Entrepreneurial Ecosystems and Regional Persistence of High Growth Firms: A 'Broken Clock' Critique

Coad, Alex and Srhoj, Stjepan

Waseda University, University of Dubrovnik

7 December 2021

Entrepreneurial Ecosystems and Regional Persistence of High Growth Firms: A 'Broken Clock' Critique

Alex Coad
Professor
Waseda Business School
Waseda University
Tokyo, Japan

Stjepan Srhoj
Assistant Professor
Department of Economics and Business
University of Dubrovnik
Dubrovnik, Croatia

This version: 7 December 2021

ABSTRACT:

The Entrepreneurial Ecosystems (EE) approach makes specific predictions regarding how EE inputs are converted into high-growth firms (HGFs) as an output. A simulation model draws out our hypothesis of regional persistence in HGF shares. Based on intuitions that EEs are persistent, we investigate whether regional HGF shares are persistent, using census data for 2 European countries taken separately (Croatia for 2004-2019, and Slovenia for 2008-2014). Overall, there is no clear persistence in regional HGF shares - regions with large HGF shares in one period are not necessarily likely to have large HGF shares in the following period. This is a puzzle for EE theory. In fact, there seems to be more persistence in industry-level HGF shares than for regional HGF shares. We formulate a 'broken clock' critique - just as a broken clock is correct twice a day, EE recommendations may sometimes be correct, but are fundamentally flawed as long as time-changing outcomes (HGF shares) are predicted using time-invariant variables (such as local universities, institutions and infrastructure).

KEYWORDS: High-Growth Firms, Persistence, Regional Persistence; Entrepreneurial Ecosystems; Clusters; Sectoral Systems of Innovation
JEL CODES: L25

ACKNOWLEDGEMENTS: We are grateful to Ben Spigel for helpful comments. The usual disclaimer applies. This work was supported by the Croatian Science Foundation under the project IP-CORONA-2020-12-1064.

1. Introduction

High Growth Firms (HGFs) make a huge contribution to job creation, innovation, and economic dynamism. As a result, they have received lots of attention from academics (Henrekson and Johansson, 2010; Coad et al., 2014; Demir et al., 2017) and also policy makers (Grover Goswami et al., 2019; Flachenecker et al., 2020). In particular, HGFs are considered to be the output of the Entrepreneurial Ecosystem (EE) approach for entrepreneurship and innovation policy (Stam, 2015; Spigel, 2017; Leendertse et al., 2021). The study of HGFs has led to the emergence of a number of stylized facts, featuring prominently the robust finding that there is little persistence in HGFs (Holzl, 2014), such that individual HGFs are unlikely to repeat their growth performance in subsequent periods, thus earning them the label "one hit wonders" (Daunfeldt and Halvarsson, 2015). The policy literature therefore speaks of "high-growth episodes" rather than high-growth firms (Grover Goswami et al., 2019). The lack of persistence of HGFs has been highlighted as a near-fatal problem for HGF policy. HGFs are prohibitively hard to predict, and when an HGF is observed it is already too late to facilitate its emergence, thus making HGFs a problematic policy target.

HGF policy may still have hope, however. We seek to contribute to the possible emergence of a new empirical stylized fact. While HGFs lack persistence at the firm level, there might be persistence at the regional level. For example, the region of Silicon Valley has made a disproportionately large contribution to high-growth entrepreneurship that has made it the envy of regional policy-makers across the globe. Inspired by the case of Silicon Valley and other success stories, the EE approach has sought to uncover the secrets of their success, focusing in particular on the local institutions and networks facilitate the emergence of a vibrant entrepreneurial culture that can be expected to lead to consistently high HGF shares (Stam, 2015; Spigel, 2017; Leendertse et al., 2021).

The EE approach considers that regional level institutions, actors and factors are significant determinants of entrepreneurship outcomes. EE builds upon insights that entrepreneurship is a regional-level phenomenon (Feldman, 2001) that depends upon the context of local institutions (Autio et al., 2014), with considerable variation in entrepreneurship between regions within countries (Fritsch and Wyrwich, 2014), and with entrepreneurial activity drawing upon regionally-embadded growth-enhancing institutions (Bosma et al., 2018). The EE literature actually makes explicit claims regarding the expected persistence of HGFs across regions. Spigel (2017, p49) writes that "[e]ntrepreneurial ecosystems have emerged as a popular concept to explain the persistence of high-growth entrepreneurship within regions." Relatedly, Spigel and Harrison (2018, p155) explain that: "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." However, the evidence base for these claims of regional-level persistence of HGFs is almost non-existent, which we consider to be a genuine gap in the literature. The best available evidence, to our knowledge, regarding regional persistence of HGFs was suggested by Friesenbichler and Holzl (2020). They find evidence of moderate autocorrelation:

> "A simple autoregressive ordinary least squares (OLS) regression using information on NUTS-3 regions in Austria shows that the lagged HGF share explains 18% of the variance of HGF shares. The β-coefficient points at medium levels of autocorrelation ($\beta = 0.40$, $p = 0.075$, regionally clustered standard errors)." (Friesenbichler and Holzl, 2020, p1586).

However, their analysis focused on just one country, they didn't focus specifically on the issue of HGF persistence at the regional level. Also, while Friesenbichler and Holzl (2020) used data that was rich in some dimensions, it was limited in some other dimensions (e.g. HGF shares are measured using a

categorical rather than a continuous variable, and also their data cannot distinguish between enterprises and establishments).

We contribute to the literature by presenting an in-depth investigation focusing squarely on investigating the regional persistence of HGFs, applying transparent and uncomplicated techniques to derive insights that are of considerable policy interest. We begin with a simulation model, that closely follows ideas in the EE literature, to derive some precise predictions for region-level persistence of HGF shares. We then present results for 2 countries, using data that cover a relatively long time period (2004-2019 for Croatia, 2008-2014 for Slovenia). Using a mix of scatterplots and regressions for various years, we find mixed evidence for persistence of HGFs across regions. Overall, there is no clear persistence of HGF shares in Croatia, although there is persistence of HGF shares in Slovenia. Our results are consistent with notions that persistence of HGF shares depends on the business cycle in Croatia, thus suggesting that the importance of EEs varies over the business cycle. The case for persistence of HGFs across regions is weaker than has been suggested in previous theoretical research. Our mixed results are therefore a challenging puzzle for the EE approach.

We also contribute to the literature by looking at the persistence of industry-level HGF shares, to investigate intuitions from the SSI (sectoral systems of innovation) approach (Malerba, 2002). Intuitions based on the SSI suggest that sector-specific innovation regimes (constrained by factors such as technological opportunity conditions, appropriability regimes, cumulativeness of the knowledge base, and characteristics of knowledge and its transmission; Breschi and Malerba, 1997) may shape the prevalence of HGFs. To our knowledge, industry-level persistence of HGF shares has not been explicitly investigated before. We find stronger persistence at the industry-level than at the regional-level, perhaps hinting that the SSI approach does better than EE in predicting HGF emergence.

Another valuable contribution of ours is that we provide rigorous quantitative analysis (simulation model, and longitudinal analysis from 2 census datasets from 2 countries) to the EE literature. Previous EE scholars have commented on how the EE literature has, thus far, leaned towards conceptual papers and qualitative papers, to the detriment of quantitative papers (Rocha et al., 2021, Table 2). EE theorizing has tended to be conceptual and descriptive, rather than formalized, which has led some to reject EE as a vague conceptual framework with "amorphous" qualities (Brown and Mawson, 2019, p358). Closer empirical scrutiny of EE predictions is needed (Brown and Mawson, 2019, p362), in particular, moving from a static dimension towards a longitudinal dimension (Malecki, 2018, p12).

The paper unfolds as follows. Section 2 concretizes previous intuitions from the EE approach into a simulation model that makes clear predictions regarding the expected sign and magnitude of the persistence of regional-level HGF shares. This simulation model therefore helps us formulate our testable hypotheses. Section 3 presents the data for Croatia (2004-2019) and Slovenia (2008-2014). Section 4 presents our analysis, first for Croatia (subsection 4.1) then for Slovenia (subsection 4.2). Section 5 discusses our findings, and Section 6 concludes.

2. Background

Entrepreneurial Ecosystems (EE) thinking is emerging as a popular framework for contemplating the innovative and entrepreneurial performance of regions. Brown and Mawson (2019, p347) refer to EE as "the latest industrial policy blockbuster". Instead of being a unique independent discovery, in a Mertonian sense, the EE perspective stands atop the shoulders of giants, in that it draws upon a large number of previous systemic approaches to considering how the co-location of firms, their supporting actors and factors, as well as public organizations and institutions can create an environment that stimulates innovation performance, entrepreneurship, and economic growth. Previous theoretical approaches in this vein include 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 many more (Malecki, 2018). Relatedly, other scholars have put sector-specific and technology-specific boundaries to their theoretical approaches to clusters of innovative and entrepreneurial activity, such as the Sectoral Systems of Innovation (Malerba, 2002).[1] Following on from these earlier milestones in the literature, EE is now gaining momentum such that "the concept of entrepreneurial ecosystem has become a 'trendy' topic within academic and policy communities" (Lafuente et al., 2021, p1).

The main features of the ecosystem approach that we keep in mind are that there are many supporting dimensions or contributing factors, that these are interconnected and coevolving, and that these factors are persistent over time. These features are used to build a dynamic simulation model. Simulation models have been described as particularly useful in the context of entrepreneurial ecosystem research (Abootorabi et al., 2021, p18).

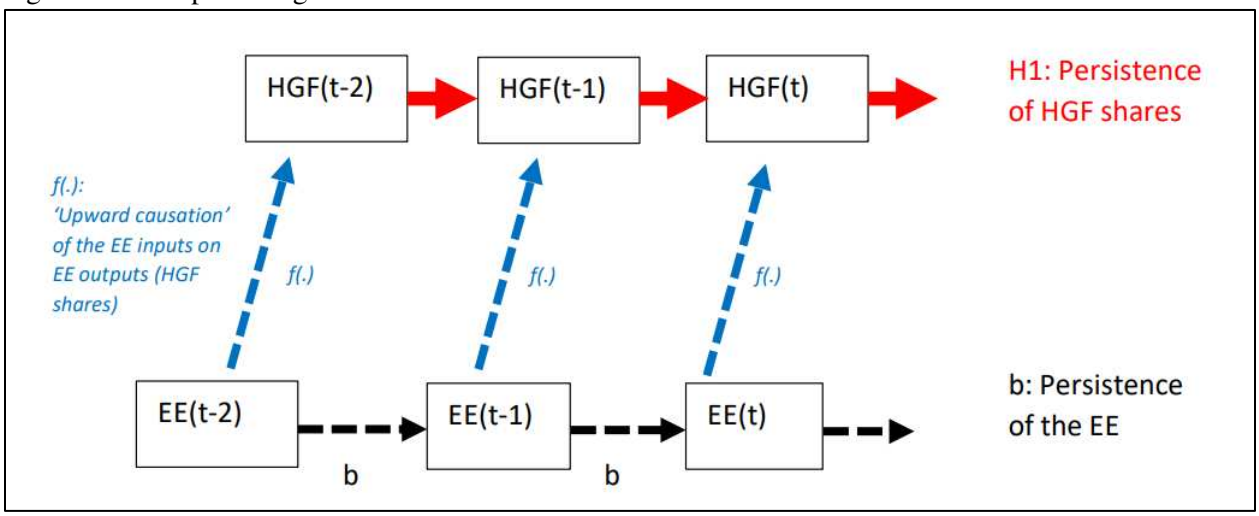
The outcome variable of the ecosystem approach is high-growth firms at a regional level (Spigel, 2017; Stam and Van de Ven, 2021). In a nutshell, our simulation model shows that high persistence in the inputs is expected to lead to high persistence in the output (i.e. high persistence in regional-level share of high-growth firms). Hence, we derive an intuitive and yet relatively unexplored prediction of EE theory: that HGF shares should be persistent across regions.

[1] A comparison of the EE approach with these previous concepts can be found in Spigel and Harrison (2018).

2.1 Intuition

Figure 1 provides a simple illustration of the intuition behind our analysis.

Figure 1: Conceptual diagram



Colour online. We test the hypothesis of HGF persistence represented by the thick red arrows.

The EE elements at time t are the inputs that are converted into the output HGF(t) by the homogenous function f(.), which is assumed to be constant over time:

$$HGF_{t-2} = f(EE_{t-2})$$

EE elements are highly persistent, i.e. $b \approx 1$, where:

$$EE_{t-1} = bEE_{t-2}$$

Hence:

$$HGF_{t-1} = f(EE_{t-1})$$

Then, by substitution:

$$HGF_{t-1} = f(EE_{t-1}) = f(b.EE_{t-2}) = b.f(EE_{t-2}) = b.HGF_{t-2}$$

Where $b \approx 1$. Hence, under plausible assumptions regarding the function f(.), persistence in EE inputs implies that there is persistence in regional HGF shares.

The following section builds a more complicated simulation model that builds more closely on EE theory, although the basic idea is similar.

2.2 Simulation model of an entrepreneurial ecosystem

2.2.1 Static model

Leendertse et al (2021) give a description of an entrepreneurial ecosystem that is sufficiently detailed that we can closely follow their suggestions to construct a simulation model.

The entrepreneurial ecosystem is a function that maps inputs X_{ik} into an output Y_i for region i.

$$Y_i = f(X_1, X_2, \cdots X_k \cdots X_K)$$

The natural choice of an output (Y_i) is "productive entrepreneurship" which is best captured in terms of high growth firms (Leendertse et al., 2021).[2] Indeed, a widely-held view on entrepreneurship ecosystems is that high-growth firms are the output (Spigel, 2017; Spigel and Harrison, 2018; Stam and Van de Ven, 2021). Leendertse et al. (2021) lack data on HGF share per region, therefore they use the number of unicorns from Crunchbase, but they mention that a problem with their indicator of HGFs is that the Crunchbase data they use is unrepresentative and only covers 0.2% of firms. Instead, we use an indicator which is much more widely used in HGF policy, which is the OECD HGF indicator (Friesenbichler and Holzl, 2021).

There are k inputs $(X_1, X_2, \cdots X_k \cdots X_K)$. The number k reflects the many different dimensions of entrepreneurial ecosystems (Lafuente et al., 2021), and is chosen to be between 5 and 14 in the literature,[3] for simplicity we take k=5.

For empirical analysis, these k inputs need to be given names, measured, and analysed. Stam (2015), Stam and Van de Ven (2021) and Leendertse et al (2021) suggest that k=10 and name these inputs as follows: Formal Institutions; Entrepreneurship culture; Networks; Physical Infrastructure; Finance; Leadership; Talent; New Knowledge; Demand; and Intermediate Services.

For the purposes of our model, we can remain agnostic regarding the names and labels ascribed to the various EE inputs or elements. Giving our own labels to these elements could unnecessarily introduce potentially controversial issues that could be a distraction from the main message of our simulation model. If we had to give names to the factors, or choose variables to proxy for these factors, this might leave us open to criticism. For example, factors such as "leadership" are hard to measure at a regional level, hence any proxy we might have to refute how leadership affects HGF share could be side-stepped by either criticizing our proxy for "leadership", or criticizing our choice of "leadership" as a label for X_k. The variables being used are vaguely defined,[4] which to some extent may make it hard to derive and evaluate testable hypotheses of a theory that remains elusive and abstract.

---

[2] Note however that Lafuente et al (2021) take high growth entrepreneurship as an input rather than an output (i.e. in the context of our model, as X_k instead of as Y). Instead, the outputs Y in Lafuente et al (2021) are GDP per capita or venture capital investments. This contrasts with Leendertse et al (2021) who take venture capital as an input, and high growth entrepreneurship as an output. Clearly, we can say that entrepreneurship ecosystems approach faces concerns about endogeneity.

[3] Other frameworks consider frameworks with five (Vedula and Kim, 2019), six (Isenberg and Onyemah, 2016), seven (Radosevic and Yoruk, 2013, RP), ten (Stam and Van de Ven, 2021; Leendertse et al., 2021) and fourteen elements (Acs et al., 2014).

[4] Leendertse et al. (2021, p5): "some elements such as institutions are multi-faceted and hard to capture in one variable."

Instead, we evaluate the predictions of EE by remaining at an abstract level. We merely assume that these EE elements are slow-changing in the sense that they do not change much (in the perspective of a regional-level ranking) from one year to the next. This assumption of persistence is probably not controversial. Indeed, one of the most basic facts of institutions (formal and informal) is that they are surprisingly inert and persistent over time (North, 1990).

We now need details regarding the functional form $f(.)$ that maps the inputs to the outputs. The inputs X_k are given an equal weight in terms of their contribution to the final outcome (Leendertse et al, 2021).[5] Also, the inputs X_k are interdependent and inter-related (Stam and Van de Ven, 2021; Leendertse et al, 2021). Correlations between these inputs in the order of 0.5 or 0.6 are usual (Leendertse et al., 2021, p10). We therefore generate values of X_k that are designed to be correlated with each other.

Regarding the function that maps the inputs X_k onto the output Y, there is a discussion in the EE literature whether this should be additive or multiplicative (Stam and Van de Ven, 2021; Leendertse et al., 2021). Leendertse et al., (2021, p8) decide that the best function is a function whereby the values of the inputs X_k are symmetrically distributed around a mean value of 1 (i.e. μ=1), and whereby the functional form of the interactions between the inputs X_k is multiplicative. This is consistent with the idea that improving the lowest-scoring dimension of the ecosystem will have the greatest impacts on the overall outcome (in line with the bottleneck method in Acs et al., 2014 and Lafuente et al, 2021).[6] We therefore adopt the multiplicative specification which introduces strong interdependence between the inputs X_k. Furthermore, we assume that the least controversial choice of a symmetric distribution for the inputs X_k is a Gaussian (μ=1, σ=1) distribution.

$$Y_i = f(X_1, X_2, \cdots X_k \cdots X_K) = X_1 \times X_2 \times \cdots \times X_k \times \cdots \times X_K = \prod_{k=1}^{K} X_k \tag{1}$$

2.2.2 Dynamic model

"Much of the extant research on EE is static and cross-sectional rather than longitudinal in nature" and this is considered to be a limitation that is hampering progress in EE research (Spigel and Harrison, 2018, p165). We therefore move from a static to a dynamic model which, given our focus on regional HGF persistence, is a necessary step. A dynamic model is well in line with the spirit of entrepreneurial ecosystems research (Abootorabi et al., 2021; Leendertse et al., 2021).[7]

Equation (1) can be rewritten using indices t to denote the time period:

[5] 15: "The current index is formed under the assumption that each element is equally important for the quality of the ecosystem."

[6] For the system $Y = \prod_{k=1}^{K} X_k$, we have $\frac{dY}{dX_k} = \prod_{j \neq k}^{K} X_j$, which implies that the largest marginal effects are observed for the dimension that has the lowest value.

[7] For example, in their conclusion, Leendertse et al (2021, p15) write "In sum, we need to move from a comparative static analysis to a dynamic analysis."

$$Y_{i,t} = f(X_{1it}, X_{2it}, \cdots X_{kit} \cdots X_{Kit}) \tag{2}$$

The dynamics of X_{kit} are represented thus:

$$X_{kit} = \beta_k X_{kit-1} + \varepsilon_{it} \tag{3}$$

2.2.3 Simulation model

Regarding the dynamics of X_{ik} from one period to the next: we parameterize the persistence parameters β_k such that they are highly persistent. Few would disagree that the inputs X_{ik} are highly persistent. "Universities are perhaps the most frequently identified actor/institution in entrepreneurial ecosystems after entrepreneurs themselves, and a large subset of research focuses on universities as hubs of such ecosystems" (Malecki, 2018, p9). The presence of universities in a region is, of course, highly persistent from one year to the next. Another example would be that one of the inputs X_{ik} mentioned specifically in Leendertse et al (2021) is "entrepreneurship culture." Fritsch and Wyrwich (2014) report that region-specific entrepreneurship culture is surprisingly persistent over the period 1925-2005, with a correlation coefficient of 0.290 between self-employment rates in 1925 and self-employment rates 2000-2005 in their sample of East German firms. An autocorrelation coefficient of 0.290 over 80 years is equivalent, in terms of a standard AR(1) time-series model[8] to an autocorrelation coefficient of about $\beta_k = 0.985$ on an annual basis.[9]

Yet another example of the high persistence of EE elements could be the particular formal institutions have been directly linked to the emergence and sustained regional advantage of Silicon Valley (Gilson, 1999; Fallick et al., 2006). The state of California is unique among US states in that no-compete agreements (also known as covenants not to compete)[10] are not legally recognized, whether it be the case of employees moving between Californian firms, or employees moving to California from other states (Gilson, 1999). This has led to California being an unusually rich ecosystem in terms of knowledge spillovers, entrepreneurial spinouts, reallocation of employees towards high-potential firms, and also even for attracting talented employees from other US states. This crucially-important legal element of the Silicon Valley EE is, of course, highly autocorrelated and persistent ($\beta_k \approx 1.00$) in the sense that the law does not generally change from one year to the next, or perhaps even from one century to the next.

Our simulation model makes the assumption that the value of β_k is the same for each of the elements of the EE. Therefore, by dropping the index k, the notation can be simplified from β_k to β. Our initial choice is $\beta = 1$, which corresponds to an $I(1)$ integrated and hence non-stationary time series process which is better known as the random walk (Shumway and Stoffer, 2017). The random walk process has the properties that the mean and variance are not restricted to taking the same values over time, thus allowing for strategic

[8] I.e. a first-order autoregressive model (Shumway and Stoffer, 2017).

[9] Consider a simple first-order autoregressive (AR(1)) time series autocorrelation model which is extended by iteratively substituting lags, x(t) = 0.985(x(t-1)) = 0.985(0.985×x(t-2)), etc, to get x(t) = (0.985^80)x(t-80) = 0.30(x(t-80)).

[10] No-compete agreements are clauses in employment contracts that firms use to prevent individuals from working for a competitor, so that firms can protect their trade secrets, and guard against knowledge spillovers to rivals.

investments and chance to move certain institutions and assets to higher or lower values over time. Random walk models seem appropriate for modelling the dynamics of ecosystems because they have been applied previously to the evolution of firm size (e.g. Gibrat, 1931) and the evolution of city size (e.g. Gabaix, 1999).

We also explore how our results change regarding alternative parameter values for β: 0.9, 0.8, 0.7 and 0.6. Values lower than 0.6 would correspond to implausibly low persistence in the institutions underpinning entrepreneurial ecosystems, and are deemed to be irrelevant for our current purposes.

"The essence of ecosystems is the interaction among its elements." (Stam and Van de Ven, 2021, p826). This interaction between the elements is built into our model in terms of the multiplicative specification in equation (1), and also in the sense that the random shocks ε_{it} are not independent, but they have a correlation across elements i of magnitude 0.5, that is generated using a multivariate Gaussian random number generator (R package `mvtnorm`).

Due to the multiplicative nature of the model (equation (1)), the outcomes $Y_{i,t}$ take extreme values that can be either positive or negative (see e.g. Appendix Figure OSM.1.1). One approach would be to take logarithms, which is not possible in our case because $Y_{i,t}$ sometimes takes negative values. We therefore apply the Inverse Hyperbolic Sine (IHS) transformation, which is a familiar technique in the literature on HGFs (McKenzie, 2017), regional clusters (Moretti, 2021) and innovation (Arora et al., 2021).

We calibrate the simulation model to our Croatian data, by generating a panel data frame with 21 cross-sectional entities tracked over t=200 time periods. The first 100 periods are considered to be a spin-up run, and are discarded, so we focus on analyzing the last 100 periods.

These parameter values provide the information we need to simulate the persistence of the outcome Y, i.e. the value of γ, by performing regressions on this from:

$$Y_{i,t} = \gamma Y_{i,t-1} + \epsilon_{it}. \tag{4}$$

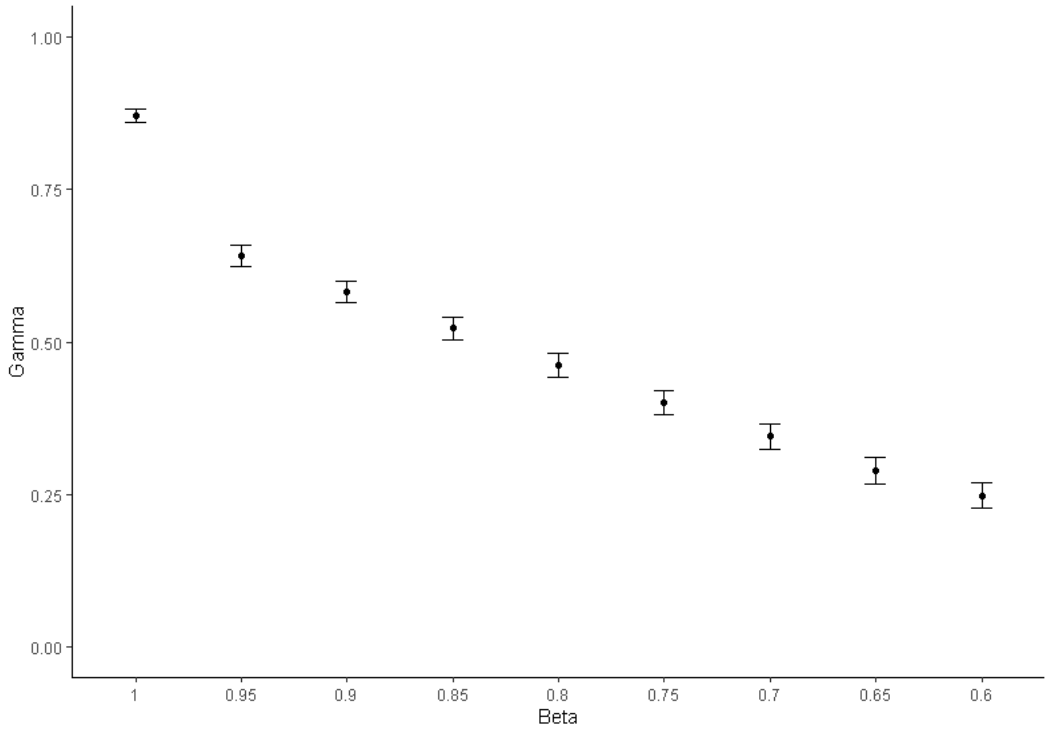
We therefore explore how variation in γ varies with variation in β.

Table 1 below show the corresponding values of γ obtained from OLS regressions on the simulated data. Table 1 shows that, when X_{kit} is modelled as a random walk (i.e. when $\beta = 1$), the persistence of Y_{it} is very high ($\gamma = 0.870$). Even when X_{kit} is modelled as a relatively low-persistence series, Y_{it} is still quite highly persistent, at a level which is highly statistically significant. Some of the values of β_k in our simulation model are implausibly low, for example it seems unlikely that the persistence of slow-changing institutional variables is only 0.6 from one period to the next, nevertheless to give the ecosystems approach the benefit of the doubt, we also consider this scenario. If our empirical analysis does not show any statistically significant persistence of Y_{it}, this suggests that there is a problem with current theorizing.

Table 1: simulation results

Meaning	Parameter									
Persistence of X_{kit}	Value of β	1	0.95	0.9	0.85	0.8	0.75	0.7	0.65	0.6
Persistence of Y_{it}	Value of γ	0.870	0.641	0.582	0.522	0.462	0.400	0.345	0.289	0.248
	Standard Error of γ	0.011	0.017	0.018	0.019	0.020	0.020	0.021	0.021	0.021

Figure 2: Graph of the simulation results in Table 1. Estimated persistence parameter γ for the output y, given different values of the persistence parameter β.



Notes: Error bars extend 1 standard error in each direction. Error bars obtained from estimating on simulated data with 21 regions (i.e. calibrated to our Croatian data) over the last 100 periods (i.e. discarding the first 100 periods as a spin-up run.

Based on our simulation model results, we therefore suggest:

Hypothesis 1: There is statistically significant positive persistence of the shares of HGFs at the level of regions

According to the values in Table 1, we would presumably expect an estimated persistence (i.e. an auto-regression coefficient) in the range of 0.4-0.6. Friesenbichler and Holzl (2020) report an autoregression coefficient of 0.40, which aligns well enough with our predictions, although it is not statistically significant at conventional levels.

Lack of support for Hypothesis 1 would cast doubt on the mapping of inputs to outputs. It would cast doubt on the relationship $\frac{\Delta Y_{it}}{\Delta X_{kit}}$ as portrayed in the model derived above, which emerges from our model that is closely derived from the entrepreneurial ecosystems literature (i.e. Leendertse et al., 2021). These doubts would be so strong that, even if it were possible for policymakers to manipulate X_{kit} , these would not have the expected effects on the outcome Y_{it}. As such, if the intended manipulations do not lead to the expected outcomes, this would suggest that the framework is not an effective approach for guiding policy.

2.3 Sectoral systems of innovation and entrepreneurship

A distinct but related strand of literature has focused on deriving policy implications for supporting innovation and entrepreneurship, not at the level of regional innovation systems, but at the level of sectoral systems of innovation (Malerba, 2002). The Sectoral Systems of Innovation (SSI) approach, to our knowledge, does not specifically make predictions about HGF shares as outcomes of the system, nevertheless for the reasons discussed below this seems worth pursuing here.

There are compelling reasons to suspect persistence in the inputs of a sectoral innovation system, when we reflect upon the various relevant dimensions that underpin sectoral innovation systems (Breschi and Malerba, 1997). These can be summarized by referring to four groups, following closely Breschi and Malerba (1997). The following discussion highlights how sectors differ systematically with regards to the innovation regimes in which they operate, with the various dimensions being relatively time-invariant and also constraining the expected growth of innovative firms.

First, there are technological opportunity conditions that refer to the likelihood of innovating for a given amount of investment in search activities. At a basic level, technological opportunity levels can be higher in some sectors than in others. Sectors may also differ with regards to the variety of technological search activities, approaches, and solutions available. The pervasiveness of technological opportunity varies across sectors, as the R&D outcomes can be applied to many or few products depending upon the opportunity conditions. Furthermore, the sources of technological opportunities may vary, with some sectors drawing on codified scientific knowledge, whereas others draw on internal R&D, or relations with suppliers, or perhaps collaborative relationships with universities and research institutes.

Second, appropriability conditions are expected to vary across sectors. Appropriability refers to the ease with which the fruits of R&D and technological search may leak to rivals. Sectors with weak appropriability conditions may contain firms that are reluctant to invest in R&D. Other sectors may draw more heavily on specific means of appropriation, such as pharmaceutical firms depending heavily on patents, while software firms rely on lead time advantages (Hall et al., 2014).

Third, sectors vary with regards to the cumulativeness of technological knowledge. Sectors vary with regards to the barriers to entry in terms of technological knowledge, ease of diffusion of knowledge, development of production capabilities and operating procedures, and so on. Scientific and technological domains vary with regards to the depth of knowledge required to approach the knowledge frontier, and the depreciation of knowledge as new breakthroughs are made.

Fourth, sectors vary with regards to characteristics of the relevant knowledge base. Technological knowledge may have product-specific applications or broadly-defined applications across various product categories. Knowledge varies according to tacitness and ability to transfer the knowledge. Knowledge also varies in terms of complexity (drawing in some cases on a large number of scientific disciplines) and in terms of independence (whether the knowledge is embedded in a larger system or whether it is easily extracted and shared).

It is plausible that many of the persistent differences across sectors in terms of sector-specific innovation regimes will shape the HGF shares of sectors. For example, HGF shares may be higher in sectors that emphasize first mover advantages and lead time benefits, and in sectors where technological breakthroughs are pervasive in the sense that they resemble general purpose technologies that can be applied to many products (rather than relating to a single product of narrow scope). HGF shares may be higher in sectors where technological knowledge is low-complexity, such that capabilities regarding production and operation can be scaled up relatively fast, and such that employees can be assembled and trained relatively fast. HGF shares may be higher in sectors that have regular cycles of new product introductions, and where customer bases are less firmly established. Some sectors may be better positioned to fuel sales growth by installing capital rather than hiring/training employees (e.g. capital-intensive industries such as cigarette-rolling).

HGF shares may be lower in less dynamic sectors where technological progress relies on slow-paced sure-footed accumulation of scientific knowledge (as opposed to sectors characterized by rapid depreciation of knowledge stocks that are more hospitable for fast-growth entrants). Sectors with high fixed costs of entry (either in terms of the cumulativeness of the knowledge base, whether patents or patent thickets inhibit production, or in terms of the costs of setting up or establishing a brand name through advertising efforts, etc) can be expected to have lower shares of HGFs.

There is a debate in the EE literature about whether sector matters. "Cluster and RIS frameworks are primarily concerned with the flows of technical knowledge within a particular industrial sector or between sectors that spur innovation. However, ecosystem research has remained largely industry agnostic." Spigel and Harrison (2018, p156). One reason for this could be that the EE elements refer to broad categories (e.g. entrepreneurial culture, finance, leadership) that do not focus specifically on characteristics of the scientific and technological knowledge base and the associated sector-specific innovation regimes. Nevertheless, for the reasons discussed above, it seems worth investigating whether some sectors are more dependable seed-beds for HGFs than others. It will also be interesting to compare sector-specific HGF persistence and region-specific HGF persistence. To our knowledge, this is the first investigation of possible persistence in sector-specific HGF shares.

Hypothesis 2: There is statistically significant positive persistence of the shares of HGFs when comparing sectors

3. Data

3.1 Description of data sources

Leendertse et al (2021) measure high-growth firms (i.e. "productive entrepreneurship") without using microdata from national statistical offices. This is presumably because they focus on a large number of countries, and it remains notoriously difficult to get access to detailed microdata from national statistical offices for cross-country comparisons across Europe.[11] Instead, Leendertse et al (2021, p8) use data from Crunchbase, which only covers 0.2% of all new European firms.[12] Research into the policy implications regarding high-growth firms has repeatedly emphasized the need for representative data on HGFs (Nightingale and Coad, 2014; Aldrich and Ruef, 2018). This low representativeness of Crunchbase data, which covers only 1 in 500 new firms, seems problematic. Instead, we use nationally representative microdata.

We use two census datasets in our analysis representing the economy of Croatia, and of Slovenia, the same as in Coad and Srhoj (2020). For our primary analysis we focus on Croatia,[13] and then for additional analysis focus separately on Slovenia. Census dataset in Croatia stems from the Financial Agency (FINA). All limited liability firms and publicly listed firms are obliged by law to report their balance sheets as well as their profit and loss statements to FINA. The advantage of having a census dataset is coverage of firms from all industries and of all sizes, while at the same time missing values do not pose a serious issue. The dataset spans 2004–2019, during this period the financial reporting is homogenous at FINA and we choose 2019 as the last reported year in order to avoid bias due to the 2020 Covid-19 pandemic. The initial dataset has 1,636,987 firm-year observations (in the period 2004-2019). The data includes reliable firm information such as revenue, employment, and the headquarters location at NUTS 2 and 3 level, city/municipality level and settlement level.[14]

3.2 Regional unit of analysis

The EE literature is not always clear about what should be the unit of analysis in terms of the regional dimension. This can be explained, in part, by data constraints. On the one hand, the ideal unit of analysis would be relatively disaggregated, to have a finer-grained analysis of distinct regions. On the other hand, data on variables such as entrepreneurial attitudes (used in some empirical EE analysis such as Stam and

[11] Some recent heroic efforts are here: https://ec.europa.eu/eurostat/cros/content/data-without-boundaries-0_en

[12] See also Leendertse et al (2021, p8): "We also explored using the ORBIS data of Bureau Van Dijk as an alternative (Bureau van Dijk, 2020; Dalle et al., 2017). However, we perceived this data to be inadequate for our purposes. First, the serial correlation between the different years in the database was very low."

[13] For those unfamiliar with Croatia, Croatia is home to unicorns and some very innovative firms, despite having only about 4 million inhabitants. For example, Infobip is an IT unicorn founded in 2006 in the city of Vodnjan, part of tourism-centered Istria county. Rimac Automobili is a unicorn founded in 2009 in the city of Sveta Nedelja, part of Zagreb county (different from the city of Zagreb), 24% of Rimac Automobili is owned by Porsche, while they acquired the famous Bugatti Automobiles, and are home to some of the world's most innovative projects like developing autonomous robotaxis. With two unicorns for a population of only 4 million, Croatia has one of the highest rates of unicorns per million citizens in the European Union. However, there are other examples of very high-growth firms with unicorn potential, such as Nanobit founded in 2008 in the city of Zagreb or Gideon Brothers founded in 2017 in the city of Osijek (Osijek-Baranja county) which develop autonomous mobile robots powered by AI and 3D vision.

[14] There are two NUTS 2 regions (changed to four recently), 21 NUTS 3 regions, 556 cities/municipalities, and about 3000 settlements.

Van de Ven (2021) and Leendertse et al. (2021)) are difficult to obtain at a disaggregated sub-national level and may only be available at the national or aggregated (e.g. NUTS 2) level.

We follow the pioneering paper by Friesenbichler and Hölzl (2020) and look at the NUTS 3 level, which in Croatia is 21 counties.[15] These 21 NUTS-3 regions are administrative units, governed by a "county", which have a separate budget that aims to maximize the welfare of the citizens and firms in its NUTS3 region. There is substantial heterogeneity among the NUTS3 regions within a NUTS2 region (e.g., the wealthy Dubrovnik county and poorer Licko-senjska county). Appendix OSM.3 contains maps of the NUTS 3 regions in Croatia and Slovenia.

In Slovenia we use Agency of the Republic of Slovenia for Public Legal Records and Related Services (AJPES) dataset. Firms of all sizes and types registered in Slovenia are obliged to deliver their annual financial statements to AJPES. Data encompasses years 2008-2014 enabling us to focus on two three-year periods: 2008-2011 and 2011-2014. Again, we focus on NUTS 3 level, which in Slovenia is 12 statistical regions. While Slovenia is a small country with only 12 statistical regions, nevertheless 12 is also the number of regions in Stam and Van de Ven's (2021) analysis of EEs in the Netherlands.

3.3 HGF definition

The output of an EE corresponds to "productive entrepreneurship" which can be proxied by the share of HGFs in a region's population of firms (Stam and Van de Ven, 2021, p817). The best-known HGF indicator comes from the Eurostat-OECD definition (Eurostat-OECD, 2007; Henrekson and Johansson, 2010). This HGF indicator is a dummy variable that takes the value 1 for firms with 10 or more employees (*E*) in the initial period (*t=0*), and a geometric average of at least 20% growth per year over 3 years. In other words, $HGF_t = 1$ if the following conditions are satisfied:

$$\left(\frac{E_{t+3}}{E_t}\right)^{\frac{1}{3}} - 1 \geq 20\% \qquad (5)$$

With a restriction on initial size:

$$E_{t=0} \geq 10 \qquad (6)$$

Hence, the growth rate threshold to be an HGF is 72.8% over a period of 3 years. This indicator (share of HGFs at the NUTS-3 level) is chosen to make our results closely comparable to those in Friesenbichler and Holzl (2020). The restriction of 10 employees at start is relaxed in some of the subsequent robustness analysis. HGFs sometimes measured as 5+ employees instead of 10+ employees (OECD 2013, p49), so we take an alternative threshold of 5+ employees.

[15] Leendertse et al (2021, p5) focus on NUTS 2 level data for some of their variables, because NUTS 3 level data is not always available. In a few cases, NUTS 2 level data is not available, and they use country-level data.

3.4 Methodology

Using the data (subsection 3.2) and HGF definitions (subsection 3.3) we construct our main variable. HGF definition can be calculated based on employment or based on revenue. We do both. The procedure is the same, but the variable to calculate growth is different. Let us define *HGF employment*. We split the dataset into five periods: 2004-2007; 2007-2010; 2010-2013; 2013-2016; and 2016-2019. We keep only those firms with at least 10 employees in the first period (i.e. in the period 2016-2019, the firm has to have at least 10 employees in 2016). Among these firms, we divide their employment at time t+3 (i.e. for the period 2016-2019; t+3 is 2019) with their employment at time t (thus in our example 2016). If the growth is larger than 72.8% we give this firm an indicator *HGF employment*. To calculate the share of *HGF employment* in each of the 21 regions, we divide the number of HGF employment with the total number of firms (only firms with 10 or more employees are included) in period t. We repeat this procedure for the *HGF revenue* where we use total sales instead of employment to calculate the growth rate. We also use an alternative HGF definition, where we take those firm with at least 5 employees (instead of 10 employees), to respond to concerns that focusing on firms with 10+ employees in the initial period could be too restrictive (Daunfeldt et al., 2015).

4. Analysis

4.1 Croatian data

4.1.1 Correlations and Scatterplots

Table 2 shows some basic correlations between HGF share for each region in one period and the preceding period. The persistence of HGF share is explored from various angles: HGFs in terms of employment growth or sales growth; Pearson correlations or Spearman's rank correlations (where the latter are more robust to outliers); for 4 possible pairings of consecutive periods (ranging from the period 2004-2010 until the period 2013-2019); and with two alternative thresholds for initial size (10 or 5 employees).

Pooling together the years, we never observe the theoretically-predicted positive correlation. In fact, the correlation between HGF share in one period and the next is negative and statistically significant for HGFs defined in terms of sales (both for 10-employee and 5-employee thresholds, both for Pearson correlations and Spearman's rank correlations) in columns (3) and (4).

Focusing on specific years rather than the pooled results, in a few cases we observe statistically significant positive correlations.[16] In most cases, however, Table 2 shows that there is no statistically significant correlation between the HGF shares of regions over two consecutive periods. This lack of serial correlation suggests that there is no persistence in regional HGF shares over time, and hence overall does not support H1.

Stronger support for persistence of HGF shares is found in Table 3, which focuses on HGF shares in specific sectors, and thereby tests Hypothesis 2. In most, but not all cases, there is a significant positive autocorrelation in industries' HGF shares.

Table 2: regional HGF share correlations: Croatia

	HGF Employment		HGF Sales	
	(1)	(2)	(3)	(4)
	Pearson correlation [p-value]	Spearman's rank correlation [p-value]	Pearson correlation [p-value]	Spearman's rank correlation [p-value]
10 employee thresholds				
Pooled (2004-2019)	-0.123 [0.263]	-0.132 [0.230]	**-0.432 [0.000]**	**-0.250 [0.022]**
Period: 2004-2007 & 2007-2010	-0.324 [0.152]	-0.088 [0.703]	-0.068 [0.770]	-0.195 [0.396]
Period: 2007-2010 & 2010-2013	**0.414 [0.062]**	0.270 [0.235]	**0.482 [0.027]**	**0.447 [0.044]**
Period: 2010-2013 & 2013-2016	-0.350 [0.120]	-0.209 [0.361]	0.260 [0.256]	0.106 [0.645]

[16] These are Employment growth HGFs in the period 2013-2019 and also Employment growth HGFs (10 employee threshold, Pearson correlation) in the period 2007-2013, and also Sales HGFs for the period 2007-2013 when the initial threshold is 10 employees.

Period: 2013-2016 & 2016-2019	**0.659** **[0.001]**	**0.449** **[0.041]**	-0.009 [0.969]	0.077 [0.741]
5 employee thresholds				
Pooled (2004-2019)	-0.168 [0.126]	-0.129 [0.240]	**-0.427** **[0.000]**	**-0.289** **[0.008]**
Period: 2004-2007 & 2007-2010	0.067 [0.772]	0.155 [0.502]	0.043 [0.854]	0.149 [0.517]
Period: 2007-2010 & 2010-2013	0.341 [0.131]	0.264 [0.247]	0.341 [0.130]	0.318 [0.160]
Period: 2010-2013 & 2013-2016	-0.275 [0.228]	-0.190 [0.409]	-0.208 [0.366]	-0.149 [0.517]
Period: 2013-2016 & 2016-2019	**0.561** **[0.008]**	**0.438** **[0.049]**	0.222 [0.333]	0.216 [0.346]

Note: p-values reported in the brackets.

Table 3: industrial HGF share (NACE 1-digit) correlations: Croatia

	HGF Empl.		HGF Sales	
	(1)	(2)	(3)	(4)
	Pearson correlation [p-value]	Spearman's rank correlation [p-value]	Pearson correlation [p-value]	Spearman's rank correlation [p-value]
10 employee thresholds				
Pooled (2004-2019)	0.078 [0.502]	0.164 [0.157]	-0.023 [0.842]	0.123 [0.288]
Period: 2004-2007 & 2007-2010	**0.573** **[0.010]**	0.333 [0.163]	-0.336 [0.160]	0.111 [0.652]
Period: 2007-2010 & 2010-2013	-0.129 [0.598]	-0.196 [0.422]	0.325 [0.174]	0.212 [0.381]
Period: 2010-2013 & 2013-2016	-0.086 [0.727]	**0.471** **[0.042]**	**0.730** **[0.000]**	**0.728** **[0.001]**
Period: 2013-2016 & 2016-2019	**0.674** **[0.002]**	**0.582** **[0.009]**	**0.923** **[0.000]**	**0.898** **[0.000]**
5 employee thresholds				
Pooled (2004-2019)	**0.260** **[0.023]**	**0.248** **[0.031]**	0.028 [0.809]	0.058 [0.618]
Period: 2004-2007 & 2007-2010	**0.696** **[0.001]**	**0.526** **[0.022]**	0.019 [0.939]	0.248 [0.306]
Period: 2007-2010 & 2010-2013	0.291 [0.228]	0.111 [0.652]	**0.437** **[0.061]**	0.389 [0.100]
Period: 2010-2013 & 2013-2016	0.329 [0.168]	**0.444** **[0.058]**	**0.555** **[0.014]**	**0.451** **[0.054]**
Period: 2013-2016 & 2016-2019	**0.712** **[0.001]**	**0.630** **[0.004]**	**0.811** **[0.000]**	**0.733** **[0.001]**

Note: p-values reported in the brackets.

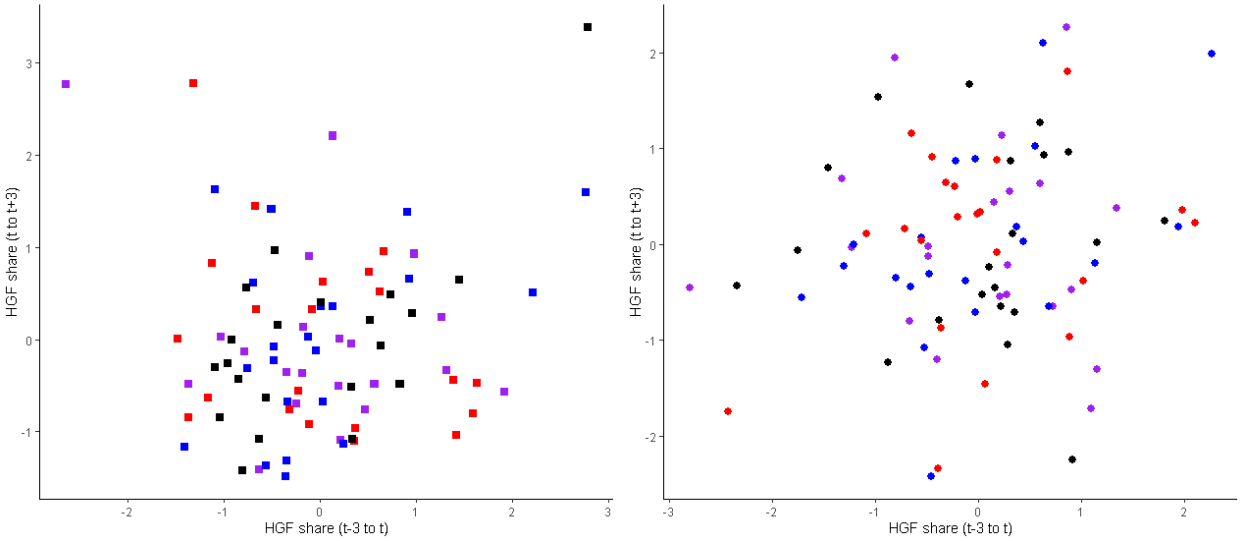
We continue our analysis of autocorrelation in regional HGF shares using scatterplots, that show the actual datapoints and thereby allow us to assess the role of phenomena such as outliers and non-linear relationships.

The datapoints from different years are normalized by year (such that each year has mean = zero), and then pooled together (but with different colours for different years) and shown in Figure 3.

Figure 3 shows that there is no clear strong positive relationship in HGF shares from one period to the next. Regions with a high HGF share in one period are not particularly likely to repeat this performance in the following year. We do not see the strong positive correlation that was predicted by our theoretical model. A few outliers are visible, but they do not appear to be driving our results.

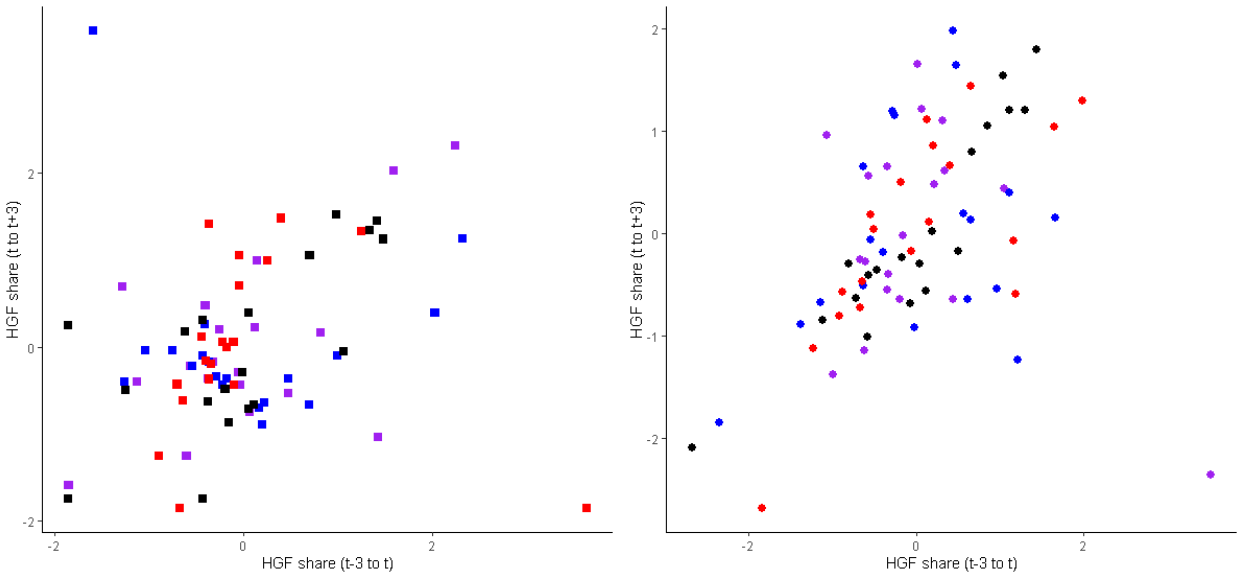
Figure 4 shows the corresponding scatterplot for investigating persistence of HGF shares within sectors. Positive correlation is easier to see here. Hence, so far, there seems to be more support for H2 (from Figure 4) than there is support for H1 (from Figure 3). Nevertheless, we continue our analysis using regressions.

Figure 3: persistence of HGF share at the regional level. Left: employment growth. Right: sales growth.



Notes: squares are for employment growth (left) and circles are for sales growth (right). Black: t=2016; Red: t=2013; Blue: t=2010; Purple: t=2007. The datapoints from different years are normalized by year (such that each year has mean = zero), and then pooled together. Colour online.

Figure 4: persistence of HGF share at the industrial level. Left: employment growth. Right: sales growth.



Notes: squares are for employment growth (left) and circles are for sales growth (right). Black: t=2016; Red: t=2013; Blue: t=2010; Purple: t=2007. The datapoints from different years are normalized by year (such that each year has mean = zero), and then pooled together. Colour online.

4.1.2 Regressions

Our regression equation is:

$$HGF\_share_{it} = \rho HGF\_share_{i,t-1} + \theta_{it}. \quad (7)$$

Where ρ is the estimate of the persistence parameter γ in equation (4), and θ_{it} is the usual error term.

Table 4 pools together our observations from different years, to get results that are comparable to those of Friesenbichler and Holzl (2020). While we aspire to compare our results to theirs, nevertheless this is not entirely possible because they focus on different time periods, in a different country, and also they transform their dependent variable (regional HGF share) from a continuous variable into discrete categories. Friesenbichler and Holzl observe positive autocorrelation of 40%, although it is not statistically significant at the 5% level. In our data, pooling the years together, there is no persistence in regional HGF shares when HGFs are measured in terms of employment or sales (columns (2) and (3)). The coefficients are small (0.087 and 0.090, respectively) and indistinguishable from zero. Taking an alternative definition of HGFs, in terms of firms with 5+ employees at start, the results are also insignificant (columns (6) and (7)). These results do not support H1. With regards to industry HGF shares, Table 4 contains no statistically significant results for the baseline HGF indicator of 10+ employees at start (columns (4) and (5)), although using the alternative definition of HGFs (5+ employees at start) does yield statistically significant autocorrelation of 43% and 42% respectively (for employment and sales HGFs, respectively) that is in line with the hypothesized autocorrelation magnitudes (columns (8) and (9)). Table 4 therefore offers nuanced support for H2. Overall, therefore, H1 is not supported, while H2 receives slightly stronger support.

Table 4: Regression results. Comparing our results with 10 and 5 employee threshold definition to F&H2020. OLS with regionally-clustered standard errors.

	10 employee thresholds					5 employee thresholds			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	F&H 2020: NUTS 3	Our data: NUTS 3	Our data: NUTS 3	Our data: sectors	Our data: sectors	Our data: NUTS 3	Our data: NUTS 3	Our data: sectors	Our data: sectors
	HGF empl. share	HGF empl. share	HGF sales share	HGF empl. share	HGF sales share	HGF empl. share	HGF sales share	HGF empl. share	HGF sales share
Regression beta coefficient	0.40	0.087	0.090	0.189	0.164	0.157	0.055	0.432	0.418
P-value; regionally (industry) clustered s.e.	0.075	0.552	0.312	0.345	0.125	0.243	0.500	0.000	0.003
R-squared	0.18	0.408	0.678	0.136	0.240	0.614	0.755	0.381	0.355
Observations	.	84	84	76	76	84	84	76	76

Notes: the number of observations for specification (1) is not clearly reported. Columns 2-9 include year dummies. We provide models (2-9) without year dummies in the Online Appendix.

Tables 5 and 6 below complement the correlations in Table 2 with regressions for employment growth HGFs (Table 5) and sales growth HGFs (Table 6). In most cases, the regional persistence of HGF shares is not significantly different from zero, and in a few cases it is even negative. For employment growth HGFs in Table 5, only 2 out of 10 of the regression models has the predicted positive and significant coefficient.

In one case, there is a weakly significant *negative* correlation between regional HGF share in 2004-2007 and the regional HGF share in 2010-2013. This sits awkwardly with the suggestion that time-invariant regional inputs are important determinants of HGF shares. Table 6 shows that only 3 out of 10 models have the expected positive and significant coefficient. In the remaining 7 out of 10 models of Table 6, the persistence is statistically indistinguishable from zero. Moreover, the low R^2 values cast doubt on whether HGF shares can be accurately predicted from HGF shares in the previous period. The R^2 values were higher in the pooled regressions in Table 4, although after a closer inspection we can confirm that this is driven by the inclusion of year dummies.

Further investigation of H2 on industry-level persistence of HGF shares in Croatia is in Table 7 (employment HGFs) and Table 8 (sales HGFs). In 5/10 cases in Table 7, and 4/10 cases in Table 8, do we find statistically significant persistence of HGF shares across industries. While we do not interpret this as strong evidence of industry-level persistence of HGF shares (H2), nevertheless there still appears to be more support for H2 than for H1.

Table 5: Predicting the share of HGFs in regions by employment indicator in time t, based on previous periods (t-1, t-2, or t-3)

	Dependent variable:									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	HGF (16-19)	HGF (13-16)	HGF (10-13)	HGF (07-10)	HGF (16-19)	HGF (16-19)	HGF (16-19)	HGF (13-16)	HGF (13-16)	HGF (10-13)
HGF (13-16)	**1.011**** **(0.383)**									
HGF (10-13)		-0.373 (0.236)			-0.329 (0.470)					
HGF (07-10)			**0.445**** **(0.189)**			-0.248 (0.390)		0.066 (0.216)		
HGF (04-07)				-0.157 (0.142)			0.209 (0.152)		0.035 (0.107)	-0.162* (0.092)
Constant	0.014 (0.015)	0.054*** (0.009)	0.019*** (0.007)	0.039*** (0.011)	0.066*** (0.017)	0.063*** (0.012)	0.041*** (0.013)	0.040*** (0.007)	0.039*** (0.008)	0.043*** (0.007)
Observations	21	21	21	21	21	21	21	21	21	21
R^2	0.434	0.123	0.171	0.105	0.041	0.020	0.060	0.003	0.004	0.096
Res. St. Error (df = 19)	0.014	0.012	0.011	0.010	0.019	0.019	0.019	0.012	0.012	0.011

Note: * p < 0.1; ** p < 0.05; *** p < 0.01. Standard errors reported in the brackets.

Table 6: Predicting the share of HGFs in regions by sales indicator in time t, based on previous periods (t-1, t-2, or t-3)

	Dependent variable:									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	HGF (16-19)	HGF (13-16)	HGF (10-13)	HGF (07-10)	HGF (16-19)	HGF (16-19)	HGF (16-19)	HGF (13-16)	HGF (13-16)	HGF (10-13)
HGF (13-16)	-0.008 (0.163)									
HGF (10-13)		0.257 (0.191)			0.036 (0.139)					
HGF (07-10)			**0.694**** **(0.254)**			-0.005 (0.320)		0.157 (0.166)		
HGF (04-07)				-0.027 (0.086)			-0.051 (0.103)		**0.281***** **(0.089)**	**0.260*** **(0.134)**
Constant	0.121*** (0.018)	0.087*** (0.019)	0.053*** (0.013)	0.059*** (0.015)	0.117*** (0.013)	0.120*** (0.018)	0.129*** (0.018)	0.101*** (0.012)	0.059*** (0.018)	0.044*** (0.024)
Observations	21	21	21	21	21	21	21	21	21	21
R^2	0.0001	0.067	0.232	0.005	0.002	0.00002	0.011	0.012	0.249	0.210
Res. St. Error (df = 19)	0.018	0.020	0.018	0.014	0.018	0.018	0.018	0.020	0.018	0.019

Note: * p < 0.1; ** p < 0.05; *** p < 0.01. Standard errors reported in the brackets.

Table 7: Predicting the share of HGFs in industry by employment indicator in time t, based on previous periods (t-1, t-2, or t-3)

	Dependent variable:									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	HGF (16-19)	HGF (13-16)	HGF (10-13)	HGF (07-10)	HGF (16-19)	HGF (16-19)	HGF (16-19)	HGF (13-16)	HGF (13-16)	HGF (10-13)
HGF (13-16)	**1.047*** (0.316)**									
HGF (10-13)		-0.046 (0.198)			-0.092 (0.272)					
HGF (07-10)			-0.258 (0.790)			**0.861*** (0.252)**		0.352 (0.243)		
HGF (04-07)				**0.284** (0.132)**			**0.338* (0.190)**		**0.315*** (0.097)**	-0.138 (0.428)
Constant	1.776 (1.677)	4.482*** (0.915)	5.405 (3.248)	1.140 (0.975)	6.661*** (1.292)	3.321** (1.355)	3.565*** (1.854)	3.076*** (0.942)	1.773*** (1.078)	5.622 (4.015)
Observations	19	19	19	19	19	19	19	19	19	19
R^2	0.455	0.007	0.017	0.328	0.021	0.264	0.165	0.106	0.345	0.019
Res. St. Error (df = 17)	2.711	2.358	4.341	1.797	3.650	3.149	3.355	2.237	1.915	4.336

Note: * p < 0.1; ** p < 0.05; *** p < 0.01. Standard errors reported in the brackets.

Table 8: Predicting the share of HGFs in industry by sales indicator in time t, based on previous periods (t-1, t-2, or t-3)

	Dependent variable:									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	HGF (16-19)	HGF (13-16)	HGF (10-13)	HGF (07-10)	HGF (16-19)	HGF (16-19)	HGF (16-19)	HGF (13-16)	HGF (13-16)	HGF (10-13)
HGF (13-16)	**1.262*** (0.122)**									
HGF (10-13)		**0.634*** (0.169)**			**0.728*** (0.239)**					
HGF (07-10)			0.565 (0.379)			0.832 (0.505)		**0.735* (0.406)**		
HGF (04-07)				-0.080 (0.066)			-0.105 (0.150)		-0.137 (0.111)	-0.078 (0.123)
Constant	-2.321 (1.498)	5.926*** (1.704)	5.531* (2.683)	8.774*** (1.615)	5.836*** (1.914)	6.940* (3.642)	15.125*** (3.225)	6.810*** (3.047)	15.060*** (2.341)	11.241*** (2.665)
Observations	19	19	19	19	19	19	19	19	19	19
R^2	0.852	0.533	0.106	0.113	0.376	0.163	0.046	0.238	0.146	0.036
Res. St. Error (df = 17)	2.405	3.126	4.984	2.859	4.939	5.722	6.110	3.993	4.228	5.176

Note: * p < 0.1; ** p < 0.05; *** p < 0.01. Standard errors reported in the brackets.

4.2 Slovenian data

4.2.1. Correlations and Scatterplots

We continue our analysis with evidence from a second country, i.e. Slovenia. Table 9 contains some correlation analysis of the regional-level persistence of HGF shares, and broadly corresponds to the correlations in Table 2 for the case of Croatia. In the case of Slovenia, however, all of the region-level correlations are positive and significant (eight cases out of eight), and furthermore they are closely aligned with the expected effect sizes from the simulation model (Figure 1). The bottom half of Table 9 focuses on industry-level HGF persistence. When HGFs are measured in terms of employment, the Pearson correlations are negative but also differ markedly from the Spearman's rank correlations, suggesting that the industry-level negative correlations in column (1) are driven by outliers, and not robust. However, the industry-level correlations in columns (3) and (4) suggest that industry-level HGF persistence is in line with theoretical predictions as long as HGFs are measured in terms of sales. Hence, in the case of Slovenia, there seems to be strong support for H1, but weaker support for H2.

Table 9: regional and industrial correlations: Slovenia

	HGF Empl.		HGF Sales	
	(1)	(2)	(3)	(4)
	Pearson correlation [p-value]	Spearman's rank correlation [p-value]	Pearson correlation [p-value]	Spearman's rank correlation [p-value]
REGION-LEVEL				
10 employee thresholds				
Regions Period: 2008-2011 & 2011-2014	**0.697** **[0.012]**	**0.685** **[0.017]**	**0.806** **[0.002]**	**0.860** **[0.001]**
5 employee thresholds				
Regions Period: 2008-2011 & 2011-2014	**0.732** **[0.007]**	**0.678** **[0.019]**	**0.791** **[0.002]**	**0.797** **[0.003]**
INDUSTRY-LEVEL				
10 employee thresholds				
Industry Period: 2008-2011 & 2011-2014	**-0.702** **[0.001]**	0.078 [0.751]	**0.673** **[0.002]**	**0.549** **[0.016]**
5 employee thresholds				
Industry Period: 2008-2011 & 2011-2014	**-0.654** **[0.002]**	0.156 [0.523]	**0.587** **[0.008]**	**0.526** **[0.022]**

Note: p-values reported in the brackets.

4.2.2. Regressions

Table 10 presents the Slovenian results, which correspond to Table 4 for Croatia. Table 10 shows positive and significant persistence in HGF shares across regions, which contrasts sharply with the results for

Croatia. Hence, in the case of Slovenia, there seems to be support for H1. Table 11 investigates H2 on industry-level HGF persistence. Focusing on results that are statistically significant at the 5% level or above, we observe the expected positive persistence only when HGFs are measured in terms of sales. Hence, in the case of Slovenia, there seems to be more support for H1 than for H2, which contrasts with the results for Croatia. However, our results for Slovenia cover only one period of observation (i.e. how HGF shares 2011-2014 are associated with HGF shares 2008-2011) and are therefore less reliable than the results for Croatia, because they could potentially be driven by e.g. temporary business cycle effects.

Table 10: Predicting the regional share of HGFs by employment and sales indicator in time t, based on previous period

	Dependent variable			
	10 employee thresholds		5 employee thresholds	
	(1)	(2)	(3)	(4)
	HGF employment (2011-2014)	HGF sales (2011-2014)	HGF employment (2011-2014)	HGF sales (2011-2014)
HGF employment (2008-2011)	0.509** (0.174)		0.512*** (0.158)	
HGF sales (2008-2011)		0.344*** (0.088)		0.358*** (0.075)
Constant	4.752** (1.575)	7.534*** (0.949)	4.841*** (1.463)	7.502*** (0.830)
Obs.	12	12	12	12
R^2	0.485	0.650	0.536	0.626
Res. St. Error (df = 19)	0.619	0.470	0.576	0.515

Note: * p < 0.1; ** p < 0.05; *** p < 0.01. Standard errors reported in the brackets.

Table 11: Predicting the industrial share of HGFs by employment and sales indicator in time t, based on previous period

	Dependent variable			
	10 employee thresholds		5 employee thresholds	
	(1)	(2)	(3)	(4)
	HGF employment (2011-2014)	HGF sales (2011-2014)	HGF employment (2011-2014)	HGF sales (2011-2014)
HGF employment (2008-2011)	-1.619* (0.840)		-1.397 (0.856)	
HGF sales (2008-2011)		0.226*** (0.024)		0.216*** (0.023)
Constant	26.855*** (8.878)	9.405*** (0.708)	25.624** (9.316)	9.871*** (0.755)
Obs.	19	19	19	19
R^2	0.493	0.453	0.427	0.344
Res. St. Error (df = 17)	5.234	2.270	5.502	2.722

Note: * p < 0.1; ** p < 0.05; *** p < 0.01. Standard errors reported in the brackets.

5. Discussion

In the context of our model, policy is effective if $\frac{\delta Y_{i,t}}{\delta X_{ikt}} > 0$. In other words, policy can intervene upon the EE element X_{ikt} to have an influence on the desired outcome $Y_{i,t}$ in region i in year t. EE elements X_{ikt} are highly persistent over time: $X_{ikt} \approx X_{ik,t+1}$. It is assumed that EE elements X_{ikt} have an influence on outcome $Y_{i,t}$. Hence, we expect that there is high persistence in outcomes $Y_{i,t}$ also. Overall, however, our results do not show consistent support for predictions of high persistence in HGF shares.

Persistence in outcomes $Y_{i,t}$ is measured as persistence of HGF shares for NUTS-3 regions, using census data for 2 countries (Croatia, then Slovenia). Our main results for Croatia show that there is no significant persistence of HGF shares, in any of the 4 regression models in Table 4, which is in contrast to theoretical predictions. We repeat the analysis on several alternative samples (measuring HGFs in terms of employment or sales growth; 10 employees vs 5 employees threshold; different years; graphical techniques as well as correlations and regressions) and our main findings hold: in general, there is no persistence of HGF shares. In a minority of cases, though, we observe the expected positive autocorrelation in regional HGF shares, when we disaggregate by years. In other words, regional persistence in HGF shares is found in some specific years but not others (Tables 5 and 6). A possible explanation could be business cycle effects. There is considerable variation in HGF shares over the business cycle, for example the HGF share in Dubrovnik-Neretva county changes from 0.078 (expansion, 2004-2007) to 0.027 (recession, 2007-2010) to 0.030 (recession, 2010-2013) to 0.030 (expansion, 2013-2016) to 0.056 (expansion, 2016-2019). Business cycle effects may also have been detected in other EE analyses (Stam and Van de Ven, 2021, p827).[17] If regional HGF shares do not fluctuate in parallel, but some regions fluctuate more than others, this could help to explain the lack of regional persistence of HGF shares over the business cycle. So far, the EE approach has not explicitly considered how the business cycle could affect EE outputs (i.e .HGF shares), in particular how the business cycle could affect some ecosystem regions more strongly than others, although this could be a fruitful line of future theoretical development.

In Croatia during our period of observation, there was a long recessionary period from 2009 to mid-2014. It may be that regional HGF persistence is higher for periods from recession-to-recession (moderate autocorrelation in Table 2 for 2007-2010 to 2010-2013; see also Table 5 column (3) and Table 6 column (3)) or expansion to expansion (higher persistence for employment HGFs in Table 2 for 2013-2016 to 2016-2019) than when the period studied is recession-to-expansion or expansion-to-recession (e.g. the lack of persistence observed in the periods 2010-2013 to 2013-2016, and 2004-2007 to 2007-2010, in Table 2). HGF persistence across the business cycle is beyond the scope of our current paper but would benefit from future research, perhaps with an analysis of several countries over a long time period.

The evidence for Slovenia, however, does find significant positive autocorrelation of regional HGF shares, and the effect sizes are in line with theoretical expectations. However, the evidence from Slovenia comes from a single period of observation (2008-2011 to 2011-2014, i.e. a recession-to-recession period) and may not be robust to other phases of the business cycle.

Overall, therefore, we find isolated pockets of support for H1 although this is certainly not an established regularity. This calls for further research from other countries (to complement our findings for Croatia and

[17] Stam and Van de Ven (2021, p827) observe variation in how their EE index explains HGF share, with an R^2 statistic ranging from 0.4370 (in 2009) to 0.6245 (in 2012), and 0.5256 (in 2015).

Slovenia, and results from Friesenbichler and Holzl (2020) for Austria), also giving a careful eye to possible business cycle effects.

Overall, we find more support for persistence in HGF shares at the level of sectors (H2) than for persistence in HGF shares at the level of regions (H1). Hence, the SSI approach seems better able to predict the persistence of HGF shares than the EE approach. This is surprising, because the SSI approach has not, to our knowledge, specifically suggested that HGF shares are the primary output of an SSI.

Empirical investigations involving volatile time series are sometimes aggregated over time to smooth out any random fluctuations. This often happens in analysis of high-frequency financial data, such as daily stock returns. In the context of National Systems of Entrepreneurship, Acs et al. (2014, p484) smooth out random fluctuations in their variables by taking two-year moving averages instead of using annual data. However, given that the standard indicator of HGFs (used here) is essentially an average growth rate over a three-year period (Eurostat-OECD, 2007), further smoothing this variable over a longer time period (e.g. a decade) seems excessive. Measuring HGF shares over a period of longer than three years will reduce the number of consecutive periods that can be used for measuring persistence of HGF shares.

Our results present a puzzle for the EE framework. The theory is relatively young and still needs refinement. Furthermore, empirical stylized facts still need to be collected. At present, the persistence of inputs is statistically incongruent with a lack of persistence of outputs, which casts doubt on the causal influence of EE elements on outputs. Further work could also investigate ways to evaluate other aspects of the function f(.) by which EE inputs are mapped into outputs. For example, does the function f(.) vary over the business cycle?

We therefore formulate a "broken clock" critique of the ability of the entrepreneurial ecosystems (EE) approach to give valuable recommendations for policy. Just as a broken clock is correct twice a day, EE predictions and recommendations may sometimes be correct, but are fundamentally flawed as long as time-changing outcomes (such as the time) are predicted using time-invariant variables (such as motionless clock hands). Taking the analogy further, we note that clock hands are associated with the passage of time, and that distinguishing between associations and causality is difficult for cointegrated series (e.g. if X and Y co-evolve, it may be difficult to distinguish between competing causal explanations $X \rightarrow Y$ or $Y \rightarrow X$), nevertheless if motionless clock hands are used to predict time-varying outcomes then it is easier to rule out that the clock is having any causal effect on the passage of time. Hence, EE variables seem unable to give reliable predictions of regional HGF shares, and furthermore EE variables appear not to have a causal influence on regional HGF shares.

6. Conclusion

The Entrepreneurial Ecosystems (EE) approach has become a fashionable and influential framework for thinking about entrepreneurship and innovation policy (Stam, 2015; Spigel, 2017; Leendertse et al., 2021). EE makes explicit predictions for the mapping of EE inputs into the persistence of shares of High Growth Firms (HGFs) at the regional level. Put differently, because many drivers of EE success are region-specific and slow-changing, regions with large shares of HGFs in one period are expected to have large shares of HGFs in the following period. We investigate whether there is any persistence in regional HGF shares. This allows us to test a prediction of EE theory, whereby time-invariant inputs are put forward to explain regional HGF shares. If EE is correct, then holding things constant (ceteris paribus): we can expect more HGFs from where HGFs emerged in the last period.

Surprisingly, we observe there is negligible persistence in HGFs at the regional level. This is incongruous with the observation that there is high persistence in the inputs to an entrepreneurial ecosystem. The relationship between inputs and outputs is so noisy that we conclude that the entrepreneurial ecosystem approach, according to its most recent formulations (i.e. Leendertse et al 2021) is not a useful approach for policymakers with regards to generating the main outcome of ecosystems, i.e. HGFs.

We therefore formulate a "broken clock" critique of EE in its current formulation. A broken clock tells the correct time twice a day, but overall it is not useful for telling the time. Moreover, manually manipulating the clock hands does not have any causal effect on the outcome (moving forwards or backwards the passage of time). In our case, policy interventions to modify EE inputs are statistically incongruent with the expected outcomes, because the inputs are highly persistent but the outputs are volatile over time. Moreover, manipulating the policy levers relating to the EE inputs is therefore unlikely to have any causal effect on the outcome (region-level HGF shares). While EE is emerging as a popular approach for practitioners and policymakers, nevertheless the research field remains young and underdeveloped, and the growing interest in the area should be met with stronger theoretical development and more rigorous empirical scrutiny (Malecki, 2018; Brown and Mawson, 2019). We hope to contribute to the broader task of rigorously evaluating EE policy with a view to making it more effective.

References

Abootorabi, H., Wiklund, J., Johnson, A. R., & Miller, C. D. (2021). A holistic approach to the evolution of an entrepreneurial ecosystem: An exploratory study of academic spin-offs. Journal of Business Venturing, 36(5), 106143.

Acs, Z.J., Autio, E., Szerb, L., 2014. National Systems of Entrepreneurship: measurement issues and policy implications. Research Policy, 43, 476–494.

Aldrich, H. E., & Ruef, M. (2018). Unicorns, gazelles, and other distractions on the way to understanding real entrepreneurship in the United States. Academy of Management Perspectives, 32(4), 458-472.

Arora, A., Belenzon, S., & Sheer, L. (2021). Knowledge spillovers and corporate investment in scientific research. American Economic Review, 111(3), 871-98.

Autio, E., Kenney, M., Mustar, P., Siegel, D., & Wright, M. (2014). Entrepreneurial innovation: The importance of context. Research Policy, 43(7), 1097-1108.

Bosma, N., Content, J., Sanders, M., Stam, E., (2018). Institutions, entrepreneurship, and economic growth in Europe. Small Business Economics 51, 483–499. https://doi.org/10.1007/s11187-018-0012-x

Breschi, S., & Malerba, F. (1997). Sectoral innovation systems: technological regimes, Schumpeterian dynamics, and spatial boundaries. Chapter 6, pages 130-156 in: Edquist, C (Editor), Systems of innovation: technologies, institutions, and organizations. Pinter: London, UK.

Brown, R., & Mawson, S. (2019). Entrepreneurial ecosystems and public policy in action: a critique of the latest industrial policy blockbuster. Cambridge Journal of Regions, Economy and Society, 12(3), 347-368.

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

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

Cooke, P., Gomez Uranga, M., & Etxebarria, G. (1997). Regional innovation systems: Institutional and organisational dimensions. Research Policy, 26, 475–491.

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

Daunfeldt, S. O., Johansson, D., & Halvarsson, D. (2015). Using the Eurostat-OECD definition of high-growth firms: a cautionary note. Journal of Entrepreneurship and Public Policy, 4(1), 50-56.

Demir, R., Wennberg, K., & McKelvie, A. (2017). The strategic management of high-growth firms: a review and theoretical conceptualization. Long Range Planning, 50(4), 431-456.

Delgado, M., Porter, M. E., & Stern, S. (2010). Clusters and entrepreneurship. Journal of Economic Geography, 10, 495–518.

Fallick, B., Fleischman, C. A., and Rebitzer, J. B. (2006). Job-hopping in Silicon Valley: some evidence concerning the microfoundations of a high-technology cluster. Review of Economics and Statistics, 88(3), 472–481.

Feldman, M. P. (2001). The entrepreneurial event revisited: firm formation in a regional context. Industrial and Corporate Change, 10(4), 861-891.

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

Freeman, C (1995). The 'National System of Innovation' in historical perspective. Cambridge Journal of Economics 19, 5-24.

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. (2001). Co-operation in regional innovation systems. Regional Studies, 35(4), 297–307

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

Furman, J. L., Porter, M. E., & Stern, S. (2002). The determinants of national innovative capacity. Research policy, 31(6), 899-933.

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. Research policy, 29(2), 109-123.

Gilson, R. (1999). The Legal Infrastructure of High Technology Industrial Districts: Silicon Valley, Route 128, and Covenants Not to Compete. New York University Law Review 74 (3), 575-629.

Grover Goswami A., Medvedev D., Olafsen E. (2019). High-Growth Firms: Facts, Fiction, and Policy Options for Emerging Economies. World Bank. Washington, DC.

Hall, B., Helmers, C., Rogers, M., & Sena, V. (2014). The choice between formal and informal intellectual property: a review. Journal of Economic Literature, 52(2), 375-423.

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

Henrekson, M., Johansson, D. and Stenkula, M. (2010). Taxation, Labor Market Policy and High-Impact Entrepreneurship, Journal of Industry, Competition and Trade 10(3): 275-296.

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

Isenberg, D., & Onyemah, V. (2016). Fostering scaleup ecosystems for regional economic growth (innovations case narrative: Manizales-Mas and Scale Up Milwaukee). Innovations: Technology, Governance, Globalization, 11(1-2), 60-79.

Lafuente, E., Ács, Z. J., & Szerb, L. (2021). A composite indicator analysis for optimizing entrepreneurial ecosystems. Research Policy, 104379.

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

Lundvall, B.-A. (ed.) 1992. National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning, London, Pinter.

Malecki, E. J. (2018). Entrepreneurship and entrepreneurial ecosystems. Geography compass, 12(3), e12359. https://doi.org/10.1111/gec3.12359

Malerba, F. (2002). Sectoral systems of innovation and production. Research policy, 31(2), 247-264.

Moretti, E. (2021). The effect of high-tech clusters on the productivity of top inventors. American Economic Review, 111(10), 3328-75.

Nightingale P., Coad A., (2014). "Muppets and Gazelles: Ideological and Methodological Biases in Entrepreneurship Research." Industrial and Corporate Change 23 (1), 113–143.

North, D., (1990). Institutions, Institutional Change and Economic Performance. Cambridge University Press.

Radosevic, S., & Yoruk, E. (2013). Entrepreneurial propensity of innovation systems: Theory, methodology and evidence. Research Policy, 42(5), 1015-1038.

Rocha, A., Brown, R., & Mawson, S. (2021). Capturing conversations in entrepreneurial ecosystems. Research Policy, 50(9), 104317.

Shumway, R. H., & Stoffer, D. S. (2017). Time Series Analysis and Its Applications: With R Examples. 4th Edition. Springer Verlag.

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

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

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

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

Vedula, S., Kim, P.H., 2019. Gimme shelter or fade away: the impact of regional entrepreneurial ecosystem quality on venture survival. Industrial and Corporate Change 28, 827–854. https://doi.org/10.1093/icc/dtz032.

ONLINE SUPPLEMENTARY MATERIALS

Appendix OSM.1: Output from the simulation model

Appendix Table OSM.1.1: summary statistics for the simulated data.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
err1	4200	-0.004	1.003	-3.741	-0.688	0.69	3.555
err2	4200	-0.003	1.008	-3.688	-0.683	0.698	3.295
err3	4200	-0.001	0.995	-2.991	-0.664	0.679	3.49
err4	4200	0.002	0.991	-3.131	-0.673	0.662	3.657
err5	4200	-0.012	0.993	-4.024	-0.695	0.665	3.888
x1	4200	0.341	8.923	-21.466	-5.796	6.232	26.428
x2	4200	-0.411	9.924	-30.702	-6.97	5.51	28.785
x3	4200	-1.349	10.083	-27.302	-7.811	2.79	36.634
x4	4200	-0.127	8.667	-32.816	-4.128	5.378	25.987
x5	4200	0.524	8.948	-21.597	-5.609	5.274	30.112
y	4200	15097.359	420884.13	-4938690.11	-2253.286	2302.854	7055488.43
y_ihs	4200	0.141	8.915	-16.106	-8.413	8.435	16.462

Notes: N=21 simulated regions, t = 200 time periods.

Appendix Table OSM.1.2: Correlations between the 5 dimensions x_k.

	x1	x2	x3	x4	x5
x1	1	0.34	0.57	0.37	0.44
x2	0.34	1	0.43	0.59	0.24
x3	0.57	0.43	1	0.32	0.49
x4	0.37	0.59	0.32	1	0.18
x5	0.44	0.24	0.49	0.18	1

Figure OSM.1.1: Time series of y for the 21 simulated regions, 200 time periods. The data are extremely skewed, which leads us to prefer the IHS transformed variable (shown in Figure OSM.1.2).

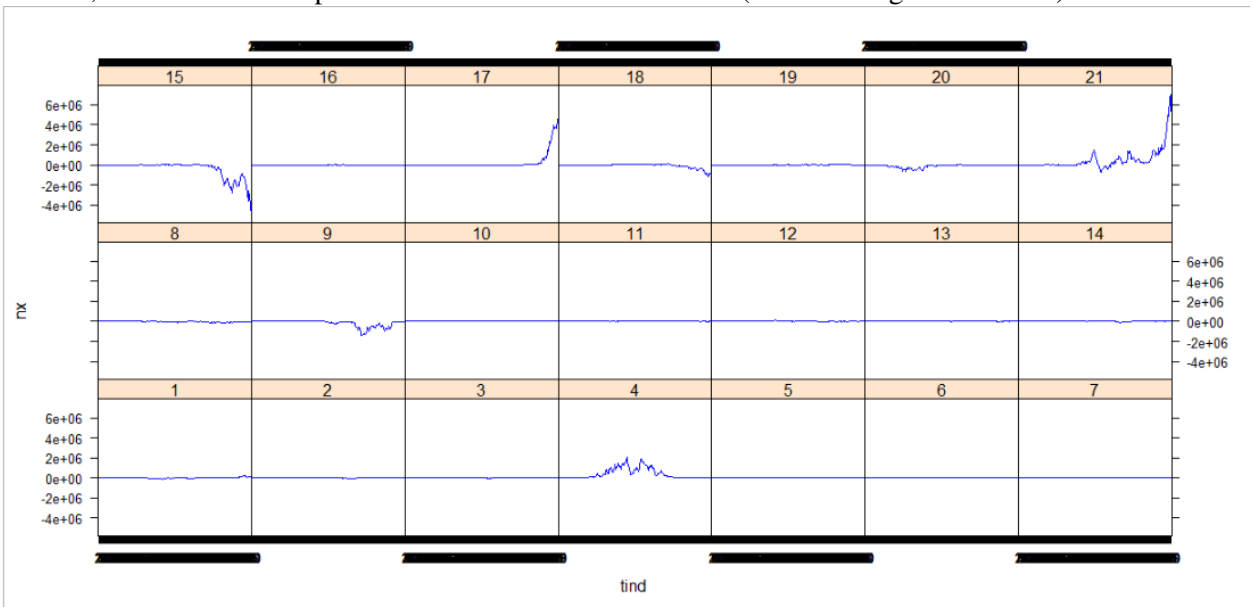
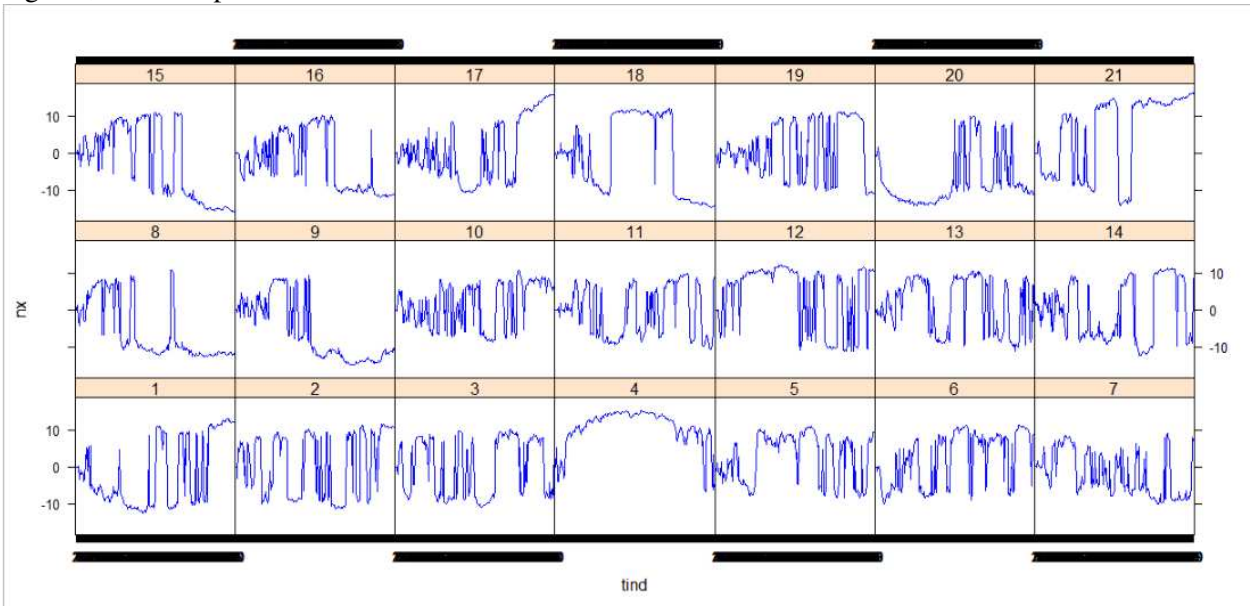


Figure OSM.1.2: Time series of y_IHS (i.e. the IHS-transformed version of y) for the 21 simulated regions, 200 time periods.



Appendix OSM.2: Extra results for persistence of HGF shares

Table 3. Pooling observations over time, full sample (1-2) and excluding single period (3-10).

	Dependent variable:									
	HGE_empl_t Full (1)	HGE_sales_t Full (2)	HGE_empl_t Exclude 13-16 (3)	HGE_sales_t Exclude 13-16 (4)	HGE_empl_t Exclude 10-13 (5)	HGE_sales_t Exclude 10-13 (6)	HGE_empl_t Exclude 07-10 (7)	HGE_sales_t Exclude 07-10 (8)	HGE_empl_t Exclude 04-07 (9)	HGE_sales_t Exclude 04-07 (10)
HGE_empl_t_1	-0.095		-0.162***		-0.063		-0.249***		0.666***	
	(0.066)		(0.045)		(0.097)		(0.058)		(0.189)	
HGE_sales_t_1		-0.258***		-0.271***		-0.249***		-0.482***		0.426***
		(0.040)		(0.039)		(0.035)		(0.077)		(0.077)
Constant	0.043***	0.122***	0.041***	0.114***	0.041***	0.117***	0.054***	0.156***	0.021***	0.071***
	(0.003)	(0.006)	(0.003)	(0.006)	(0.005)	(0.006)	(0.003)	(0.012)	(0.005)	(0.007)
Observations	84	84	63	63	63	63	63	63	63	63
R2	0.015	0.186	0.102	0.295	0.006	0.198	0.100	0.442	0.237	0.306
Adjusted R2	0.003	0.176	0.087	0.284	-0.010	0.185	0.085	0.433	0.224	0.294
Residual Std. Error	0.017 (df = 82)	0.028 (df = 82)	0.012 (df = 61)	0.025 (df = 61)	0.019 (df = 61)	0.029 (df = 61)	0.017 (df = 61)	0.025 (df = 61)	0.015 (df = 61)	0.019 (df = 61)
Note:										*p<0.1; **p<0.05; ***p<0.01

> |

Appendix OSM.3: NUTS 3 regions in Croatia and Slovenia

Figure: NUTS 3 regions of Croatia and Slovenia

Sources: https://en.wikipedia.org/wiki/Counties_of_Croatia and https://en.wikipedia.org/wiki/Statistical_regions_of_Slovenia

Appendix OSM.4: R code for the simulation model

```
set.seed(1)   # set the random seed
library(plm)
library(dplyr)

# generate the panel data frame
# ID: we have 21 NUTS 3 regions; and let's say 100 time periods
timeperiods <- 200
number_of_id <- 21
time <- rep((seq(1:timeperiods)),number_of_id)
id <- rep(1:number_of_id, each=timeperiods)

# set the persistence of x (1.00, 0.95, 0.90, ...)
xpers = 0.8

# error terms e that are correlated, using mvtnorm
library(mvtnorm)
ercov = 0.8  # this is our parameter for the covariance of the error terms
sigma_e <- matrix(
c(1,  ercov,ercov,ercov,ercov,
  ercov,  1,  ercov,ercov,ercov,
  ercov,ercov,  1,  ercov,ercov,
  ercov,ercov,ercov,  1,  ercov,
  ercov,ercov,ercov,ercov,  1), ncol=5)
err <- rmvnorm(n=length(id), mean = rep(0, nrow(sigma_e)), sigma=sigma_e)
colMeans(err)
var(err) # sanity check: check the covariance of the k error terms -- yes, fine.
cor(err)


# put these variables as a data frame
rp <-  data.frame(id, time, err) # rp: Regional Persistence
head(rp)
attach(rp)
# create lags in panel format
rp <- pdata.frame(rp, index=c("id","time"))
class(rp)


rp <- rename(rp, err1 = X1)
rp <- rename(rp, err2 = X2)
rp <- rename(rp, err3 = X3)
rp <- rename(rp, err4 = X4)
rp <- rename(rp, err5 = X5)


# rp <- rp[with(rp,order(id,time)),]
class(rp)

rp$x1 <- NA
rp$x2 <- NA
rp$x3 <- NA
rp$x4 <- NA
rp$x5 <- NA

# initializing for each id
for(j in 1:number_of_id) {
  rp$x1[((j-1)*timeperiods)+1] <- 0
  rp$x2[((j-1)*timeperiods)+1] <- 0
  rp$x3[((j-1)*timeperiods)+1] <- 0
  rp$x4[((j-1)*timeperiods)+1] <- 0
  rp$x5[((j-1)*timeperiods)+1] <- 0
}

for(j in 1:number_of_id) {
for(i in (((j-1) * timeperiods)+2):(j * timeperiods)) {
  rp$x1[i] <-   xpers * rp$x1[i-1] + rp$err1[i]
  rp$x2[i] <-   xpers * rp$x2[i-1] + rp$err2[i]
  rp$x3[i] <-   xpers * rp$x3[i-1] + rp$err3[i]
  rp$x4[i] <-   xpers * rp$x4[i-1] + rp$err4[i]
  rp$x5[i] <-   xpers * rp$x5[i-1] + rp$err5[i]
  }
```

```
}

plot(rp$x1)
plot(rp$x2)
plot(rp$x3)
plot(rp$x4)
plot(rp$x5)


#####--- generate the outcome y

rp$y <- rp$x1 * rp$x2 * rp$x3 * rp$x4 * rp$x5
plot(rp$y)

# since it is skewed, and we can't take logs (negatives) we calculate the IHS: inverse hyperbolic sine
rp$y_ihs = asinh(rp$y)
plot(rp$y_ihs)

# install.packages("vtable")
library(vtable)
vtable(rp)

# sanity check - AR should be found using arima
arima(rp$x1,    order = c(1,0,0))
arima(rp$y_ihs, order = c(1,0,0))
# yes, no problem here

# sanity check: autocorrelation of x1 ... x5, should be highly persistent
library(plm)  # for plm regression on panel data (pooling means pooled OLS)
library(Hmisc) # for Lag
model1 <- plm(rp$x1 ~ Lag(rp$x1,1),
        data = rp, index = c("id","time"), model = "pooling")
summary(model1)

sanity check: plm::lag gives the same results
model2 <- plm(rp$x1 ~ plm::lag(rp$x1,1), data = rp, index = c("id","time"), model = "pooling")
summary(model2)

model3 <- plm(rp$y_ihs ~ plm::lag(rp$y_ihs,1), data = rp, index = c("id","time"), model = "pooling")
summary(model3)

model4 <- plm(rp$y ~ plm::lag(rp$y,1), data = rp, index = c("id","time"), model = "pooling")
summary(model4)

# y is very skewed
hist(rp$y)
plot(density(rp$y))
# It is natural that we should take IHS (or logs) of y, because it is generated by a multiplicative process


# CONCLUSION:
# x1-x5 are random walks, and are highly persistent
# Y, and IHS-transformed y, are highly persistent, see here for OLS on y_ihs:
summary(model3)
```