positive relationship between the quality of the EE and the share of high-growth firms using a slightly different analogy, consider the fact that initial entrepreneurial talent may be uniformly distributed across geographical space.[2] However, the appearance of high-performing EE regions comes about because some regions have better institutional frameworks, because some regions are better than others at nurturing and supporting the development of talented entrepreneurs, and also because of net migration.

All regions may have "heroes", but some regions direct their "heroes" into productive entrepreneurship, and support their ventures, spurring them on to success. Other regions may direct their "heroes" into unproductive entrepreneurship, and/or restrain their emerging ventures (through "blame" and "shame") to stifle their chances of success. Our analysis is framed in the regional ecosystem as a supporting or suppressing context: hence our focus is not on comparing regions in terms of how many "heroes" they have, but instead we focus on comparing regions in terms of what they do with the "heroes" that they get (and how they attract the immigration of heroes from elsewhere). A strong supporting ecosystem could lower the threshold between "potential hero" and "probably non-hero" to help even individuals who start off as "semi-heroes" to achieve their full potential – therefore the total number of heroes that emerge depends on the strength of the supporting ecosystem. The stronger the supporting ecosystem (input), the higher the share of local "heroes" (output).

Hypothesis 1: More developed regions will have higher regional HGF shares
Regional development can be measured in various ways. The most common indicator of economic development is arguably GDP per capita. However, given that the EE approach is often linked to discussions of the knowledge economy, innovation systems, high-tech entrepreneurship, and so on, it seems also worth investigating regional development in terms of indicators of scientific and technical knowledge, i.e. patents

[2] To assume otherwise could potentially lead to theories bordering on nationalism and racism, that are deeply problematic and best avoided.

per capita, and research and development (R&D) expenditures per capita. Finally, to better connect to previous empirical research on EEs, we take the index developed by Leendertse et al (2021, Appendix D) which is meant to directly capture the sophistication of a region's EE.

Hypothesis 1A: Regions with a higher GDP per capita will have higher regional HGF shares

Hypothesis 1B: Regions with a more patents per capita will have higher regional HGF shares

Hypothesis 1C: Regions with a more R&D investment per capita will have higher regional HGF shares

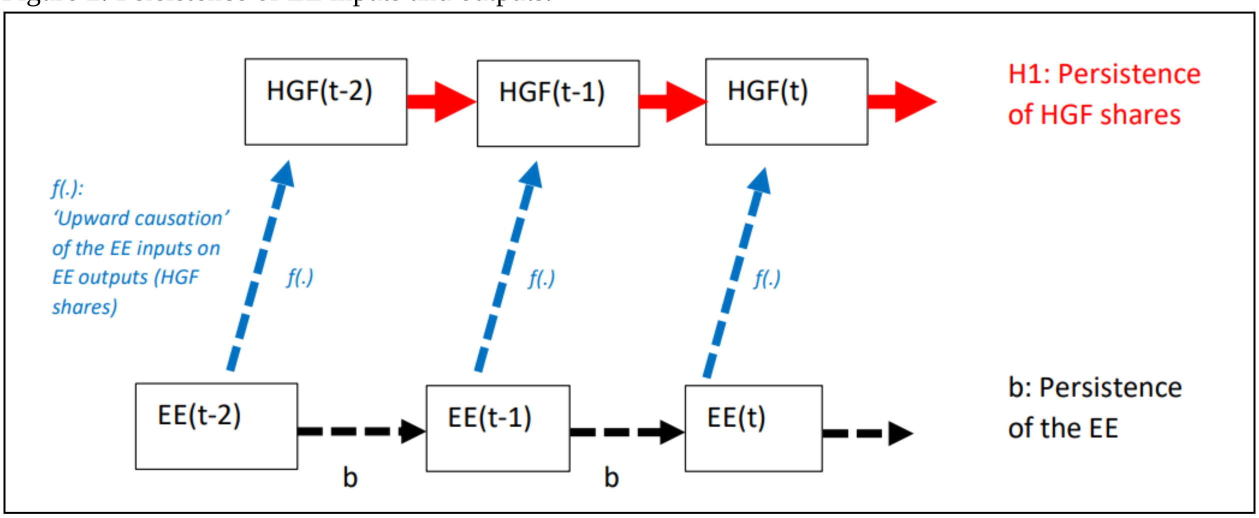
Hypothesis 1D: Regions with a higher EE index will have higher regional HGF shares

2.2 Regional persistence of HGF shares

Previous research has highlighted the persistence of entrepreneurial activity at the regional-level (Andersson and Koster, 2011; Fritsch and Wyrwich, 2014). Regarding regional HGF shares, our reasoning here proceeds along two lines: first, considering the persistence of EE inputs ; and second, considering the role of business accelerators.

First, the standard view of an EE is that the outputs depend on a number of inputs (Figure 1) such as institutions, infrastructure, and culture. A key characteristic of these inputs, in statistical terms, is that they change little over time. Institutions, infrastructure, culture, and the factors in Figure 1 more generally, are highly persistent and often considered to be time-invariant (Coad and Srhoj, 2023). If the inputs are time-invariant, then under most functional forms that map the inputs into outputs, we can also expect the outputs to be time-invariant (Figure 2).

Figure 2: Persistence of EE inputs and outputs.



Source: Coad and Srhoj (2023). We test the hypothesis of HGFs persistence represented by the thick red arrows.

Second, an alternative line of reasoning relates to the empirical literature that has found rigorous evidence that effective institutions can boost the number of so-called "gazelles" (i.e. high-growth young firms). Evidence from Chile on the evaluation of Business Accelerators has identified a causal effect of business accelerators on firm performance and rapid growth (González-Uribe and Leatherbee, 2018; González-Uribe and Reyes, 2021).

We do not dispute that these studies are showing causal effects of supporting institutions (i.e. business accelerators) on gazelles. Let us take these findings further, in the context of a thought experiment: imagine that we can randomly decide which regions will start this business accelerator programme, and which regions will not. If the business accelerator produces additional HGFs, then randomly allocating these business accelerator programs (i.e. randomly strengthening the EE) will have a causal effect on the output (HGFs).

Unless these causal effects are somehow temporary, or somehow cancelled-out by general equilibrium effects (due to negative effects in non-supported firms, e.g. Kwon and Sorenson, 2021), then we would expect that a top-performing institution in one period will also be top-performing in the next period, leading to persistence in inputs, and as a consequence, persistence in outputs (Coad and Srhoj, 2023).

Together, these two lines of reasoning suggest that there would be regional persistence in HGF shares. Regions with consistently above-average HGF shares as output would correspond to regions with above-average EE inputs. Regions with consistently below-average HGF shares would presumably be chronically underdeveloped in terms of EE inputs. This leads to the following hypotheses:

Hypothesis 2: Given the regional persistence of the inputs, there is regional persistence of HGF shares

Hypotheses 2 can be tested in different ways, depending upon the considered time period:

Hypothesis 2A: There is high persistence in regional HGF shares in two consecutive three-year periods

Hypothesis 2B: There is high persistence in regional HGF shares even with longer lags

2.3 Regional development and persistence of HGF shares

Hypothesis 1 investigated whether more developed regions have higher regional HGF shares. The degree of regional-level economic development is usually quite stable over the time period studied here, which is 3-year periods (given the standard HGFs definition that we use). While catchup and leapfrogging do occur, they tend to be exceptions, affecting a few outliers in rare cases. In most cases, the relative level of regional economic development changes little from one period to the next.

Hypothesis 1 (i.e. better EEs in more developed regions) can therefore be combined with Hypothesis 2 (i.e. EEs are persistent) to derive a third set of hypotheses that relate to whether regional HGF shares are more persistent in more developed regions.

In more developed regions, the emergence of HGFs is not a coincidence, but comes from systematic support, stable institutions, a robust accumulated scientific knowledge base, reliable management routines and business practices, and dependable infrastructure – all of which are persistent. In this vein, the innovation literature has investigated "innovation persistence" because this corresponds to a case where outcomes are not chance events but are presumably linked to stable latent capabilities (e.g. Cefis, 2003; Raymond et al., 2010). In contrast, in less developed regions, the supporting conditions (i.e. the EE inputs) are less well developed, and HGFs are expected to be rare. In the cases where HGFs are observed, this could be due to one-off lucky breaks rather than systematic dependable regional capabilities. In less well-developed regions, the emergence of HGFs could be due to chance fluctuations that relate to the "law of small numbers" (Kahneman, 2011, chapter 10)[3] – and therefore are unlikely to persist.

We therefore posit Hypothesis 3, and then operationalize it using the same four indicators as above (Hypotheses 1a-1d).

Hypothesis 3: Regional persistence of HGF shares is higher in more developed regions

Hypothesis 3A: Regional persistence of HGF shares is higher in regions with higher GDP per capita

Hypothesis 3B: Regional persistence of HGF shares is higher in regions with more patents per capita

Hypothesis 3C: Regional persistence of HGF shares is higher in regions with more R&D investment per capita

Hypothesis 3D: Regional persistence of HGF shares is higher in regions with higher EE index

These above hypotheses speak to the sources behind the potential persistence of HGF shares. In particular, they address a mechanism envisioned by the EE literature that deals with path-dependence.[4] According to EE theory, there exists a self-reinforcing mechanism between entrepreneurial inputs and outputs (see arrow P3 in Figure 1). This path-dependent process would generate persistence in regional HGF shares.[5] If we assume the presence of such self-reinforcing mechanism, this would entail that the strength in persistence is dependent not only on having a better EE, but also on past levels of HGF shares. In other words, regions that feature higher levels of HGFs in the past, will enjoy an enduring advantage in producing HGFs in the future.

This leads us to posit a further hypothesis:

Hypothesis 3E: Regional persistence of HGF shares is higher in regions with high HGF shares in the past

[3] To illustrate the idea behind the "law of small numbers": you are more likely to have 100% heads if you flip a coin 4 times than if you flip a coin 400 times. In such cases where the outcome is a proportion, having a small sample size increases the variance, but not the mean. Note that the law of small numbers is related to the volatility of outcomes but it does not affect the persistence of outcomes. To stay within the analogy, we assume that all regions have HGF shares that are randomly-distributed, but that well-developed regions have a higher mean than less-developed regions. Hence, well-developed regions are more likely to have longer spells of relatively high HGF shares.

[4] Path dependence refers to phenomena for which later conditions are dependent on previous ones. Prior evidence has shown that path-dependence is indeed one of the sources behind persistence in entrepreneurial dynamics, such as start-up rates (see, e.g., Andersson and Koster, 2011). From the EE standpoint, Wurth et al (2021) argue for the presence of a "strong path dependence in the evolution of entrepreneurial ecosystems". In addition, they write that "EE should be treated as a system (strong path-dependency within its evolution), with overall quality positively related to entrepreneurial output, which in turn feeds back into the regional EE" (p. 764).

[5] For instance, regions that have experienced superior levels of HGFs in the past may develop a strong entrepreneurial culture that induces persistence in high-growth entrepreneurship (see, e.g., Andersson and Koster, 2011).

3. Methodology

Given the various approaches taken in the literature, and the various pitfalls and statistical fallacies that may bedevil empirical research in this area, this section provides a detailed discussion regarding the many methodological choices to be made.

3.1 Regional level of analysis

In a study of regional HGFs persistence, the initial methodological choice relates to the level of geographic aggregation. In the European context, previous studies have focused at the level of local authority districts, NUTS-3 and NUTS-2 (Nomenclature of Territorial Units for Statistics) regions, as well as at the country level (Table 1). In our view, the country level is too aggregated, and furthermore it would yield only a small number of observations for analysis of statistical significance.

Table 1. Studies on HGFs at the different territorial units of analysis

Unit of analysis	Study	Sample
LAD	Fotopoulos (2022)	UK
NUTS-3	Friesenbichler and Hölzl (2020)	Austria
	Coad and Srhoj (2023)	Croatia and Slovenia
	Audretsch and Belitski (2021)	9 countries: Bulgaria, Croatia, Czech Republic, Denmark, Hungary, Italy, Portugal, Romania, and Slovakia.
NUTS-2	Belitski et al. (2021)	21 European countries
	Leendertse et al. (2021)	28 European countries
	Mikic et al., (2021)	8 European countries
	Stam and Van de Ven (2021)	Netherlands
	Sleuwaegen and Ramboer (2020).	13 European countries
Country	Corrente et al. (2019)	24 European countries
	Bravo Biosca, Criscuolo and Menon (2016)	11 countries

Notes: LAD stands for Local Authority Districts. This is a classification used in Great Britain which identifies regions that are generally smaller than NUTS-3 regions.

Much research on EE and HGF shares is conducted at the level of NUTS-2 regions, primarily due to data availability (e.g. Leendertse et al., 2021). However, recent literature examines HGF shares at NUTS-3 or city/municipality level (Coad and Srhoj, 2023; Fotopoulos, 2022; Friesenbichler and Hölzl, 2020). The NUTS-3 level is seen as superior given the considerable heterogeneity of NUTS-3 regions among different NUTS-2 regions (see Coad and Srhoj, 2023). For our primary analysis, we provide results at the NUTS-3 level, nevertheless, given the amassed research at NUTS-2 level, we repeat the analysis at the NUTS-2 level as a robustness check.

3.2 HGF share definition

HGFs have a precise and well-known definition that comes from Eurostat-OECD (2007). The HGFs indicator is a dummy variable that equals 1 for firms with 10 or more employees (E) in the initial period (t-

3), and a geometric average of at least 20% growth (in terms of either sales or employment) per year over 3 years. More recently this growth criteria has sometimes been reduced from 20% per annum to 10% per annum (e.g. Flachenecker et al., 2020; Benedetti Fasil et al., 2021; OECD, 2021a), perhaps with the goal of facilitating statistical analysis by boosting the number of observations for HGFs or capturing more medium and larger firms for whom it is difficult to achieve a 3-year 72.8% growth (with the yearly 20% growth indicator).

In our study we leverage the official definition of HGF shares provided by Eurostat. According to this, the numerator in the HGF shares corresponds to the number of HGFs in a given region at time t. Eurostat definition[6] (details in Section 4) uses 10% growth in employment per annum. In other words, we measure HGFs in terms of employment growth using the size indicator $X = \{employees\}$, $HGF_t = 1$ if the following conditions are satisfied:

$$\left(\frac{X_t}{X_{t-3}}\right)^{\frac{1}{3}} - 1 \geq 10\% \qquad (1)$$

With a restriction on initial size:

$$E_{t-3} \geq 10 \qquad (2)$$

Hence, the growth rate threshold to be an HGF is 33.1% over a three-year period. According to the official data provided by Eurostat, the denominator of the HGF shares corresponds to all firms with 10+ employees in the region at time t. In the Appendix we discuss different adjustments to the HGF denominator, e.g. the use of all firms with ≥10 employees in the region at time t-3, and provide estimation results for these alternative ratios.

3.3 Estimation strategy

Amid the paucity of systematic evidence from representative datasets on EE and HGFs, this paper's goal is to contribute some relatively simple and transparent evidence regarding HGFs and their persistence across regions in the EU, to test ideas circulating in the theory.

Hypothesis 1 stated that more developed regions will have higher regional HGF shares. This is a fairly straightforward unconditional statement that can be investigated without a long list of control variables.[7] According to EE theory, the EE inputs (shown in Figure 1) are the theoretical determinants of the EE output (i.e. regional share of HGFs). Previous empirical research has sought to explain the regional share of HGFs

[6] Eurostat provides data on HGF shares at the regional-level based on employment, whereas HGF shares based on both revenues and employment are available only at the national-level.

[7] For example, Hypothesis 1, and the literature on which it draws, do not make statements about partial relationships such as "Holding constant a set of background factors such as regional size and quality of regional governance, more developed regions will have higher regional HGF shares." Instead of making conditional statements, EE theory aspires to broad-based predictions between regional development and outputs that hold unconditionally. Note also that the addition of explanatory variables sometimes corresponds to the introduction of new biases and complications such as collider bias (Elwert and Winship, 2014).

in terms of EE inputs, although using idiosyncratic indicators of HGF shares (Stam and Van de Ven, 2020; Leendertse et al., 2021). Instead, our goal is to investigate the relationship between HGF shares and regional economic development at a more basic level. We do not wish to engage in detailed debates about what are the correct empirical operationalizations of the various EE inputs, instead we wish to remain at a more fundamental level and test more basic predictions that are often taken for granted. We do not control for a region's size, because our dependent variable is a proportion and therefore can be considered to be scale-invariant. When pooling together years, however, we include time dummies to control for potential macroeconomic fluctuations, in recognition of previous work that found that HGF shares vary over the business cycle (e.g. Stam and Van de Ven, 2020).

In the regression equation below, the explanatory variable (regional development) essentially proxies for the various inputs of an EE (shown in Figure 1) that presumably are positively related to economic development.

$$HGF\ share_{jt} = a + \beta_1 Devlp_{jt-k} + e_{jt} \qquad (3)$$

Where $HGF\ share_{jt}$ is a continuous dependent variable capturing the share of HGFs in each NUTS-3 region j, for each three-year period t. $Devlp_{jt-k}$ is an independent categorical variable capturing terciles of GDP per capita (H1a), patents per capita (H1b), R&D investments per capita (H1c), and EEI (H1d). To alleviate some of the potential endogeneity, $Devlp_{jt-k}$ is measured with a lag k that varies for the various proxies that are taken, depending on data availability. For GDP, patents and R&D investments, $Devlp_{jt-k}$ is measured as a mean over the available period until 2007, therefore, prior to any HGF shares calculations. EEI exists only as a snapshot in time (Leendertse, et al., 2021), and is therefore used in the available format. These key independent variables are always interpreted taking the least developed group of regions as the reference category. Time fixed effects are included in all specifications. The term ε corresponds to errors that are not directly observed to the econometrician.

Hypothesis 2 investigates whether there is regional persistence of HGF shares:

$$HGFs\ share_{jt} = \alpha + \beta_2 HGF\ share_{jt-3} + \varepsilon_{jt} \qquad (4)$$

Investigation of Hypothesis 2, using the regression equation above, also does not require a long list of control variables. We do not control for the inputs to an EE, under the usual assumption that the inputs have already been converted into the EE output, and hence we focus merely on describing the dynamics of the output (Coad and Srhoj, 2023).

When considering lags, we take a three-year lag, such that the lagged period ends before the current period starts. Indeed, if the lagged period and the current period are overlapping, this introduces bias in the form of a positive correlation between these two periods. Appendix 2 shows mathematically that, unless we have a 3-year lag, we will introduce bias (specifically, in the form of positive autocorrelation).
Hypothesis 3 investigates how the persistence parameter β_2 (i.e. β_2 from the equation above) varies with the level of economic development of regions. Here too, economic development is proxied by regional-level GDP per capita (H3a), patents per capita (H3b), R&D investment per capita (H3c) or EEI (H3d). One way of testing H3, therefore, would be to estimate the persistence parameter β_2 from (4) on subsamples of low vs high economic development. An alternative way of testing H3 would be to introduce an interaction term that explores how the persistence of HGF shares varies with economic development:

$$HGF\ share_{jt} = \vartheta + \beta_3 HGF\ share_{jt-3} + +\beta_4 Devlp_{jt-k} + \beta_5 Devlp_{jt-k} HGF\ share_{jt-3} + \epsilon_{jt} \quad (5)$$

H3 suggests that more developed regions have higher persistence in regional HGF shares, therefore H3 suggests that the coefficient on β_5 from equation (5) will be positive.

All of the above equations are estimated by means of Ordinary Least Squares (OLS) and clustering standard errors at the regional-level. Eq. (3) and (4) are also estimated using quantile regressions (Parente and Santos Silva, 2016). These allow us to dig deeper into an analysis of the heterogeneity across regions, investigating the relationship between economic development and HGF shares at various quantiles of the (conditional) HGF shares distribution. This seems worthwhile, because there is theoretical interest in supporting lagging regions (at the bottom of the HGF shares distribution), as well as interest in studying and learning from the best-performing regions (at the top of the HGF shares distribution). Moreover, estimating Eq. (4) using quantile regressions will allow us to test Hypothesis 3e, namely, whether regions that at the top of the HGF shares distribution in one period, are able to stay at the top in the next (see, e.g. Andersson, & Koster, 2011). This would speak to possible path-dependency in regional HGFs according to EE theory (see Wurth, et al., 2021).

Overall, we also rely on complementary empirical methods, e.g. ranging from graphical representations of our underlying data to simple correlation coefficients as additional evidence base to test our three stated hypothesis.

A common methodological choice is the inclusion of unit-specific fixed effects to control for unobserved heterogeneity in panel data. This is deliberately avoided here. Regarding Hypothesis 1, including fixed effects in equation (3) would be problematic because it would control away for the time-invariant aspects of regional HGF shares, such as regional development. Any time-invariant aspect of regional development would be removed by the fixed effects, which would be problematic for our investigation of the role of regional development on HGF shares. Regarding Hypotheses 2 and 3, including fixed effects in equation (4) would be problematic because we are interested in evaluating the degree of persistence of HGF shares. If the persistent component of HGF shares is removed via fixed effects, this would give a biased view of the persistence of HGF shares, exaggerating the time-varying component. Hence, we do not include region-specific fixed effects. Other authors (e.g. Fotopoulos, 2022), too, have expressed concerns about including fixed effects in studies of the persistence of regional HGF shares.[8]

[8] A potential additional concern for our statistical analysis could be a discontinuity in the distribution of HGF shares, such that there might be a mass point at zero. Small regions may have a small number of firms with 10+ employees in general, and perhaps in some years they may (due to chance fluctuations) have zero HGFs, hence an HGF shares of zero. This could lead to various econometric difficulties associated with truncated variables such as non-Gaussian residuals. Relatedly, if small regions have a small number of firms overall, then this could lead to higher volatility in HGF shares, according to the "law of small numbers" (Kahneman, 2011). We therefore pay attention to whether regions have HGF shares of zero, but fortunately this does not appear to be a problem for our analysis.

4. Data

Our analysis is based on Eurostat data (see Table 2 for the variables' description). This is not microdata (unlike some previous work such as Friesenbichler and Hölzl (2020) and Coad and Srhoj (2023)) but it is micro-aggregated and publicly available.[9] These data provide two main advantages when studying HGF shares: i) they are based on representative data and avoid the tipycal pitfalls concerning selection bias attributable to e.g. commercial data sources[10]; ii) they are available for a sizeable share of European countries, thus allowing for a cross-country analysis.

Table 2: Description of variables

Variable	Description	Available at:	
		NUTS-3	NUTS-2
	Dependent variable		
HGF shares	Share of high growth enterprises measured in employment *(in t):* number of high growth enterprises divided by the number of active enterprises *(in t)* with at least 10 employees. Includes industry, construction and services except insurance activities of holding companies. Obtained from Eurostat (code: bd_hgnace2_r3). HGF shares capture three-year periods, which is why the variable is used only in years 2008, 2011, 2014, 2017, and 2020.	✓	✓
	Independent variables		
GDP per capita	Gross domestic product (GDP) in purchasing power parity (PPP, EU27 from 2020) per inhabitant. Obtained from Eurostat (code: nama_10r_3gdp). Constructed as predetermined variable capturing mean GDP per capita in the period 2000-2007.	✓	✓
Patents per capita	Number of patent applications filed annually in a given region based on the address of applicants/inventors. Constructed as a predetermined variable capturing mean patents per capita in the period 1997-2007. Source: OECD, REGPAT database, August 2022. Mean patents in the period 1997-2007 is divided by the average annual population (obtained from Eurostat, code: nama_10r_3popgdp).	✓	✓
R&D investment per capita	R&D expenditure purchasing power parity (PPP) per inhabitant at constant 2005 prices. Obtained directly from Eurostat (code: rd_e_gerdreg). Constructed as predetermined variable capturing mean R&D expenditure per inhabitant in the period 1997-2007.		✓
Entrepreneurial Ecosystem Index	Quality of EEs, developed by Leendertse, Schrijvers and Stam (2021) (see Table D1). The EE index is a snapshot in time, and available in two different versions: the EE index additive and the EE index log.		✓

The main dependent variable, HGF shares at the NUTS-3 and NUTS-2 levels, is not available for all EU countries, and regional coverage increases with time (Table 3). Our sample is an unbalanced panel covering up to 20 European countries throughout the period 2008-2020. In the results section we present NUTS 3 level results, however, in the robustness analysis we provide also NUTS-2 level results. In comparison to existing studies (e.g. Friesenbichler and Hölzl (2020) for Austria; Coad and Srhoj (2023) for Croatia and

[9] Data available at this link: https://ec.europa.eu/eurostat/web/structural-business-statistics/business-demography [last accessed: 12th Jan 2023]

[10] An example of this is the use of datasets such as ORBIS, which often do not provide exhaustive coverage of balance-sheet data for young and small firms thus introducing bias when it comes to computing HGF shares.

Slovenia; Fotopoulos (2022) for the UK) our persistency analysis of HGF shares has the broadest geographical scope and largest sample size. Importantly, two variables (R&D investment *per capita*, and Entrepreneurial Ecosystem Index - EEI) are measured at NUTS-2 level, but not at the NUTS-3 level. We still use these two variables in examining the heterogeneous persistency at the NUTS 3 level by merging their NUTS-2 level values to each NUTS 3 region within a NUTS-2 region. Table 4 presents some summary statistics for NUTS-3 regions.

Table 3: Country coverage over time

Year	N	# countries	Countries added (cumulative)
2008	60	2	+ Austria, Portugal
2009	60	2	
2010	60	2	
2011	470	13	+ Bulgaria, Denmark, Spain, Finland, France, Hungary, Italy, Lithuania, Malta, Romania, Slovakia
2012	505	14	+ Croatia
2013	559	16	+ Czechia, Netherlands
2014	556	15	- Malta
2015	509	15	+ Estonia ; - Spain
2016	644	18	+ Spain, Latvia, Poland
2017	647	18	
2018	668	19	+ Sweden
2019	665	20	+ Malta
2020	665	20	

Notes: Malta has only two NUTS-3 regions. Due to changes in NUTS-3 regions in years 2019 and 2020 the number of observations has a small decrease although number of countries (due to Malta) increases.

Table 4: Summary statistics at NUTS-3 level

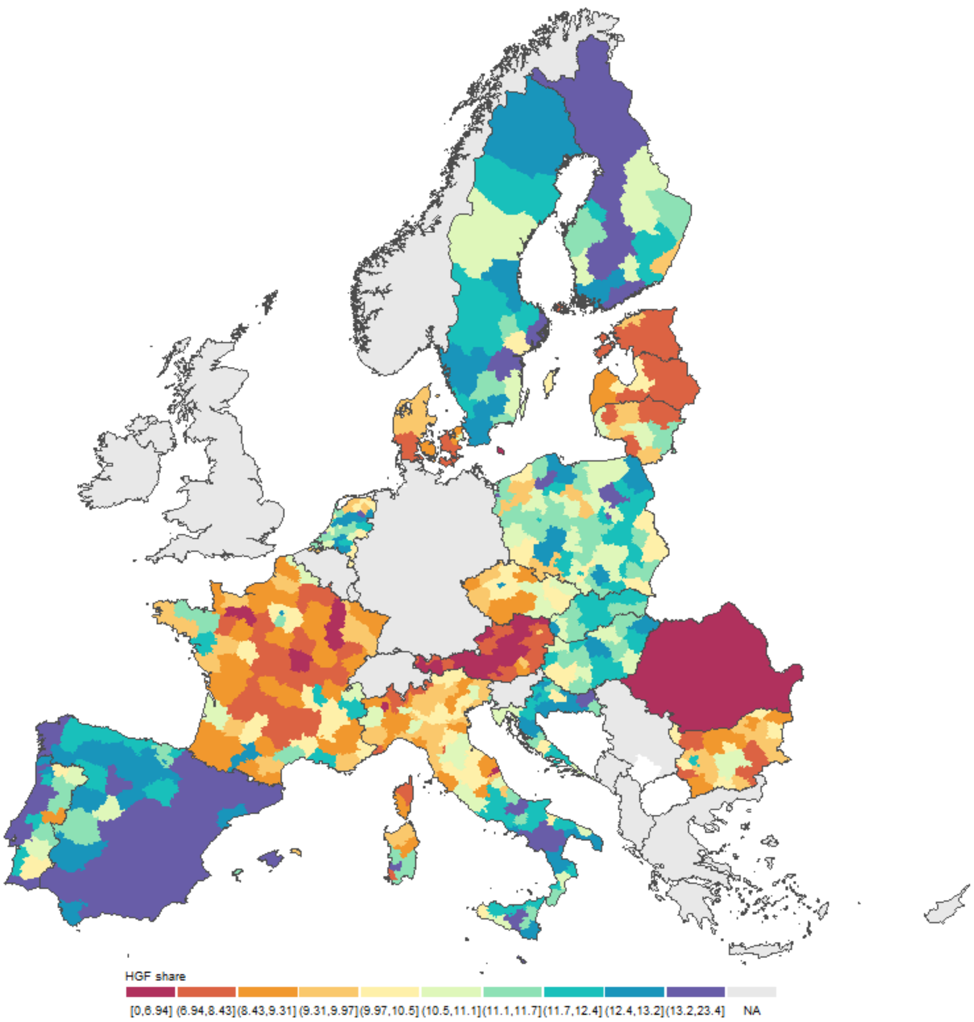
	N	Mean	St. Dev.	Min	Median	Max
Panel A						
HGF shares (in %)	6,068	8.94	3.25	0.00	9.00	27.78
GDP per capita	6,066	17,530	8,415	979	17,763	68,900
Patents per capita	6,066	0.05	0.14	0.00	0.01	2.64
R&D investment per capita	5,478	261.60	280.45	7.03	161	1,791
EEI Additive	5,994	6.38	5.68	1.26	4.19	35.08
EEI Log	5,994	-9.07	6.90	-21.96	-9.82	10.58
Panel B						
HGF shares (in %)	2,398	8.64	3.14	0.00	8.65	27.78
GDP per capita	2,397	17,585	8,388	979	17,788	68,900
Patents per capita	2,397	0.05	0.14	0.00	0.01	2.64
R&D investment per capita	2,183	259.17	278.78	7.03	161	1,791
EEI Additive	2,369	6.34	5.63	1.26	4.21	35.08
EEI Log	2,369	-9.08	6.85	-21.96	-9.82	10.58

Notes: Panel A includes all years. Panel B includes years 2008, 2011, 2014, 2017, and 2020. Number of regions with HGF shares equal to zero in Panel A is 12 (out of 6068) and in Panel B is 5 (out of 2398).

5. Results

5.1. Hypothesis 1: More developed regions will have higher regional HGF shares

Figure 3: Distribution of mean HGF shares over the period 2016-2020. This figure relates to H1



In this section we provide evidence to support or dispute Hypothesis 1, namely, whether more developed regions have higher regional HGF shares. We start by providing some basic descriptive statistics, in the form of a map (Figure 3) and Table 5 with top-15 performing regions. Figure 3 shows distribution of mean HGF shares over the period 2016-2020 (main dependent variable) across 20 EU Member States and their 643 NUTS 3 regions. The map below highlights the conundrum: less developed regions (e.g. south-eastern and insular Spain, southern and insular Italy, southern Portugal) seem to be the EE “hotspots” in terms of having higher HGF shares, which goes against common assumptions and EE theory.[11]

[11] For instance, according to the EE additive index developed by Leendertse et al. (2021), the top five entrepreneurial ecosystems at NUTS-2 level are: 1) Copenhagen (DK01), 2) West and East Inner London (UKI3&4); 3) Berkshire,

Table 5 shows the top-ranking regions in terms of HGF shares – i.e. the regions that EE theory considers as the top-performing entrepreneurial regions, that other regions should aspire towards. These regions are mainly in Spain, Italy, Portugal, and Finland. The top five EEs in Europe, apparently, are La Gomera and Lanzarote in the Canary Islands, Valencia in Spain, Pohjois-Pohjanmaa in Finland, and Medio Campidano in Sardinia, Italy. This does not match well with popular conceptions of Europe's entrepreneurial hotspots.

Table 5: Top-ranking regions in terms of average HGF shares during 2016-2020.

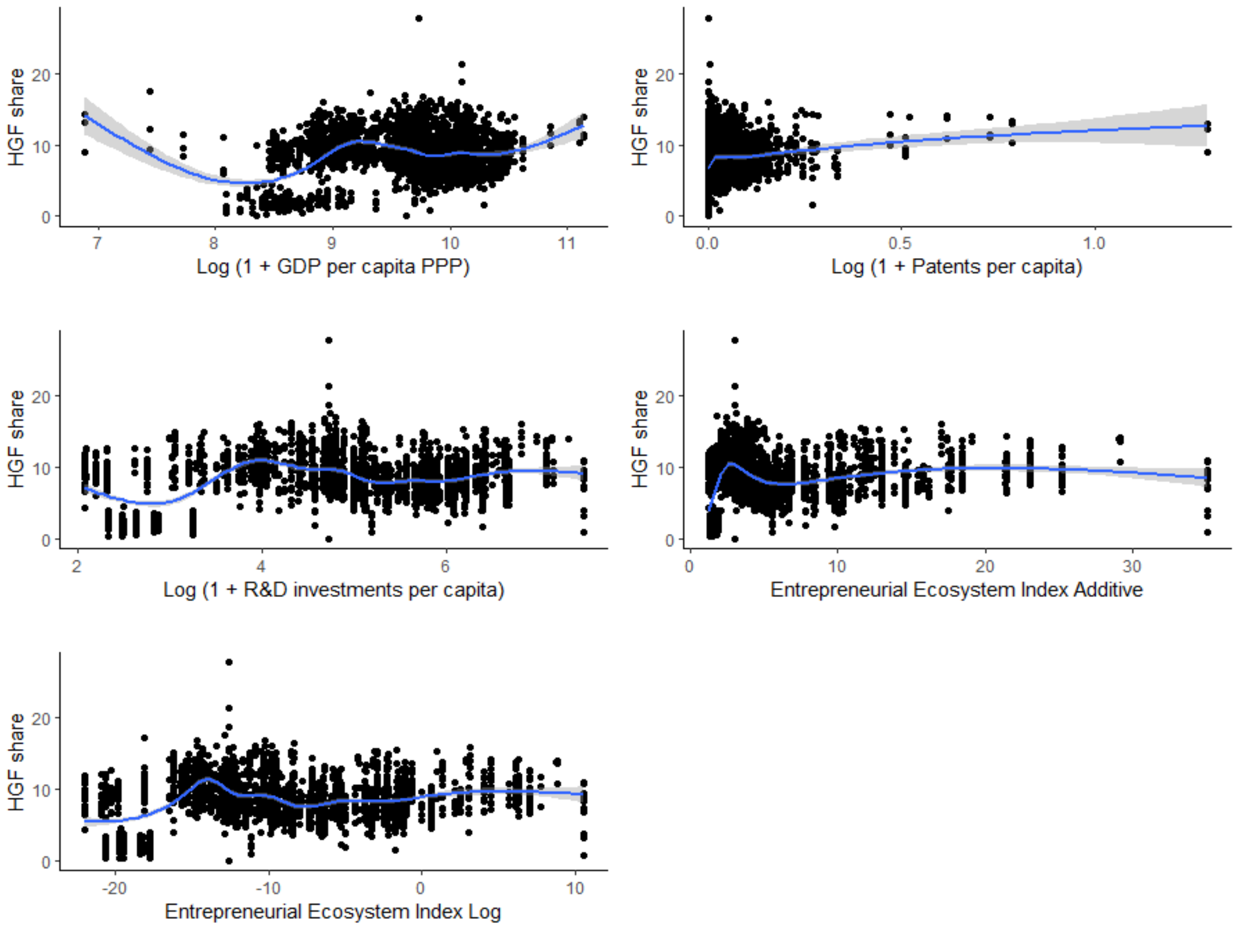
Rank	Top performing regions	Regional HGF shares
1	La Gomera (Spain)	23.44
2	Medio Campidano (Italy)	17.08
3	Lanzarote (Spain)	15.95
4	Pohjois-Pohjanmaa (Finland)	15.82
5	Valencia (Spain)	15.47
6	Oeste (Portugal)	15.31
7	Potenza (Italy)	15.06
8	Caltanissetta (Italy)	15.04
9	Fuerteventura (Spain)	15.01
10	Lleida (Spain)	14.95
11	Toledo (Spain)	14.94
12	Castellón/Castelló (Spain)	14.93
13	Região de Leiria (Portugal)	14.80
14	Madrid (Spain)	14.73
15	Murcia (Spain)	14.73

The initial descriptive information does not seem to provide support for Hypothesis 1. The rest of this subsection conducts a more formal analysis, testing whether regions with higher GDP per capita, more patents per capita, more R&D investment per capita and higher EEI have higher regional HGF shares (Hypotheses 1A, 1B, 1C and 1D).

We start the analysis with scatterplots of HGF shares and our four independent variables of interest. Figure 4 shows scatterplot of HGF shares and GDP per capita PPP. According to EE theory, we would expect a positive linear relationship. However, ourresults do not seem to show a positive correlation between HGF shares and GDP per capita PPP.

Buckinghamshire and Oxfordshire (UKJ1); 4) Helsinki-Uusimaa (FI1B); 5) Stockholm (SE11). According to other popular metrics, such as the one published in the Global Startup Ecosystem Report, the top five entrepreneurial hotspots in Europe are 1) London; 2) Amsterdam; 3) Paris; 4) Berlin; 5) Stockholm.

Figure 4: HGF shares and regional development indicators.



Notes: Results at NUTS-3 level. Loess function applied.

In general, Figure 4 shows no clear relationship between HGF shares and any of the regional development indicators. We repeat this exercise for GDP per capita PPP, patents per capita and R&D investments per capita, but take logarithm of one plus the value (Figure A1). These results suggest a positive relationship between regional GDP per capita PPP and HGF shares among the regions with low HGF shares and low GDP, but after this initial positive relationship there seems to be no correlation. No clear relationship is visible for patents or R&D investments per capita. To test these relationships in a more formal manner, Table 6 provides Pearson correlation and Spearman's rank correlation. There is no or very low magnitude of correlation between HGF shares and regional development indicators.

Table 6: Correlation: HGF shares and regional economic development indicators. This table relates to H1.

	Log (1 + GDP per capita PPP)	Log (1 + Patents per capita)	Log (1 + R&D investment per capita)	EE Index Additive	EE Index Log
Pearson correlation	0.167	0.059	0.171	0.072	0.109
P-value	0.000	0.004	0.000	0.000	0.000
Spearman's rank correlation	0.018	-0.023	0.016	-0.018	-0.018
P-value	0.374	0.252	0.444	0.390	0.382
N	2,398	2,398	2,398	2,398	2,398

Next, we move to the regression analysis and estimate Eq. (3). Table 7 shows that a change from the first (lowest) to the second GDP per capita tercile is associated with a statistically significant increase in HGF shares, at the NUTS-3 level. The first column of Table 7 suggests that regions in the lower range of the distribution of GDP per capita (countries such as Romania, with low GDP per capita) have lower HGF shares than regions with medium levels of GDP per capita. Being in the top tercile of GDP per capita does entail an increase in HGF shares relative to the lowest tercile. Yet, it does not result in any statistically significant increase in HGF shares with respect to countries located at the second tercile. This result provides only weak support for H1a.

Moving from the first to the second tercile in patents per capita is associated with a strong increase in the HGF shares. However, surprisingly, moving from the first to the third tercile in patents per capita leads to an increase that is considerably smaller in magnitude and barely statistically significant. We do not go as far as interpreting this finding as an inverted U-shape caused by diminishing returns for patents, but simply as a lack of systematic evidence of the expected positive relationship between regional patents per capita and HGF shares. We therefore reject H1b.

Relatedly, moving from the first to the second tercile in R&D investment per capita is associated with higher HGF shares, though only marginally statistically significant. Moving from the first to the third tercile in R&D investments per capita is associated with higher HGF shares, however, the effect size is smaller and this association is only weakly statistically significant. This provides nuanced support for H1c. In the cases of H1a and H1c, we find weak support, although this relationship seems largely confined to some outliers at the low end of the distribution, as shown in Figure 4.

Finally, we do not find any evidence of a statistically significant positive relationship between EEI and HGF shares. In other words, being a top performer in terms of the quality of the EE does not result in an increased incidence of HGFs in a region.

Table 7: Linear regression - HGF shares and regional economic development

	Dependent variable:				
	HGF shares				
	(1)	(2)	(3)	(4)	(5)
GDP per capita PPP 2nd tercile	0.666** (0.271)				
GDP per capita PPP 3rd tercile	0.672*** (0.259)				
Patents per capita 2nd tercile		1.053*** (0.266)			
Patents per capita 3rd tercile		0.449* (0.261)			
R&D investment per capita 2nd tercile			0.514* (0.278)		
R&D investment per capita 3rd tercile			0.463* (0.272)		
EEI Additive 2nd tercile				0.142 (0.259)	
EEI Additive 3rd tercile				0.145 (0.264)	
EEI Log 2nd tercile					0.407 (0.264)
EEI Log 3rd tercile					0.244 (0.269)
Mean	8.64	8.64	8.64	8.64	8.64
Observations	2,397	2,397	2,183	2,369	2,369
R^2	0.194	0.202	0.191	0.184	0.186
Adjusted R^2	0.192	0.200	0.189	0.182	0.184

Notes: *p<0.1; **p<0.05; ***p<0.01. Reference category is always 1st (lowest) tercile.

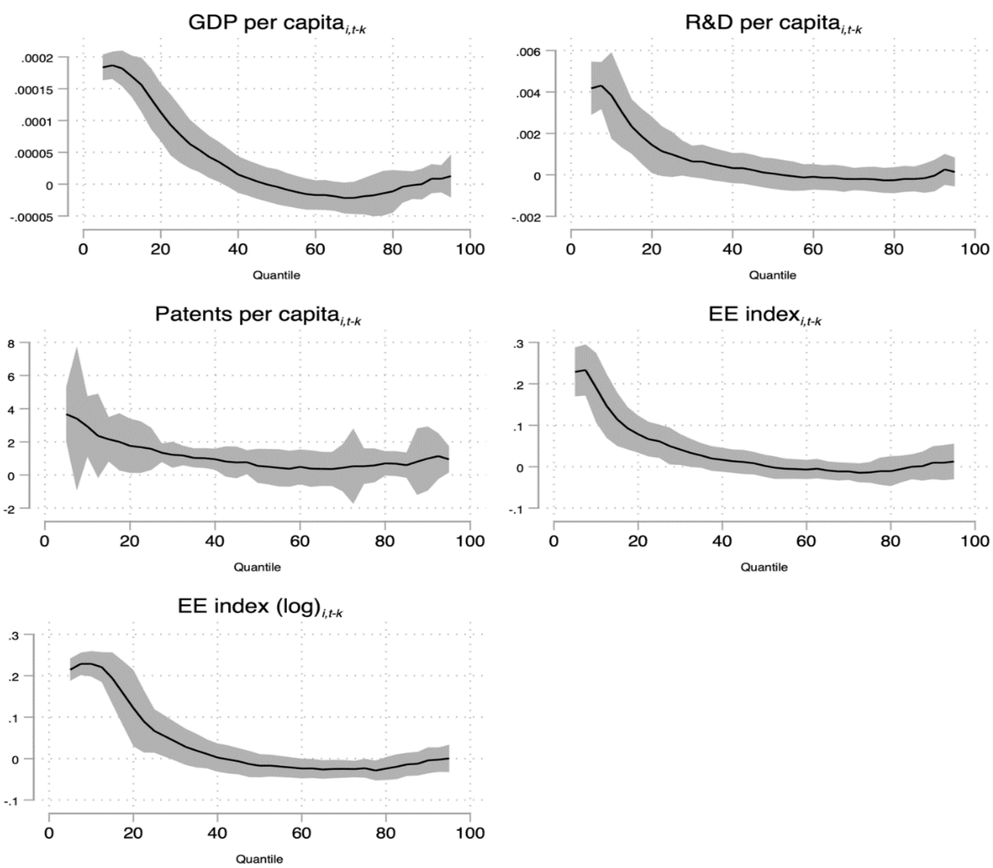
The linear regression is based on means, but there might be heterogeneity in the relationship between our dependent and independent variables. For this reason, we also estimate quantile regressions (see Figure 5 for a graphical representation of the results).

An increase in GDP per capita is associated with a small increase in the HGF shares among those regions with low HGF shares (at the 10^{th} and 20^{th} percentile). However, GDP per capita is negatively associated among regions with HGF shares at the 60^{th} or at the 70^{th} percentile. Patents per capita are positively associated with the HGF shares only among regions with very low HGF shares (10^{th} percentile), but less so for regions with higher HGF shares.

Quantile regression results for R&D investment per capita, and EEI show similar results, namely, an increase in R&D investments and EEI is associated with an increase in HGF shares only among regions with low HGF shares, but it is negatively associated or non-significant among regions with medium or higher HGF shares.

The literature on HGFs and EEs often focuses on the upper quantiles, where one finds the above-average performing firms and regions. This is in contrast with the results in Figure 5, which highlight different behaviour that is confined exclusively to the lowest quantiles. It seems that economic development matters at the lowest quantiles, presumably for removing some of the most basic obstacles and constraints to economic dynamism (as proxied by HGF shares). For most regions however, there is a negligible relationship between economic development and (conditional) HGF shares. In sum, results presented in the subsection show negligible support for our Hypothesis 1 as more developed regions are not associated with higher regional HGF shares.

Figure 5: Quantile regression plots.



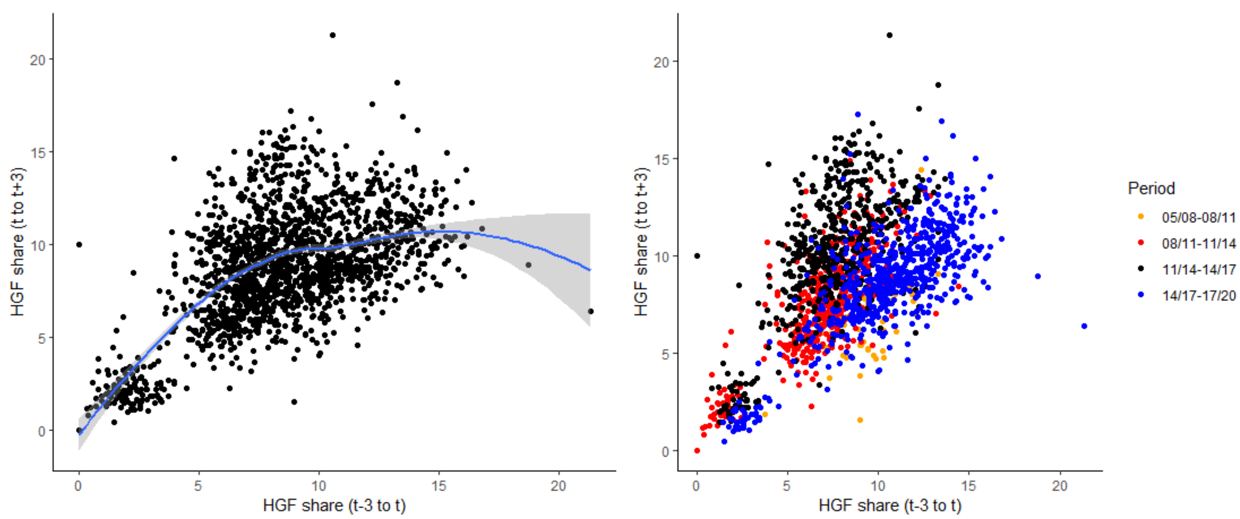
Notes: the plots report estimates from quantile regressions. All specifications include as dependent variable the HGF shares at the NUTS-3 level and control for year fixed effects. Standard errors are robust and clustered at the NUTS-3 level. Shaded areas represent 95% confidence intervals.

5.2. Hypothesis 2: Given the regional persistence of the inputs, there is regional persistence of HGF shares

In this subsection we answer the question: how strong is regional persistence of HGF shares? We start by presenting pooled regional HGF shares for two consecutive three-year periods (Figure 6, left). In the Figure 6 (right) we also show regional persistence where points are colored based on the two three-year periods.

There are two main clusters in Figure 6. Firstly, there is a smaller cluster of regions with low shares in both periods (left, bottom). More importantly, there is a large cluster in the middle of the figure indicating a variability in HGF shares between two periods, however, there seems to be some persistence between the regional HGF shares of two periods. The right graph shows that the magnitude of correlation might be affected by the time period, probably due to time-specific regional demand and macroeconomic conditions.

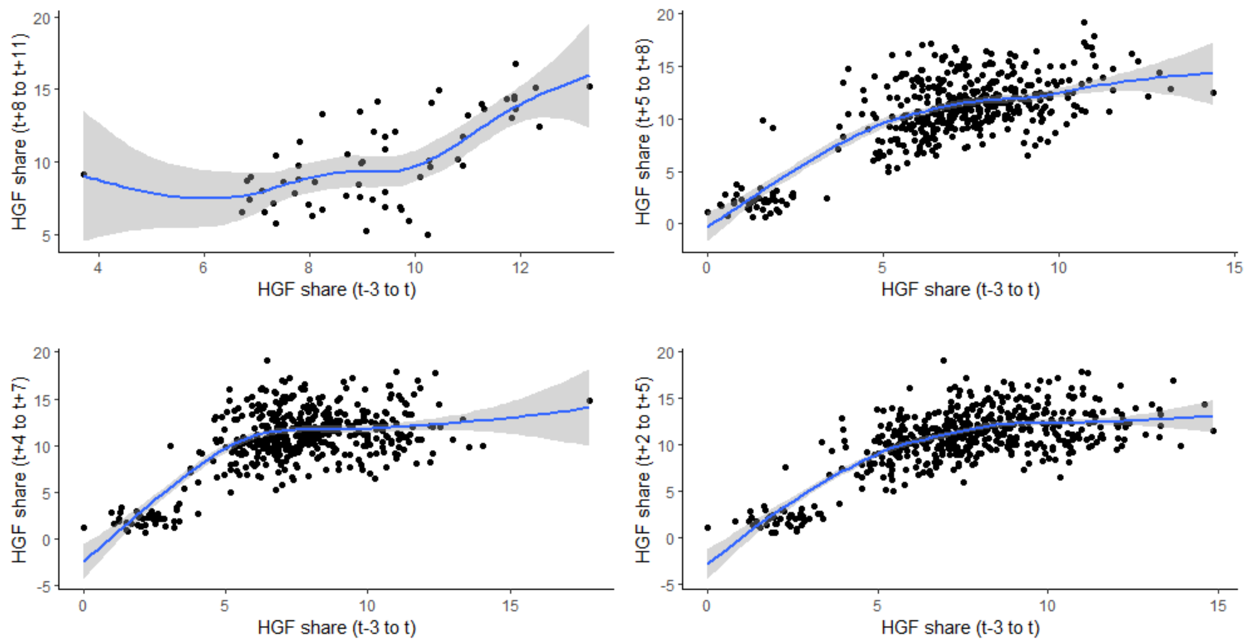
Figure 6: Persistence of regional HGF shares.



Notes: Left graph shows scatterplot of HGF shares in two consecutive three-year periods with smoothed conditional means using loess method. Right graph uses time variable to color the observations. Blue color for period 05/08 (t) and 08/11 (t+3); Red color for period 08/11 (t) and 11/14 (t+3); Black color for period 11/14 (t) and 14/17 (t+3). Color online.

Persistence might be lower with a longer lag, which is why we also plot regional HGF shares with a longer time lag (Figure 7). The dependent variable (on Y-axis) is based on HGFs information 5 to 11 years ahead of the independent variable (lagged HGF shares, on X-axis). Scatterplots are similar to the main scatterplot in Figure 6.

Figure 7: Persistence of regional HGF shares with longer lags.



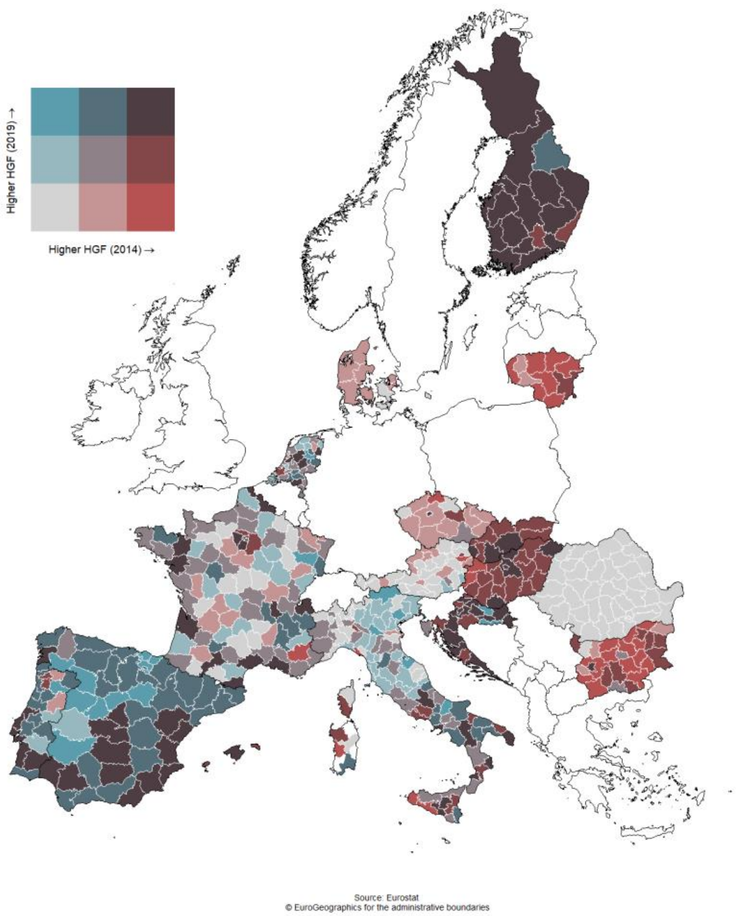
Notes: All graphs use 2016-2019 as the last period (t-3 to t). At the x-axis, upper left graph shows 05/08, upper right 08/11, lower left 09/12, and lower right 11/14.

In sum, Figures 6 and 7 show scatterplots for the persistence of regional HGF shares. Similar patterns are observed for the different years. However, HGF shares seem higher in some years than in others. The correlations generally look quite positive: more positive than what was observed previously for NUTS-3 regions in Croatia (Coad and Srhoj, 2023). There seems to be a separate cluster or regions that have low values of regional HGF shares. In other words, in the first graph, the distribution of regional HGF shares seems to be bimodal, with one mode at around 0.08 corresponding to the main group of datapoints, and then another mode at around 0.02 corresponding to a separate mini-cluster of observations corresponding to low HGF shares. Furthermore, there is a gap between these two groups, such that the regional HGF shares does not seem to be a continuous variable. Fortunately, there seems to be no problem of bunching at the lower bound (no bunching at HGF shares == 0.000), which is good news from a statistical perspective.

Figure 8 presents a bivariate map of the persistence of HGF shares. The bivariate map is shown for two periods, although the results are similar for the left and right maps. Some regions (shown with dark brown-purple shading) appear to have above-average HGF shares in consecutive periods: parts of Spain, Finland, southern Italy, Slovakia, etc. Other regions (shown with light shading) have relatively low persistency of HGF shares: Austria, Romania, and parts of France.

We continue with a more formal empirical exercise – Pearson correlation and Spearman's rank correlation. Table 8 shows correlation statistics, both the usual Pearson correlations as well as Spearman's rank correlations (that are more robust to the case of non-normally-distributed variables). The correlations are overall positive and reasonably large in magnitude; hence we conclude that they give reasonable support for H2. This is interesting considering that the corresponding relationship was weaker in the case of Croatian regions, but it is more similar to Slovenian (Coad and Srhoj, 2023) and Austrian regions (Friesenbichler and Hölzl, 2020).

Figure 8: Bivariate map of HGF shares: 2014-2019 (left) and 2014-2020 (right)



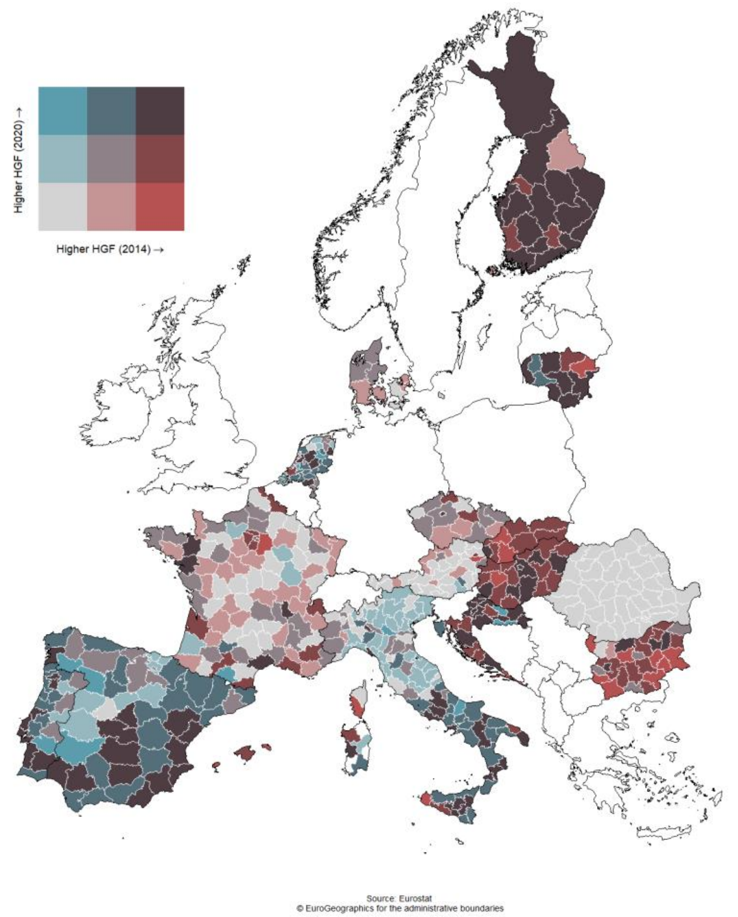


Table 8: Regional HGF shares correlations.

	HGF shares		
	(1)	(2)	(3)
	Pearson correlation [p-value]	**Spearman's rank correlation [p-value]**	**N**
Pooled	0.574 [0.000]	0.507 [0.000]	1723
2005-2008 & 2008-2011	0.564	0.510	60
2008-2011 & 2011-2014	0.781	0.721	467
2011-2014 & 2014-2017	0.633	0.538	556
2014-2017 & 2017-2020	0.757	0.678	640
2005-2008 & 2011-2014	0.627	0.670	60
2005-2008 & 2014-2017	0.695	0.689	60
2005-2008 & 2017-2020	0.657	0.753	60
2008-2011 & 2014-2017	0.665	0.533	468
2008-2011 & 2017-2020	0.712	0.578	465

Notes: All correlation estimates are significant at p-values < 0.01, but are removed for brevity.

Regression results using different lag structures are reported in Table 9. Overall, our point estimates provide support for the idea of persistence of HGF shares. These results suggest that the findings in Coad and Srhoj (2023) for the lack of persistence of regional HGF shares in Croatia might not hold for a broader sample of EU countries.

Table 9: Persistence of regional HGF shares.

	Dependent variable: *HGF share t*		
	(1)	(2)	(3)
One lag model			
HGF shares t-3	0.742*** (0.024)		
Two lag model			
HGF shares t-6		0.831*** (0.038)	
Three lag model			
HGF shares t-9			0.912*** (0.045)
N	1723	1079	523
R^2	0.570	0.483	0.501
Time FE	Yes	Yes	Yes
Regionally clustered SE	Yes	Yes	Yes

Notes: *p<0.1; **p<0.05; ***p<0.01.

5.3. Hypothesis 3: Is persistence of HGF shares higher in more developed regions and in those with higher HGF shares in the past?

We test whether regional persistence of HGF shares is higher in more developed regions using Eq. (5) (subsection 3.3). Tables 10 and 11 show non-significant or significant negative interaction terms (in bold) between previous regional HGF shares and whether region is in the middle or highest tercile of regions as measured by GDP per capita, patents per capita, R&D expenditures per capita or EEI. This indicates that regional persistence of HGF shares is not higher in more developed regions. Actually, the interaction of HGF shares and upper terciles is negative, which is evidence in favour of HGF shares persistence being higher in less developed regions.

Table 10: Regional HGFs persistence with GDP, Patents and R&D investments interactions.

	Dependent variable:		
	HGF shares (t)		
	(1)	(2)	(3)
HGF shares (t-3)	0.794^{***} (0.026)	0.771^{***} (0.030)	0.811^{***} (0.023)
GDP per capita PPP 2nd tercile	1.410^{***} (0.391)		
GDP per capita PPP 3rd tercile	2.024^{***} (0.445)		
HGF shares (t-3) X GDP per capita PPP 2nd tercile	**-0.110^{***}** **(0.040)**		
HGF shares (t-3) X GDP per capita PPP 3rd tercile	**-0.183^{***}** **(0.048)**		
Patents per capita 2nd tercile		1.860^{***} (0.420)	
Patents per capita 3rd tercile		0.445 (0.405)	
HGF shares (t-3) X Patents per capita 2nd tercile		**-0.154^{***}** **(0.042)**	
HGF shares (t-3) X Patents per capita 3rd tercile		**-0.022** **(0.044)**	
R&D investment per capita 2nd tercile			1.953^{***} (0.480)
R&D investment per capita 3rd tercile			0.742^{**} (0.365)
HGF shares (t-3) X R&D investment per capita 2nd tercile			**-0.166^{***}** **(0.052)**
HGF shares (t-3) X R&D investment per capita 3rd tercile			**-0.061** **(0.040)**

Time FE	Yes	Yes	Yes
Regionally clustered SE	Yes	Yes	Yes
Observations	1,717	1,717	1,570
R^2	0.581	0.579	0.611
Adjusted R^2	0.579	0.577	0.609

Notes: *p<0.1; **p<0.05; ***p<0.01. Reference category is always 1st (lowest) tercile.

Table 11: Regional HGFs persistence with EEI interactions.

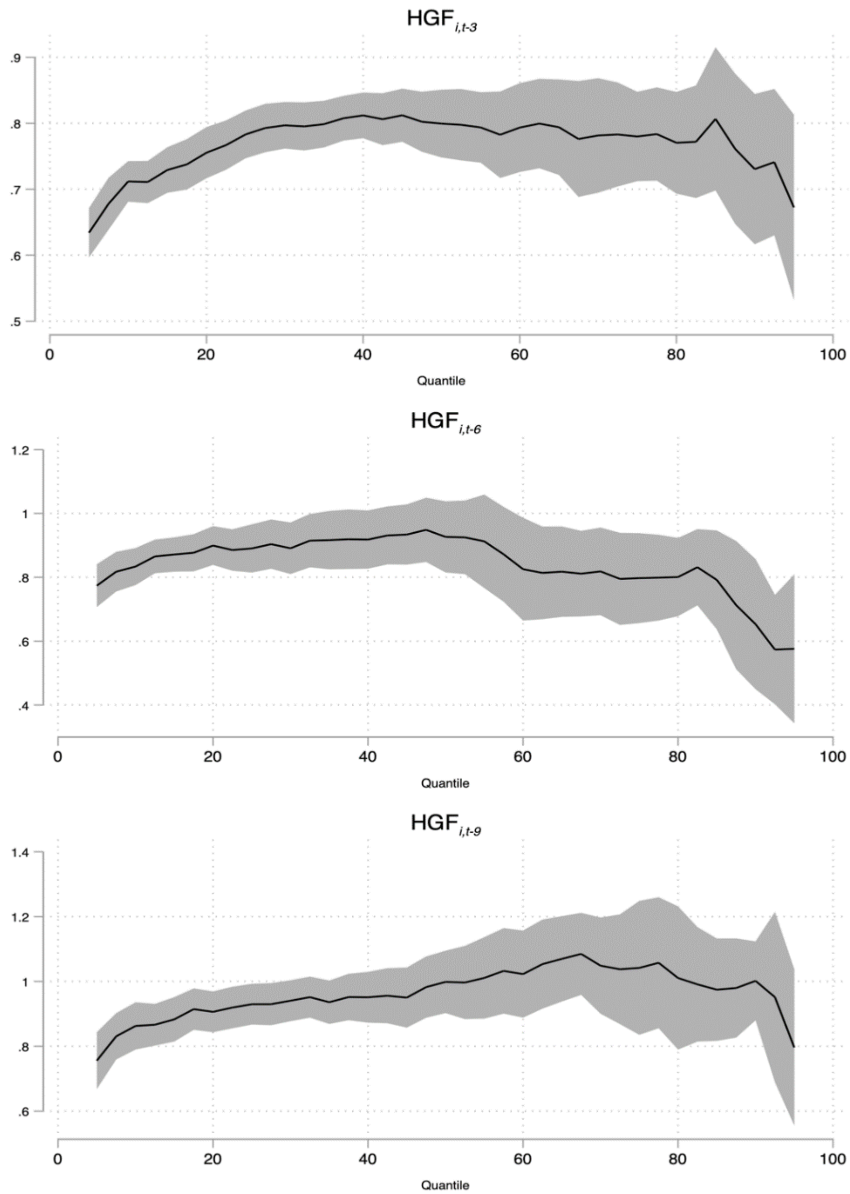
	Dependent variable:	
	HGF shares (t)	
	(1)	(2)
HGF shares (t-3)	0.797***	0.798***
	(0.029)	(0.030)
EEI Additive 2nd tercile	1.999***	
	(0.439)	
EEI Additive 3rd tercile	0.921**	
	(0.396)	
HGF shares (t-3) X EEI Additive 2nd tercile	-0.177***	
	(0.046)	
HGF shares (t-3) X EEI Additive 3rd tercile	-0.077*	
	(0.042)	
EEI Log 2nd tercile		2.054***
		(0.438)
EEI Log 3rd tercile		0.863**
		(0.398)
HGF shares (t-3) X EEI Log 2nd tercile		-0.182***
		(0.046)
HGF shares (t-3) X EEI Log 3rd tercile		-0.077*
		(0.042)
Time FE	Yes	Yes
Regionally clustered SE	Yes	Yes
Observations	1,697	1,697
R^2	0.583	0.584
Adjusted R^2	0.581	0.582

Notes: *p<0.1; **p<0.05; ***p<0.01. Reference category is always 1st (lowest) tercile.

Next, we investigate whether path-dependence can be the reason behind the observed persistence in regional HGF shares. To that end, we use quantile regressions to test whether the marginal effect of prior HGF shares on present HGF shares differs across the distribution of HGFs across regions. The presence of path-dependence would entail heterogeneity across the distribution: the relationship between past and current

HGF shares should be stronger in regions with high levels of HGFs. This would provide support for the presence of a self-reinforcing mechanism between EE inputs and outputs.

Figure 9: Quantile regression results: persistency analysis.



Notes: the plots report estimates from quantile regressions. The top figure regresses HGF shares at time t against HGF shares at time t-3; the middle figure at time t-6, and the bottom figure at time t-9. All specifications control for year fixed effects and cluster standard errors ar the NUTS-3 level. Shaded areas represent 95% confidence intervals.

The quantile regression results in Figure 9 show positive persistence, but with some degree of heterogeneity across the quantiles. Point estimates tend to resemble an inverted U-shaped relationship between current HGF shares and lagged ones. At the upper quantiles, coefficients are generally equal or smaller if compared with other quantiles, and sometimes lower than the ones obtained at bottom quantiles. This indicates that for regions with the highest HGF shares, having a large HGF shares in the previous period does not help so much to increase the HGF shares in the current period. This evidence does not provide support for the idea that the output of EEs – the HGF shares – are strongly path-dependent. If this were the case, we would expect coefficients to increase as we move to upper quantiles (see, e.g. Andersson and Koster (2011) on path-dependency of start-up rates).

5.4. Robustness

To test the sensitivity of our results, we conduct a number of robustness checks which are listed below.

5.4.1. Analysis at the NUTS-2 level

Most existing research in this strand of the literature was conducted at the NUTS-2 level, mostly due to data availability (e.g. Stam and Van de Ven, 2021; Belitski et al., 2021; Leendertse et al., 2022). NUTS-2 level regions are larger in terms of population and geographic surface, so we check whether our results change when we change the unit of observation from NUTS-3 to NUTS-2. Results in Tables C1-C4 show:

1) no support for more developed regions having higher HGF shares (H1);
2) support for regional persistence of HGF shares (H2);
3) no support for more developed regions having higher persistence of HGF shares (H3).

5.4.2. Alternative HGF shares definitions

We repeat our analysis using two alternative definitions of regional HGF shares. A drawback of the Eurostat definition of the HGF shares is that the denominator is measured in the same period as HGFs. Such a definition could potentially lead to bias, if the number of firms with 10+ employees at time t is not a perfect proxy for number of firms with 10+ employees at time t-3. We therefore recalculate the HGF shares at the NUTS-3 level and repeat the analysis for our three main hypotheses. Results in Tables D1 – D4 show:

1) no support for more developed regions having higher HGF shares (H1);
2) support for regional persistence of HGF shares (H2);
3) no support for more developed regions having higher persistence of HGF shares (H3).

Another potential drawback of the standard HGF shares definition is that the denominator with firms employing 10 or more employees may mask substantial heterogeneity in firm size distributions across regions. For instance, less developed regions might have more micro firms (i.e. firms with less than 10 employees) and fewer medium and large firms due to lower levels of technological development and productivity (Poschke, 2018).[12] We therefore recalculate the HGF shares as the regional number of HGFs in t divided by regional number of all firms in the region in period t-3, and repeat the analysis for our three main hypotheses. Results in Tables E1 – E4 show:

1) no support for more developed regions having higher HGF shares (H1);
2) support for regional persistence of HGF shares (H2), but of weaker magnitude;

[12] Another alternative indicator of HGFs "intensity" would take a region's population as the denominator. We refrain from using this measure in our empirical analysis for three main reasons. First, this HGF shares mixes together many different factors (including regional density, share of economically active individuals in the human population, share of the public sector, etc) and confounds two distinct levels of analysis: the firm-level (i.e. HGFs and firms) and the individual level (i.e. human population). Second, it is not in line with official statistics as published by Eurostat. Third, the overwhelming majority of existing studies rely on firm-based denominators, whereas population is seldom used (for an exception, see Fotopoulos (2022)).

3) weak support (mixed results) for more developed regions having higher persistence of HGF shares (H3).

6. Discussion

Table 12 summarizes the results obtained so far by using them to evaluate whether the hypotheses are supported.

Table 12: Revisiting the hypotheses

Hypothesis	Statement	Indicator	Supported?
H1	More developed regions will have higher regional HGF shares		
H1a		GDP per capita	Weak support
H1b		Patents per capita	Not supported
H1c		R&D investment per capita	Weak support
H1d		EEI	Not supported
H2	Given the regional persistence of the inputs, there is regional persistence of HGF shares		Supported
H2a		In two consecutive three-year periods	Supported
H2b		With longer lags	Supported
H3	Regional persistence of HGF shares is higher in more developed regions and in those with high shares in the past		Not supported
H3a		GDP per capita	Not supported
H3b		Patents per capita	Not supported
H3c		R&D investment per capita	Not supported
H3d		EEI	Not supported
H3e		High HGF shares in the past	Not supported

Hypothesis 1 receives weak support in some cases, while in other cases it receives no support. Figure 4 showed that support for H1 seemed to be driven by the lower end of the distribution of economic development (e.g. regions with low GDP per capita PPP). There was some support for H1a and H1c in Table 7, which seemed to be driven by the lower end of the distribution of regional HGF shares (Figure 5). Support for H1a and H1c is weaker when focusing on the NUTS-2 level instead of the NUTS-3 level (Table C1). Overall, therefore, H1 is not supported.

Hypothesis 2 receives the strongest support: there is persistence in regional HGF shares.

Hypothesis 3 investigates whether the persistence of HGF shares is higher in more developed regions. H3 is generally not supported. The results actually appear to be significantly counter to expectations (i.e. significantly negative instead of positive) in our main specification for H3 (Table 10). Similarly, we do not find support for path-dependency in HGF shares as persistence is not stronger for regions with high levels of HGFs in the past.

7. Conclusion

This paper investigated the phenomenon of regional shares of high-growth firms (HGFs), as well as their persistence. To that end, we derived some predictions from Entrepreneurial Ecosystem (EE) theory and tested three broad hypotheses using regional-level data (NUTS-3 and NUTS-2 level data) from several European countries.

According to EE theory (Stam, 2015; Spigel, 2017; Leendertse et al., 2021; Coad and Srhoj, 2023), the output of ecosystems is measured in terms of regional HGF shares. Regions whose EE is built from a stronger set of inputs can expect higher levels of outputs, and the best-performing EEs are expected to produce more outputs in terms of having higher regional shares of HGFs. On this basis, analysis of regional rankings yields the puzzling finding that the leading EEs in Europe, apparently, are in regions such as the Canary Islands, Basilicata, and Algarve. These results cast doubt that regional HGF shares are truly capturing the performance of an EE. Additional analyses confirm that regions that are highly developed, innovative or with superior quality in terms of EEs do not feature the highest incidence of HGF. However, we do find evidence that is in line with EE theory, namely, the fact that HGF shares exhibit persistence over time. Yet, we show that this persistence is not stronger for regions with higher development, nor for those that experienced high levels of HGF shares in the past, something that is not consistent with path-dependency being the main mechanism behind persistence.

These findings call for a more nuanced interpretation of regional HGF shares, including a better understanding of their nature and drivers. There is a widespread assumption among policy-makers and practitioners according to which a higher prevalence of HGFs in a given region is synonymous with vibrant regional economic dynamics. However, as shown in the paper, those regions that are considered entrepreneurial hotspots are actually the ones registering a lower incidence of HGF. Additionally, our results provide an opportunity for the refinement of EE theory. The current framework, according to which a set of inputs are converted into the EE output (i.e. regional HGF shares), does not receive strong support in our data. A prudent approach would be for policymakers to avoid investing in applying EE principles, at least until a stronger evidence base emerges regarding how the EE framework can generate the expected EE outputs in European regions. At present, there is insufficient evidence that the EE framework can effectively achieve its stated goals.

References

Andersson, M., & Koster, S. (2011). Sources of persistence in regional start-up rates—evidence from Sweden. Journal of Economic Geography, 11(1), 179-201.

Audretsch, D. B., & Belitski, M. (2021). Towards an entrepreneurial ecosystem typology for regional economic development: the role of creative class and entrepreneurship. Regional Studies, 55(4), 735-756. DOI: 10.1080/00343404.2020.1854711

Belitski, M., Korosteleva, J., & Piscitello, L. (2021). Digital affordances and entrepreneurial dynamics: New evidence from European regions. Technovation, 102442.

Benedetti Fasil, C, Domnick, C, del Rio, J-C, Fákó, P, Flachenecker, F., Gavigan, J. P., Janiri, M. L., Stamenov, B. & Testa, G., (2021). High Growth Enterprises in the COVID-19 Crisis Context demographics, environmental innovations, digitalization, finance and policy measures. EUR 30686 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-37269-1, doi:10.2760/63402, JRC124469

Borggren, J., Eriksson, R. H., & Lindgren, U. (2016). Knowledge flows in high-impact firms: How does relatedness influence survival, acquisition and exit? Journal of Economic Geography, 16(3), 637-665. https://doi.org/10.1093/jeg/lbv014

Brown R, Mason C, & Mawson S. (2014). Increasing 'The Vital 6 Percent': designing effective public policy to support high growth firms, NESTA working paper 14/01, NESTA, London. http:// www.nesta.org.uk/sites/default/files/working_paper_increasing_the_vital_6_percent.pdf

Cainelli G., Ganau R., & Giunta A., (2023). Value chain, regional institutions and firm growth in Europe. Journal of Economic Geography, forthcoming. https://doi.org/10.1093/jeg/lbad004

Cefis, E. (2003). Is there persistence in innovative activities? International Journal of Industrial Organization, 21(4), 489-515.

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. https://doi.org/10.1093/icc/dtt052

Coad, A., & Srhoj, S. (2023). Entrepreneurial Ecosystems and Regional Persistence of High Growth Firms: A 'Broken Clock' Critique. Research Policy, 52(6), 104762..

Coad, A., Harasztosi, P., Pál, R., & Teruel, M. (2022). Policy instruments for high-growth enterprises. *Questioning the Entrepreneurial State*, *273*.

Corrente, S., Greco, S., Nicotra, M., Romano, M., & Schillaci, C. E. (2019). Evaluating and comparing entrepreneurial ecosystems using SMAA and SMAA-S. Journal of Technology Transfer, 44(2), 485-519.

Cryer, J. D., & Chan, K. S. (2008). Time series analysis: with applications in R (Vol. 2). New York: Springer.

Delgado, M., Porter, M. E., & Stern, S. (2010). Clusters and entrepreneurship. Journal of Economic Geography, 10, 495–518. https://doi.org/10.1093/jeg/lbq010

Du, J., & Vanino, E. (2021). Agglomeration externalities of fast-growth firms. Regional studies, 55(2), 167-181.

Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. Annual Review of Sociology, 40, 31-53.

Eurostat - OECD (2007). Eurostat-OECD Manual on Business Demography Statistics. Office for Official Publications of the European Communities: Luxembourg.

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

Fotopoulos, G. (2022). Knowledge Spillovers, Entrepreneurial Ecosystems and the Geography of High Growth Firms. Entrepreneurship Theory and Practice, 10422587221111732.

Friesenbichler, K., & Hölzl, W. (2020). High-growth firm shares in Austrian regions: The role of economic structures. *Regional Studies*, *54*(11), 1585-1595.

Fritsch, M., & Mueller, P. (2007). The persistence of regional new business formation-activity over time–assessing the potential of policy promotion programs. Journal of Evolutionary Economics, 17, 299-315.

Fritsch M., & Wyrwich M., (2014) The Long Persistence of Regional Levels of Entrepreneurship: Germany, 1925–2005, Regional Studies, 48:6, 955-973, DOI: 10.1080/00343404.2013.816414

Gonzalez-Uribe, J., & Leatherbee, M. (2018). The effects of business accelerators on venture performance: Evidence from start-up Chile. Review of Financial Studies, 31(4), 1566-1603.

González-Uribe, J., & Reyes, S. (2021). Identifying and boosting "Gazelles": Evidence from business accelerators. Journal of Financial Economics, 139(1), 260-287.

Haltiwanger, J., Jarmin, R. S., Kulick, R., & Miranda, J. (2016). High growth young firms: contribution to job, output, and productivity growth. In *Measuring entrepreneurial businesses: current knowledge and challenges* (pp. 11-62). University of Chicago Press.

Harima, J. (2020). Public Accelerators in Entrepreneurial Ecosystems. Springer Fachmedien Wiesbaden.

Henrekson, M., & Johansson, D. (1999). Institutional effects on the evolution of the size distribution of firms. Small Business Economics, 12(1), 11-23. https://link.springer.com/article/10.1023/A%3A1008002330051

Hundt, C., & Sternberg, R. (2016). Explaining new firm creation in Europe from a spatial and time perspective: A multilevel analysis based upon data of individuals, regions and countries. Papers in Regional Science, 95(2), 223-257.

Isenberg, D. (2014). What an entrepreneurship ecosystem actually is. Harvard Business Review, 5(1), 7.

Kahneman, D. (2011). Thinking, fast and slow. Macmillan.

Leendertse, J., Schrijvers, M., & Stam, E. (2021). Measure twice, cut once: Entrepreneurial ecosystem metrics. Research Policy, 104336.

Mikic, M., Horvatinovic, T., & Kovac, I. (2021). Climbing up the regional intellectual capital tree: an EU entrepreneurial ecosystem analysis. Journal of Intellectual Capital, 22(6), 1030-1054.

OECD. (2010) High-Growth Enterprises: What Governments Can do to Make a Difference. OECD Studies on SMEs and Entrepreneurship. Paris: OECD.

OECD. (2013). An International Benchmarking Analysis of Public Programmes for High-Growth Firms. March 2013. Final report prepared by the OECD Local Economic and Employment Development Programme in collaboration with the Danish Business Authority.

OECD (2021a), Understanding Firm Growth: Helping SMEs Scale Up, OECD Studies on SMEs and Entrepreneurship, OECD Publishing, Paris, https://doi.org/10.1787/fc60b04c-en

OECD (2021b), OECD Economic Surveys: Austria 2021, OECD Publishing, Paris, https://doi.org/10.1787/eaf9ec79-en.

OECD (2022), OECD Economic Surveys: Romania 2022, OECD Publishing, Paris, https://doi.org/10.1787/e2174606-en.

Parente, P. M., & Santos Silva, J. M. (2016). Quantile regression with clustered data. *Journal of Econometric Methods*, *5*(1), 1-15.

Poschke, M. (2018). The firm size distribution across countries and skill-biased change in entrepreneurial technology. American Economic Journal: Macroeconomics, 10(3), 1-41. https://www.aeaweb.org/articles?id=10.1257/mac.20140181

Raymond, W., Mohnen, P., Palm, F., & Van Der Loeff, S. S. (2010). Persistence of Innovation in Dutch Manufacturing: is it spurious? Review of Economics and Statistics, 92 (3), 495-504.

Sleuwaegen L., & Ramboer S., (2020). Regional competitiveness and high growth firms in the EU: the creativity premium, Applied Economics, 52:22, 2325-2338, DOI:10.1080/00036846.2019.1686454

Spigel, B. (2017). The relational organization of entrepreneurial ecosystems. Entrepreneurship Theory and Practice, 41(1), 49-72.

Spigel, B. (2022). Examining the cohesiveness and nestedness entrepreneurial ecosystems: evidence from British FinTechs. Small Business Economics, 1-19. https://doi.org/10.1007/s11187-021-00589-z

Spigel, B., & Harrison, R. (2018). Toward a process theory of entrepreneurial ecosystems. Strategic Entrepreneurship Journal, 12(1), 151-168.

Spigel, B., Kitagawa, F., & Mason, C. (2020). A manifesto for researching entrepreneurial ecosystems. Local Economy, 35(5), 482-495.

Stam, E. (2015). Entrepreneurial ecosystems and regional policy: a sympathetic critique. European Planning Studies, 23(9), 1759-1769.

Stam, E., & Bosma, N. (2015). Local policies for high-growth firms. *The Oxford handbook of local competitiveness*, 286-305.

Stam, E., & Van de Ven, A. (2021). Entrepreneurial ecosystem elements. Small Business Economics, 56(2), 809-832.

Sternberg, R., von Bloh, J., & Coduras, A. (2019). A new framework to measure entrepreneurial ecosystems at the regional level. Zeitschrift für Wirtschaftsgeographie, 63(2-4), 103-117.

Wurth, B., Stam, E., & Spigel, B. (2021). Toward an entrepreneurial ecosystem research program. Entrepreneurship Theory and Practice, 1042258721998948.

Appendices

This section contains the appendices to the paper “Regional incidence and persistence of high-growth firms: Testing ideas from the Entrepreneurial Ecosystems literature” by Alex Coad, Clemens Domnick, Pietro Santoleri and Stjepan Srhoj.

The appendices contain the following information:

- **Appendix A**: provides a summary of statements taken from the the EE literature concerning regional HGF shares incidence and persistence;
- **Appendix B**: motivates the reason why growth periods should not be overlapping;
- **Appendix C**: reports additional results and re-runs our main analysis using NUTS-2 regions instead of NUTS-3 regions;
- **Appendix D**: reports results using as dependent variable the number of HGFs at time t divided by the number of active firms with ≥ 10 employees at time t-3;
- **Appendix E**: reports results using as dependent variable the number of HGFs at time t divided by the number of all active firms at time t-3;

Appendix A: Statements by EE scholars regarding HGFs

HGFs as the output of the Entrepreneurial Ecosystem	
Source	Statement
Harima (2020, p. 30)	"The idea of entrepreneurial ecosystems, however, focuses on the entrepreneurial process within regional boundaries, with all entrepreneurs serving as the focal actors who drive the ecosystem evolution, and establishing high-growth firms as the primal output of entrepreneurial ecosystems (Spigel, 2017)."
Spigel, Kitagawa and Mason (2020, p. 484)	"we define entrepreneurial ecosystems here as the regional collection of actors (such as entrepreneurs, advisors, workers, mentors, and workers) and factors (cultural outlooks, policies, R&D systems, and networks) that all contribute to the creation and survival of high-growth ventures. We focus on high-growth entrepreneurship because it is seen as a major driver of job creation and economic growth in both advanced and emerging economies"
Spigel, Kitagawa and Mason (2020, p. 485)	"'high-growth' firms – the firms that ecosystems are, in principle, designed to support."
Sleuwaegen and Ramboer (2020, p. 2326)	"we develop a conceptual model explaining the link between a virtuous entrepreneurial ecosystem and the emergence of HGFs"
Stam and Van de Ven (2021 p. 817)	"the envisaged output of the ecosystem: high-growth firms. … we have proxied productive entrepreneurship with the prevalence of high-growth firms … the share of high-growth firms of the regional business population"
Stam and Van de Ven (2021, p. 809)	"We find that the prevalence of high-growth firms in a region is strongly related to the quality of its entrepreneurial ecosystem."
Audretsch and Belitski (2021, p. 738)	"Regions with a sizeable creative industry where the creative class works are more likely to display openness to diversity and new ideas generation that spillover into new marketable products leading to high-growth firms (Audretsch et al., 2019a)."
Audretsch and Belitski (2021, p. 739)	"In measuring productive entrepreneurship (Baumol, 1993), we used the share of high-growth firms in a region (Stam, 2018; Stam et al., 2011, 2012) (Table 1)."
Wurth, Stam, & Spigel, (2021, p. 7)	"one of the defining features of entrepreneurial ecosystems research has been a focus on productive entrepreneurship. ... It is often measured as high-growth entrepreneurship"
Spigel (2022, p. 3)	"Ecosystems differ from other territorial-based theories of economic development such as clusters and innovation systems due to their focus on the types of regional environment that impact high-growth entrepreneurs"
High-performing Entrepreneurial Ecosystems should display <u>persistence</u> in HGF shares	
Spigel (2017, p. 49)	"Entrepreneurial ecosystems have emerged as a popular concept to explain the persistence of high-growth entrepreneurship within regions."
Spigel and Harrison (2018, p. 155)	"Cluster and RIS [Regional Innovation System] concepts provide well-researched frameworks that help us understand why some places enjoy persistently higher rates of high-growth entrepreneurship than others."
Entrepreneurial Ecosystems are path-dependent	
Wurth, Stam, & Spigel, (2021, p. 764)	"strong path dependence in the evolution of entrepreneurial ecosystems. EE should be treated as a system (strong path-dependency within its evolution), with overall quality positively related to entrepreneurial output, which in turn feeds back into the regional EE" (p. 764).

Source: based on Coad and Srhoj (2023).

Appendix B: Why the growth periods must be non-overlapping

This appendix is based on Coad and Srhoj (2023).

The periods over which HGF shares are measured need to be non-overlapping, otherwise we will introduce severe positive bias in our estimates of persistence.

Consider the regional share of HGFs, denoted by Y_{it} for region i and year t. Y_{it} refers to the three-year growth performance of firms, hence $Y_{it} = f\left(gr_{ijt}, gr_{ijt-1}, gr_{ijt-2}\right)$, which shows that regional HGF shares depends on the growth of firm j over a three-year period (growth gr at time t, t-1, and t-2). For didactic purposes, we assume that annual growth of firm j, gr_{ijt} is a purely random variable $(\mu = 0, \sigma^2 = 1)$ that is independent and identically distributed over years:

$$\gamma = cor\left(gr_{ijt}, gr_{imt}\right) = 1 \, for \; j = m,$$

and

$$\gamma = cor\left(gr_{ijt}, gr_{imt}\right) = 0 \, for \; j \neq m,$$

For non-overlapping periods, we have zero persistence (i.e. zero correlation of regional HGF shares). Considering the covariance:

$$\gamma = cov(Y_{it}, Y_{it-3}) = cov\left[\left(f\left(gr_{ijt}, gr_{ijt-1}, gr_{ijt-2}\right), f\left(gr_{ijt-3}, gr_{ijt-4}, gr_{ijt-5}\right)\right)\right]$$

$= 0$, given that

$$\gamma = cov\left(gr_{ijt}, gr_{imt}\right) = 0 \, for \; j \neq m,$$

For overlapping periods, however, we introduce positive bias in the persistence:

$$\gamma = cov(Y_{it}, Y_{it-1}) = cov\left[\left(f\left(gr_{ijt}, gr_{ijt-1}, gr_{ijt-2}\right), f\left(gr_{ijt-1}, gr_{ijt-2}, gr_{ijt-3}\right)\right)\right]$$

This positive bias comes from the fact that the two terms $f\left(gr_{ijt}, gr_{ijt-1}, gr_{ijt-2}\right)$ and $f\left(gr_{ijt-1}, gr_{ijt-2}, gr_{ijt-3}\right)$ are positively related because they include the same terms gr_{ijt-1} and gr_{ijt-2}. The same growth events (i.e. gr_{ijt-1} and gr_{ijt-2}) would be double-counted, because they are included in both three-year periods.

Then the use of overlapping periods increases the value of $cor(Y_{it}, Y_{it-3})$ towards 1 (i.e. positive bias that exaggerates persistence in regional HGF shares).

Consider the special case where $f\left(gr_{ijt}, gr_{ijt-1}, gr_{ijt-2}\right) = gr_{ijt} + gr_{ijt-1} + gr_{ijt-2}$

In the case of non-overlapping periods, we have:

$cov(Y_{it}, Y_{it-3}) = cov\left(\left(gr_{ijt} + gr_{ijt-1} + gr_{ijt-2}\right), \left(gr_{ijt-3} + gr_{ijt-4} + gr_{ijt-5}\right)\right) = 0$

given that

$$\gamma = cov(gr_{ijt}, gr_{imt}) = 0 \; for \; j \neq m$$

However, in the case of overlapping periods (Cryer and Chan, 2008, Section 4.2),

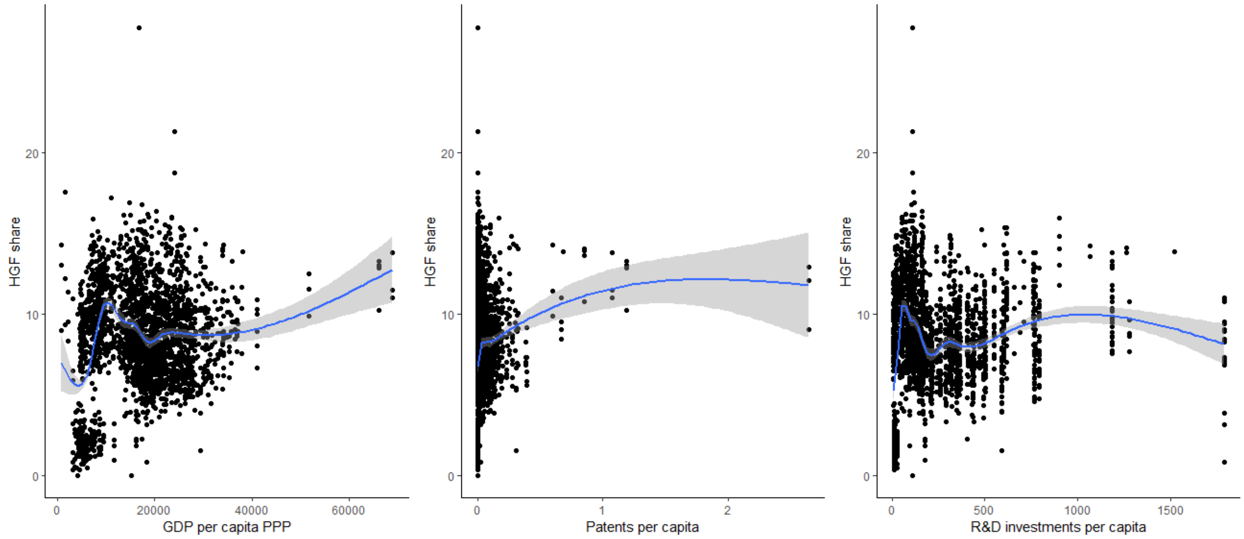
$$\begin{aligned} cov(Y_{it}, Y_{it-1}) = \; & cov\left((gr_{ijt} + gr_{ijt-1} + gr_{ijt-2}), (gr_{ijt-1} + gr_{ijt-2} + gr_{ijt-3})\right) \\ & = cov\left((gr_{ijt}, gr_{ijt-3})\right) + cov\left((gr_{ijt-1}, gr_{ijt-1})\right) + cov\left((gr_{ijt-2}, gr_{ijt-2})\right) \\ & = 0 + 1 + 1 \end{aligned}$$

The corresponding correlation coefficient can be calculated recalling that $\sigma^2 = 1$, to yield: $\frac{0+1+1}{1+1+1} = 0.6667$

Use of overlapping periods therefore leads, in this case, to an estimated persistence of 66.7%, whereas in fact we know that it should be 0% (given that gr_{ijt} was set up to be independent from one year to the next).

This might explain why some studies have found high persistence in regional HGF shares when using overlapping periods. For example, Fotopoulos (2022) uses a sample has 378 regions over a 6-year period (2012-2017), and N=2268 in Table 2. Considering that 2268/378 = 6, this suggests that the three-year periods are overlapping (e.g. the growth rates for the first time period overlap with those for the second time period), which would give a strong positive bias to the persistence of HGF shares because it is likely that the same growth events will be double-counted.

Figure A1: Scatterplot HGF shares and regional development indicators. Absolute values of GDP per capita, Patents per capita, and R&D investments per capita.



Notes: Results at NUTS-3 level. Loess function applied.

Appendix C: Additional results at the NUTS-2 level

In this section we replicate our baseline analysis using data at the NUTS-2 level.

Table C1: HGF shares and regional economic development (NUTS-2)

	Dependent variable:				
	HGF shares				
	(1)	(2)	(3)	(4)	(5)
GDP per capita PPP 2nd tercile	0.036 (0.323)				
GDP per capita PPP 3rd tercile	-0.590 (0.479)				
Patents per capita 2nd tercile		0.724 (0.470)			
Patents per capita 3rd tercile		0.079 (0.481)			
R&D investment per capita 2nd tercile			0.375 (0.490)		
R&D investment per capita 3rd tercile			0.037 (0.500)		
EEI Additive 2nd tercile				-0.063 (0.471)	
EEI Additive 3rd tercile				0.455 (0.489)	
EEI Log 2nd tercile					-0.127 (0.470)
EEI Log 3rd tercile					0.315 (0.489)
Mean	8.98	8.98	8.98	8.98	8.98
Observations	604	604	556	576	576
R^2	0.254	0.257	0.253	0.255	0.253
Adjusted R^2	0.247	0.250	0.245	0.247	0.245

*Notes: *p<0.1; **p<0.05; ***p<0.01.* Reference category is always 1st (lowest) tercile. Results at the NUTS2-level. Main analysis HGF shares definition used. Time fixed effects included, standard errors are regionally clustered.

Table C2: Regression results (NUTS-2)

Dependent variable: ***HGF shares (t)***				
	(1)	(2)	(3)	N
One lag model				
HGF shares t-3	0.511*** (0.039)	0.511*** (0.058)	0.781*** (0.043)	432
$\mathbf{R^2}$	0.285	0.285	0.651	
Two lag model				
HGF shares t-6	0.617*** (0.067)	0.617*** (0.100)	0.790*** (0.094)	270
$\mathbf{R^2}$	0.238	0.238	0.496	
Three lag model				
HGF shares t-9	0.887*** (0.072)	0.887*** (0.071)	0.917*** (0.081)	131
$\mathbf{R^2}$	0.540	0.540	0.543	
Time FE	No	No	Yes	
Regionally clustered SE	No	Yes	Yes	

Notes: *p<0.1; **p<0.05; ***p<0.01. Results at the NUTS2-level. Main analysis HGF shares definition used.

Table C3. Regression results at NUTS-2 level. Interactions with GDP, Patents and R&D investments.

	Dependent variable:		
	HGF shares (t)		
	(1)	(2)	(3)
HGF shares (t-3)	0.808^{***} (0.048)	0.784^{***} (0.070)	0.829^{***} (0.049)
GDP per capita PPP 2nd tercile	1.092 (0.819)		
GDP per capita PPP 3rd tercile	1.154 (0.734)		
HGF shares (t-3) X GDP per capita PPP 2nd tercile	**-0.063 (0.083)**		
HGF shares (t-3) X GDP per capita PPP 3rd tercile	**-0.078 (0.076)**		
Patents per capita 2nd tercile		2.231*** (0.775)	
Patents per capita 3rd tercile		-0.523 (0.697)	
HGF shares (t-3) X Patents per capita 2nd tercile		**-0.185** (0.075)**	
HGF shares **(t-3) X Patents per capita 3rd tercile**		**0.092 (0.072)**	
R&D investment per capita 2nd tercile			-0.456 (0.773)
R&D investment per capita 3rd tercile			0.900 (0.874)
HGF shares (t-3) X R&D investment per capita 2nd tercile			**0.077 (0.077)**
HGF shares (t-3) X R&D investment per capita 3rd tercile			**-0.073 (0.090)**
Observations	433	433	400
R^2	0.660	0.667	0.676
Adjusted R^2	0.653	0.661	0.669

*Notes: *p<0.1; **p<0.05; ***p<0.01.* Reference category is always 1st (lowest) tercile. Results at the NUTS2-level. Main analysis HGF shares definition used. Time fixed effects included, standard errors are regionally clustered.

Table C4. Regression results at NUTS-2 level. Interactions with EEI.

	Dependent variable:	
	HGF shares (t)	
	(1)	(2)
HGF shares (t-3)	0.821*** (0.055)	0.826*** (0.054)
EEI Additive 2nd tercile	1.084 (0.833)	
EEI Additive 3rd tercile	-0.023 (0.753)	
HGF shares (t-3) X EEI Additive 2nd tercile	**-0.091 (0.084)**	
HGF shares (t-3) X EEI Additive 3rd tercile	**0.041 (0.074)**	
EEI Log 2nd tercile		0.956 (0.842)
EEI Log 3rd tercile		-0.081 (0.759)
HGF shares (t-3) X EEI Log 2nd tercile		**-0.078 (0.085)**
HGF shares (t-3) X EEI Log 3rd tercile		**0.045 (0.075)**
Observations	413	413
R^2	0.680	0.679
Adjusted R^2	0.674	0.673

*Notes: *p<0.1; **p<0.05; ***p<0.01.* Reference category is always 1st (lowest) tercile. Results at the NUTS2-level. Main analysis HGF shares definition used. Time fixed effects included, standard errors are regionally clustered.

Appendix D. HGF shares denominator: firms with ≥ 10 employees at time t-3

If HGFs are measured over the time period t:t-3, then firms with ≥ 10 employees should be measured in the initial period, t-3. If we were to measure the HGF shares in terms of HGFs (at time t), divided by firms with ≥ 10 employees (at time t), this might lead to selection bias by dropping non-surviving firms. Number of firms with ≥ 10 employees might be stable over time, but there could also be an argument of potential HGFs being "high-risk" firm types that have higher chances of exit. In what follows, we report results using using as denominator the number of firms with ≥ 10 employees at time t-3.

Table D1. Linear regression - HGFs (divided by lagged firms >= 10) and regional economic development

	Dependent variable:				
	HGF shares				
	(1)	(2)	(3)	(4)	(5)
GDP per capita PPP 2nd tercile	0.008** (0.004)				
GDP per capita PPP 3rd tercile	0.006 (0.004)				
Patents per capita 2nd tercile		0.011*** (0.004)			
Patents per capita 3rd tercile		0.005 (0.004)			
R&D investment per capita 2nd tercile			0.011*** (0.003)		
R&D investment per capita 3rd tercile			0.009*** (0.003)		
EEI Additive 2nd tercile				0.004 (0.003)	
EEI Additive 3rd tercile				0.004 (0.003)	
EEI Log 2nd tercile					0.006** (0.003)
EEI Log 3rd tercile					0.004 (0.003)
Mean	0.09	0.09	0.09	0.09	0.09
Observations	2,154	2,154	1,966	2,133	2,133
R^2	0.189	0.194	0.269	0.245	0.247
Adjusted R^2	0.187	0.192	0.267	0.244	0.246

*Notes: *p<0.1; **p<0.05; ***p<0.01.* Reference category is always 1st (lowest) tercile. Results at the NUTS3-level. HGF shares definition computed from raw Eurostat data on the number of HGFs in a region, and on number of firms with 10 or more employees in a region, where lagged (t-3) number of firms in a region is used as a denominator. Time fixed effects included, standard errors are regionally clustered.

Table D2: Regression results. Observations pooled across years. HGFs divided by lagged firms >= 10.

Dependent variable: ***HGF shares (t)***				
	(1)	(2)	(3)	N
One lag model				
HGF shares t-3	0.516*** (0.020)	0.516*** (0.047)	0.667*** (0.021)	1507
R^2	0.317	0.317	0.568	
Two lag model				
HGF shares t-6	0.645*** (0.044)	0.645*** (0.058)	0.721*** (0.053)	948
R^2	0.182	0.182	0.315	
Three lag model				
HGF shares t-9	0.784*** (0.052)	0.784*** (0.055)	0.784*** (0.055)	448
R^2	0.335	0.335	0.335	
Time FE	No	No	Yes	
Regionally clustered SE	No	Yes	Yes	

Notes: *p<0.1; **p<0.05; ***p<0.01. Results at the NUTS3-level. HGF shares definition computed from raw Eurostat data on the number of HGFs in a region, and on number of firms with 10 or more employees in a region, where lagged (t-3) number of firms in a region is used as a denominator. Time fixed effects included, standard errors are regionally clustered.

Table D3: Regression results. Observations pooled across years. Persistency with GDP, Patents and R&D investments interactions.

	Dependent variable:		
	HGF shares (t)		
	(1)	(2)	(3)
HGF shares (t-3)	0.735*** (0.028)	0.703*** (0.023)	0.742*** (0.033)
GDP per capita PPP 2nd tercile	0.022*** (0.004)		
GDP per capita PPP 3rd tercile	0.030*** (0.005)		
HGF shares (t-3) X GDP per capita PPP 2nd tercile	**-0.184***** **(0.044)**		
HGF shares (t-3) X GDP per capita PPP 3rd tercile	**-0.288***** **(0.055)**		

Patents per capita 2nd tercile		0.026***	
		(0.005)	
Patents per capita 3rd tercile		0.005	
		(0.005)	
HGF shares (t-3) X Patents per capita 2nd tercile		**-0.213*****	
		(0.042)	
HGF shares (t-3) X Patents per capita 3rd tercile		**-0.044**	
		(0.053)	
R&D investment per capita 2nd tercile			0.027***
			(0.006)
R&D investment per capita 3rd tercile			0.002
			(0.005)
HGF shares (t-3) X R&D investment per capita 2nd tercile			**-0.240*****
			(0.057)
HGF shares (t-3) X R&D investment per capita 3rd tercile			**-0.009**
			(0.054)
Observations	1,499	1,499	1,377
R^2	0.588	0.583	0.531
Adjusted R^2	0.586	0.581	0.529

*Notes: *p<0.1; **p<0.05; ***p<0.01.* Reference category is always 1st (lowest) tercile. Results at the NUTS3-level. HGF shares definition computed from raw Eurostat data on the number of HGFs in a region, and on number of firms with 10 or more employees in a region, where lagged (t-3) number of firms in a region is used as a denominator. Time fixed effects included, standard errors are regionally clustered.

Table D4: Regression results. Observations pooled across years. Persistency with EEI interactions.

	Dependent variable:	
	HGF shares (t)	
	(1)	(2)
HGF shares (t-3)	0.742***	0.739***
	(0.044)	(0.044)
EEI Additive 2nd tercile	0.025***	
	(0.005)	
EEI Additive 3rd tercile	0.008***	
	(0.006)	
HGF shares (t-3) X EEI Additive 2nd tercile	**-0.252*****	
	(0.057)	
HGF shares (t-3) X EEI Additive 3rd tercile	**-0.099**	
	(0.062)	
EEI Log 2nd tercile		0.026***

		(0.016)
EEI Log 3rd tercile		0.006
		(0.006)
HGF shares (t-3) X EEI Log 2nd tercile		**-0.249*****[***]
		(0.057)
HGF shares (t-3) X EEI Log 3rd tercile		**-0.087**
		(0.063)
Observations	1,486	1,486
R^2	0.509	0.511
Adjusted R^2	0.506	0.508

*Notes: *p<0.1; **p<0.05; ***p<0.01.* Reference category is always 1st (lowest) tercile. Results at the NUTS3-level. HGF shares definition computed from raw Eurostat data on the number of HGFs in a region, and on number of firms with 10 or more employees in a region, where lagged (t-3) number of firms in a region is used as a denominator. Time fixed effects included, standard errors are regionally clustered.

Appendix E. HGF shares denominator: all firms in the region at time t-3

A potential drawback of the baseline HGFs indicator is that firms with fewer than 10 employees are not taken into consideration. An alternative indicator could therefore be the following:

$$\frac{HGFs}{Firms} = \frac{Firms\ with\ \geq 10\ empl}{Firms} \cdot \frac{HGFs}{Firms\ with\ \geq 10\ empl} \tag{4}$$

This indicator arguably seeks to describe the overall dynamism of an economy, by showing the frequency of HGFs as a proportion of all firms. However, the inclusion of the first right-hand-side component, i.e.

$$\frac{Firms\ with\ \geq 10\ empl}{Firms}$$

is potentially problematic because firms with fewer than 10 employees cannot possibly become HGFs and therefore should not be taken into consideration as the sample from which HGFs emerge. This indicator penalizes regions with lots of firms with <10 employees. Also, it could be considered to be "unfair", because we are looking at the ratio of HGFs with firms that could not possibly become HGFs even if they wanted to. Compared to the indicator in 3.1.2.1, this indicator would be akin to the following ratio:

$$\frac{lottery\ winners}{Everyone\ including\ those\ that\ did\ not\ buy\ and\ were\ not\ even\ eligible\ to\ buy\ a\ lottery\ ticket}$$

A potential problem here is the denominator is not such a relevant source group for the numerator. Nevertheless, in what follows, we re-run our analysis using this HGF shares definition.

Table E1. Linear regression - HGFs (divided by lagged all firms) and regional economic development

	Dependent variable:				
	HGF shares (t)				
	(1)	(2)	(3)	(4)	(5)
GDP per capita PPP 2nd tercile	-0.001*** (0.0003)				
GDP per capita PPP 3rd tercile	-0.001*** (0.0003)				
Patents per capita 2nd tercile		-0.001*** (0.0003)			
Patents per capita 3rd tercile		-0.0002			

		(0.0003)			
R&D investment per capita 2nd tercile			-0.0004^{**} (0.0002)		
R&D investment per capita 3rd tercile			0.0004^{*} (0.0002)		
EEI Additive 2nd tercile				-0.001^{**} (0.0003)	
EEI Additive 3rd tercile				0.0002 (0.0002)	
EEI Log 2nd tercile					-0.001^{***} (0.0003)
EEI Log 3rd tercile					0.0001 (0.0003)
Mean	0.004	0.004	0.004	0.004	0.004
Observations	2,156	2,156	1,966	2,133	2,133
R^2	0.077	0.068	0.086	0.061	0.065
Adjusted R^2	0.075	0.066	0.084	0.059	0.063
Residual Std. Error	0.003 (df = 2150)	0.003 (df = 2150)	0.002 (df = 1960)	0.003 (df = 2127)	0.003 (df = 2127)

Notes: *p<0.1; **p<0.05; ***p<0.01. Reference category is always 1st (lowest) tercile. Results at the NUTS3-level. HGF shares definition computed from raw Eurostat data on the number of HGFs in a region as numerator, and on lagged number of all firms in a region as a denominator. Time fixed effects included, standard errors are regionally clustered.

Table E2: Regression results. HGFs divided by all firms.

	Dependent variable: ***HGF shares (t)***			
	(1)	(2)	(3)	N
One lag model				
HGF shares t-3	0.684*** (0.015)	0.684*** (0.032)	0.716*** (0.039)	1509
R^2	0.588	0.588	0.687	
Two lag model				
HGF shares t-6	0.434*** (0.022)	0.434*** (0.068)	0.446*** (0.067)	950
R^2	0.287	0.287	0.310	
Three lag model				
HGF shares t-9	0.494*** (0.033)	0.494*** (0.038)	0.494*** (0.038)	448
R^2	0.333	0.333	0.333	
Time FE	No	No	Yes	
Regionally clustered SE	No	Yes	Yes	

Notes: *p<0.1; **p<0.05; ***p<0.01. Results at the NUTS3-level. HGF shares definition computed from raw Eurostat data on the number of HGFs in a region as numerator, and on lagged number of all firms in a region as a denominator. Time fixed effects included, standard errors are regionally clustered.

Table E3: Persistency with GDP, Patents and R&D investments interactions.

	Dependent variable:		
	HGF shares (t)		
	(1)	(2)	(3)
HGF shares (t-3)	0.696*** (0.046)	0.671*** (0.043)	0.762*** (0.019)
GDP per capita PPP 2nd tercile	-0.0004 (0.0003)		
GDP per capita PPP 3rd tercile	-0.0003 (0.0003)		
HGF shares (t-3) X GDP per capita PPP 2nd tercile	**0.054 (0.065)**		
HGF shares (t-3) X GDP per capita PPP 3rd tercile	**0.086 (0.062)**		
Patents per capita 2nd tercile		-0.001*** (0.0002)	
Patents per capita 3rd tercile		-0.001*** (0.0002)	
HGF shares (t-3) X Patents per capita 2nd tercile		**0.170*** (0.062)**	
HGF shares (t-3) X Patents per capita 3rd tercile		**0.165*** (0.055)**	
R&D investment per capita 2nd tercile			0.0003** (0.0002)
R&D investment per capita 3rd tercile			-0.0003* (0.0002)
HGF shares (t-3) X R&D investment per capita 2nd tercile			**-0.046 (0.043)**
HGF shares (t-3) X R&D investment per capita 3rd tercile			**0.103** (0.043)**
Observations	1,501	1,501	1,377
R^2	0.689	0.694	0.704
Adjusted R^2	0.688	0.693	0.702

Notes: *p<0.1; **p<0.05; ***p<0.01. Reference category is always 1st (lowest) tercile. Results at the NUTS3-level. HGF shares definition computed from raw Eurostat data on the number of HGFs in a region as numerator, and on lagged number of all firms in a region as a denominator. Time fixed effects included, standard errors are regionally clustered.

Table E4: Persistency with EEI interactions.

	Dependent variable:	
	HGF shares (t)	
	(1)	(2)
HGF shares (t-3)	0.930***	0.929***
	(0.022)	(0.022)
EEI Additive 2nd tercile	0.001***	
	(0.0001)	
EEI Additive 3rd tercile	0.0004*	
	(0.0002)	
HGF shares (t-3) X EEI Additive 2nd tercile	**-0.428*****	
	(0.031)	
HGF shares (t-3) X EEI Additive 3rd tercile	**-0.114****	
	(0.046)	
EEI Log 2nd tercile		0.001***
		(0.0001)
EEI Log 3rd tercile		0.0003
		(0.0002)
HGF shares (t-3) X EEI Log 2nd tercile		**-0.427*****
		(0.031)
HGF shares (t-3) X EEI Log 3rd tercile		**-0.096****
		(0.047)
Observations	1,486	1,486
R^2	0.743	0.742
Adjusted R^2	0.741	0.740

*Notes: *p<0.1; **p<0.05; ***p<0.01.* Reference category is always 1st (lowest) tercile. Results at the NUTS3-level. HGF shares definition computed from raw Eurostat data on the number of HGFs in a region as numerator, and on lagged number of all firms in a region as a denominator. Time fixed effects included, standard errors are regionally clustered.

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub

ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub