Drivers of high-growth businesses across U.S. states: policy implications[1]

Nadine Jeserich[2]
GrowthEconomics, Inc.

Hugh Kelley[3]
National University of Ireland, Galway

Graham Toft[4]
GrowthEconomics, Inc.

Gregg Cole[5]
Edward Lowe Foundation

April 2012

Keywords: entrepreneurship, innovation policy, high-growth, determinants, US states

---

[1] Presented at the Kauffman Foundation International Research & Policy Roundtable, Liverpool, UK, March 2012. We would like to thank the participants of the roundtable for their valuable feedback.
[2] Corresponding author; Head of Research, GrowthEconomics, Inc. 2425 Gulf of Mexico Dr., Longboat Key, FL 34228, USA, Email: nadine@growtheconomics.com
[3] Associate Professor, Department of Economics, Cairnes School of Business and Economics, National University of Ireland, Galway, Galway, Ireland. Email: hugh.kelley@nuigalway.ie. Tel.:00353.91.49.5087, Fax: 00353.91.52.4130.
[4] Owner and President, GrowthEconomics, Inc. 2425 Gulf of Mexico Dr., Longboat Key, FL 34228, USA, Email: graham@growtheconomics.com
[5] Edward Lowe Foundation, Institute for Exceptional Growth Companies, Edward Lowe Foundation, 58220 Decatur Road, P.O. Box 8, Cassopolis, MI 49031-0008, USA

Abstract

The lack of consistent empirical evidence on drivers of high growth businesses and on the success of business growth policies leaves policy makers uncertain and apprehensive about high growth strategies. This study tests a wide-ranging array of regional policy relevant drivers using a unique combination of principal components and regression analysis, which allows us to establish their causality with respect to the share of high growth establishments (HGE) in all 50 U.S. states at the end of the last business cycle in 2007. For both employment and sales defined HGEs, the results highlight the importance of domestic and international labor supply as well as the role of the existing industrial structure. The analysis furthermore highlights several contrasts in determinants HGEs depending upon definitions and data years. The results confirm that effective and enabling business growth policies need to be flexible in their approach, adaptable over time and broad-based across multiple and diverse policy areas.

1. Introduction

In the last decade, the attention of policy makers has shifted increasingly to the role of high growth firms and enabling policies (OECD, 2002 and 2010; NESTA, 2011). The limitation of most earlier empirical research on high-growth performance has become increasingly problematic in this respect as policy choices have relied on insufficient quantitative evidence resulting in either too general (e.g. quality of life) or too specific policy targets (e.g. high-tech small and medium enterprises).

This paper quantifies the extent to which variations in state-level entrepreneurial conditions and policies can account for the cross-state variation in the share of U.S. high-growth businesses. As a federal system of government with decentralized state economic development responsibilities and long-established, technology, innovation and entrepreneurial policies, careful empirical examination of their performance in supporting and attracting high growth businesses can provide policy insights of international relevance and utility.

Since Birch (1979**),** empirical studies have tried to identify the type of firms that have the most significant impact on economic performance. Few studies have investigated the drivers of different stages of firm growth, or drivers of HGF performance or prevalence. Even fewer studies have empirically addressed the role of a whole range of public policies in support of high growth businesses or their effectiveness. Some empirical studies have focused providing descriptive statistics such as high-growth business shares across regions or industries (Acs et al., 2008; Hoffman and Junge, 2006; Teruel and deWit, 2011). Nearly absent are studies investigating high growth firms or more generally, entrepreneurial activity, that include tests for the direction of causality of the potential driver variables (an exception is Kreft and Sobel, 2005).

This study contributes to the empirical literature by investigating the conditional policy relevant and causal drivers of U.S. state level shares of high growth businesses, importantly moving beyond general statements of association. In particular, we investigate what causes the cross-sectional variation in the share of these businesses. Such a focus on the aggregate prevalence of a firm type, rather than focusing on individual or representative firm growth rates, will be more instructive to policy makers in devising generalizable and effective economic development policies. Specifically, by focusing on the share of high-growth businesses rather than on drivers of individual firm growth rates, a wider range of public policies can be considered. To the extent that business growth rates are volatile and non-enduring, especially for young/small firms (Hoelzl, 2009), the policy focus needs to be targeted on time relevant drivers that create an environment that allows as many firms as possible to relocate, emerge and remain with as many high growth spurts as possible and with long term sustainable profitability thereafter. To provide an empirical contribution with broader relevance, we employ a uniquely large database (close to 150 variables) comprising the full cross-section of the 50 US states, representing a plethora of institutional structures and policy approaches. Finding common drivers of high growth businesses with this approach will provide a more precise and adaptable policy target, which can be more easily generalized to other regions or countries.

This study utilizes a unique data set, developed by GrowthEconomics, Inc. for state benchmarking purposes, in combination with multiple sophisticated statistical techniques to investigate the following research questions:

1. Do states with comparable shares of HGE show similarities with respect to their innovation and entrepreneurship performance?
2. Do those states cluster geographically to form 'innovation/high-growth regions'?
3. Are there innovation and entrepreneurship policy elements that significantly cause the occurrence of HGE, or are more fundamental framework conditions such as physical infrastructure or amenities the key drivers?
4. And to what extent are identified causal drivers stable across time?
5. Are there policies that show positive interactions or counteract each other's impact in the short run?

To maximize robustness of the results, our analysis is repeated for a set of years and with multiple definitions of the dependent variable, high-growth establishments (HGE). Most importantly, our method provides evidence for the direction of causality using classic Granger causality tests, which is so often neglected. We focus on a time frame of less than five years to identify short-term/fast-impact policy targets. As exceptional firm activity is very short lived, policies have to continuously adapt and it is therefore more useful for policy makers to conduct a wide cross section analysis with a larger set of explanatory variables that can repeated and updated every few years. The results provide key insights into the causal effect of common empirically identified policy drivers such as small business innovation support, amenities and infrastructure. We also consider less investigated policy relevant variables such as business liability costs, and the possible interaction of different policies on the share of high growth businesses in an economy. As many of the explanatory variables can be considered "intermediate" outcomes between actual policies and resulting firm and economic growth, such as educational attainment, this approach leaves the flexibility to investigate different combinations of policies to achieve particular outcomes.

Section 2 of the paper proceeds with an overview of the determinants of firm growth and high-growth firms that has informed the selection of explanatory variables and Section 3 provides a general analytical framework and description of the data sources. Section 4 presents the empirical methodology and Section 5 summarizes the statistical results regarding the drivers of the share of high-growth businesses across states. Based on these results, Section 6 discusses the policy implications in the context of current international policy interest. Finally, Section 7 presents concluding remarks.

2. Literature review

A key contrast of this study to the prevalent literature on high growth businesses is its focus on the establishment level (see Section 3.1). The vast majority of empirical studies on high growth businesses summarized below have used the firm as their level of analysis.

While Birch's research spurred investigations into small firms as job contributors, later studies have observed the importance of firm age over size for job creation (Praag and Versloot, 2008; Audretsch and Keilbach, 2004; Haltiwanger et al., 2010; Stangler and Kedrosky, 2010; Henrekson and Johansson 2010). However, quantifying the importance of firm attributes such as age and exceptional growth performance, i.e. high-growth firms (HGF) or "gazelles", is still in development, and the literature is plagued with many methodological limitations (Delmar et al., 2003).[6] Praag and Versloot (2008) review empirical studies of the impact of small, young firms and find positive associations with employment growth, innovation, and productivity growth. Further, a recent survey of studies (Henrekson and Johanssen, 2010) identifies that it is not so much either small or young firms but rather a relatively small number of high-growth firms that are responsible for all or a large share of net new jobs, and generally increased economic performance (Birch and Medoff, 1994; Storey, 1994; Daunfeldt et al., 2010). At the same time, heightened economic development has shifted practitioner interest in the U.S. to the stage of firm growth as a core strategy.[7]

Micro-level determinants of firm growth rates and high growth firms

The management literature has identified a variety of significant firm level characteristics as key growth drivers, such as CEO/founder characteristics, organizational structure and management style (Storey 1994; Achtenhagen et al. 2010; Barringer et al. 2005; Parker et al. 2010) though the explanatory power of these variables is unclear (Coad 2009). Furthermore, different types of high-growth firms, defined by, e.g. sales rather than employment growth, show different growth drivers (Achtenhagen et al., 2010). Recent empirical investigations using techniques such as quantile regression to analyze the drivers of firm growth across the full growth distribution highlight that different factors might explain fastest growing firms compared to the more moderately growing firms (Coad and Rao, 2008; Hoelzl, 2009; Stam and Wennberg, 2009), i.e. factors influence not just the mean growth rate but the dispersion. However, though these internal/firm-level factors are key to explaining firm-level entrepreneurial activity and advance a comprehensive theoretical framework for the field, they do not provide sufficient grounds for a comprehensive policy framework.

A limited number of empirical studies have specifically investigated the characteristics of high growth businesses. The first policy focus came from the early empirical findings that *small* high-growth firms are key job contributors and that therefore policies should focus on promoting exclusively small businesses. Later research established that high-growth firms tend to be younger than non-high-growth firms, though with no clear consensus on the typical size (Henrekson and Johahnssen, 2010), shifting the policy focus towards nascent entrepreneurs and startup activities. Similarly, studies have tried to narrow down whether high-growth firms are more prevalent in certain industries, which would support the policy practice of industry targeting. However,

6 Studies have focused almost exclusively on the contribution to job growth (Storey, 1997; Henrekson and Johanssen, 2010, Schreyer, 2000) with a few exception (see Daunfeldt et al. 2010 for an overview).

7 Edward Lowe Foundation, 2011. Second-stage entrepreneurs. http://www.edwardlowe.org/secondStage

although their prevalence certainly differs between industries, they seem to be present across all sectors (Daunfeldt et al., 2010). These very general firm characteristics therefore do not provide clear policy guidelines.

Macro policies and institutional factors

A few studies have focused on establishing the possible policy and institutional determinants of variations in high-growth firms. Davidsson and Henrekson (2002) correlate the presence of high growth firms and startups in Sweden with the strength of institutional factors compared to other countries, such as taxation on labor, savings and wealth accumulation, "public production of goods and services" (crowding out of private activity/monopolization), and labor market regulations. Other studies have investigated interspatial variations of high-growth firms and potential policy drivers. There are similar to studies of entrepreneurial activity represented by self-employment rates, startup rates, or of attracting and retaining high-tech businesses, identifying broadly issues of finance, regulations, infrastructure, and human capital (Reynolds et al., 1994, Armington and Acs, 2002; Audretsch and Fritsch, 2002, Audretsch and Keilbach, 2004; Goetz and Freshwater, 2011; Lee, 2011). Some have focused on identifying clear spatial patterns and linking them to a few general spatial characteristics such as industry structure, urbanization or startup activity rates (Stam, 2005, Hoffman and Junge, 2006; Acs et al., 2008). Friedman's (1995) study of 200 urban areas in the 1980s points among other factors to industrial diversity, local venture capital, educational attainment and infrastructure as key drivers of the presence of young high-growth firms. Teruel and deWit (2011) preliminary investigation of the share of high growth firms across 17 countries points to important educational and institutional factors but based on a limited dataset and only a handful of independent variables.

Some additional insights into potential policy relevant drivers for high growth firms can be gained from the broader field of entrepreneurship studies, since high-growth firms can be seen as an outcome of successful entrepreneurial activity (see Daunfeldt et al. 2010 for a recent overview of the literature). Empirical entrepreneurship studies indicate the importance a number of policy factors. The following recur in the literature, while not confirming the direction of impact: general entrepreneurial climate and institutional environment (Fritsch and Mueller, 2007; Hall and Sobel, 2008, Henrekson, 2005; Henrekson and Johanssen, 2009); certain forms of laws and regulations (van Stel et al., 2007; Klapper et al., 2006); access to capital (Goetz and Rupasingha, 2009; Grilo and Thurik, 2005), taxes (Gordon, 1998; Cullen and Gordon, 2002; Bruce and Deskins, 2010), local amenities (Goetz and Rupasingha, 2009), low administrative burdens (van Stel and Stunnenberg, 2004), support for R&D/innovative activity (Fritsch and Mueller, 2007; Hoelzl, 2009), human capital (Armington and Acs, 2002), business support infrastructure (McFarland et al., 2010), and physical and digital infrastructure (Goedhuys and Sleuwaegen, 2010, Coad, 2009).

These empirical studies tend to look at similar policy relevant areas that might drive entrepreneurship and high growth firms, but the variety of definitions used, the specific choice of explanatory variables, and the choice of statistical techniques result in a mixed conclusion on the significance and certainly direction of impact of specific variables. However, they do give guidance, together with feedback from practitioners and policy makers, as to topic areas to be included in a comprehensive investigation of the kind pursued in the previous state benchmarking work of GrowthEconomics, Inc., and Section 3.2 of this paper.

3. Data

3.1 Identifying high-growth businesses

Data to investigate the state share of HGEs were obtained from the National Establishment Time Series (NETS) by Walls & Associates, provided by the Institute for Exceptional Growth Companies (IEGC) at the Edward Lowe Foundation. The underlying Dun & Bradstreet business information was first used by Birch (1981), and this reconstructed dataset by Walls & Associates is a radical improvement to the original raw data (Kolko and Neumark, 2007; Neumark et al., 2011). One highly desirable feature of the NETS database is that it covers the largest set of establishments across all ownership types and industries. This reflects the fact that it is designed to capture the universe rather than a sample of establishments. The dataset under investigation for the period 2003 to 2009 includes information across all 50 states on approximately 17 to 22 million establishments per year.

Definitions

The study focuses on both standalone establishments and branch facilities of larger firms that reflect purely organic growth and are therefore mostly small.[8] An establishment has a distinctive line of business (producing a good or service) at an identified location. The term 'firm' refers to a corporate entity, either standalone, multi-facility or multi-national. Penrose (1959) first pointed out that firms that grow organically will tend to show a smoother growth pattern (Wiklund and Davidsson, 1999; Levie, 1997) and are more likely to present genuine job creation (Delmar and Davidsson, 2003).[9] Only continuing firms that existed during the whole period under investigation are included. Much of the research to date has been firm-focused. From the perspective of national and sub-national growth strategy, accelerating the presence, survival and sustainability of business establishments is of primary concern and thus the focus of this analysis.

A more complex methodological decision relates to the choice of outcome measures and definitions on what constitutes a high growth business. A recurring criticism of studies on the impact or determinants of high-growth firms has been the inconsistent or imprecise choice of empirical methodology (Delmar 1997, Davidsson et al. 2006, Delmar et al. 2003). Each outcome measure implies a certain theory or approach to firm growth (Wiklund, 1998) and each has its own advantages and disadvantages (Davidsson et al, 2006). This study analyzes separately the most common performance measures, sales and employment growth (Achtenhagen et al., 2010, Weinzimmer et al., 1998) and reports results separately.

---

[8] The median of the median employment size across our 2006 and 2007 data on HGEs, which does not include startups, is 4-5 employees.

[9] However, acquired growth is an important mechanism for the efficient reallocation of scarce economic resources and high growth businesses through acquired growth will present a large share of all these businesses, (Delmar et al. 2003).

This study in its first research phase will test a high growth definition of an annual average growth of 20% or more to provide stronger comparability across time periods than an absolute threshold of for example the top 5% of businesses (Hoelzl 2009).[10] Using an annual average acknowledges some of the likely short-run fluctuations in firm growth (Schreyer, 2000). Even though 5 years usually present a tipping point for firm survival rates, due to our shorter time series of explanatory variables in a tradeoff for larger variable variety, we focus primarily on a period of three years of growth with three years of lags in the explanatory variables. This is consistent with evidence that above average firm growth rates are very short-lived phenomena. It represents a valid starting point for short-term/immediate policy impact analysis, as effective policies will most likely have to continuously adapt through time (Parker et al., 2010).[11]

3.2 Determinants of high-growth businesses

The source of explanatory variables is GrowthEconomics' state benchmarking database developed over the last decade under the guidance of policy makers and business associations on key drivers of economic growth and entrepreneurial activity. The choice of these variables is less driven by a narrow theoretical framework than by the opinion of practitioners and empirical studies (see Section 2). The purpose of this study is to investigate the widest possible set of explanatory variables in a field that has been limited to very general policy related measures such as tertiary educational attainment or to a very small set of specific variables such as the impact of small business grants. We are presupposing entrepreneurial activity, as reflected in high-growth establishments, as an important input to state economic growth. We then analyze the various policy-relevant state-level drivers of this intermediate variable based on detailed measures of business climate, research and innovation activity, access to capital, legal and regulatory environment, education and workforce, and physical and digital infrastructure (see Appendix Table 1 for a full list of independent variables). Where possible, explanatory variables are policy outcomes (university spin-offs) rather than direct policy measures (government expenditures), in order to maintain more flexibility and transferability of results to how those outcomes can be achieved. The information is collated from a variety of established public and private data sources and continuously updated and revised. For the current analysis, the GrowthEconomics database has been extended by several policy related variables after a thorough review of literature on firm growth. All variables are scaled to adjust for the size of the respective state economy, e.g. per capita, per gross domestic product, per worker, etc., to create comparable metrics for benchmarking purposes. The consequent collinearity of the some of the explanatory variables is addressed by our use of the principal components method, see Section 4. The industry

---

[10] Absolute change in the outcome measure would bias results in favor of large firms, while a percentage measure would advantage small businesses. A combined, measure though lacking in conceptual foundation (Davidsson and Wiklund, 2000), is often used as a compromise (Hoelzl, 2009; Ace et al., 2008; Coad and Hoelzl, 2009).

[11] Summary statistics on the dependent variable indicate clear differences in the definitions and years under observation. The share of employment defined HGEs in 2007 is over two standard deviations from zero (almost 3 stdev in 2006), while the share for the sales defined HGEs is just around two standard deviations from zero (more than 2 stdev in 2006).

structure and the median size and age of high growth establishments in each state were added as control variables as is common in studies of business growth.[12]

We acknowledge that some explanatory variables will have a more direct impact on high growth establishments than others. There might therefore be a certain hierarchical structure embedded in the independent variables that might lead to an aggregation bias, apparently stronger relationships between variables than are actually the case, and probable violation of the assumption of independence of observations. At this stage of the research, this is not explicitly taken into account through, e.g. with the use of a hierarchical linear modeling approach, but we are controlling for most of the potential presence of spurious correlations with the use of our unique principal components approach (see Section 4).

4. Empirical methodology

The goal of our empirical methodology is to derive inferences about how a large set of independent-variable metrics **X** drive a small set of dependant-variable performance criteria, **Y**. Ideally one could simply regress *Y* on *X* and obtain a set of regression parameters from the raw data b^{Raw}, i.e.

$$Y = b^{\mathrm{Raw}} \cdot X + \varepsilon \tag{1}$$

Unfortunately, there are a number of problems with such a large dataset that must be addressed. Most obviously, our number of predictors greatly exceeds the size of our 50-state cross-section. In order to achieve our research goal of identifying unbiased estimates of the degree of causality among drivers and high growth entrepreneurial activity we sequentially employ six statistical techniques. First, in order to allow direct comparison of the relative effects of the various 'raw' metric drivers **X** and account for the differences in magnitude, the presence of outliers, and different distributional characteristics we normalize all data using median- and z-scores. This yields a panel dataset denoted **x** and a vector of dependant variables denoted **y**. Second, controlling for the likely presence of collinearity among our normalized drivers and reducing our large set of predictors into a smaller set to allow for sufficient statistical power for inference is accomplished by principal components analysis. Components **c(x,w) are** based on the normalized metric data **x** and a set of principal component weights matrix **w**. Third, we employ linear regression analysis which relates the derived principal components **c(x,w)** to the normalized dependent variable **y. T**his yields a set of estimated parameters **b**. Fourth, to recover the importance of the association among drivers **x** and the dependant variable **y**, we employ a technique which combines the estimated component parameters **b** and component weights **w** and then sums the product of these parameters across components. Fifth, to identify whether a causal relation exists among important drivers **x** and the dependant variable **y** we conduct a Granger causality test comparing each driver within **x** to **y**. Finally, we conduct robustness checks to investigate the stability of the causal relations.

[12] As this dataset is continuously updated, some of the results with differ slightly from an earlier draft of the paper discussed at the Liverpool Roundtable.

4.1 Data normalization

In order to normalize our data we create both modified median- and z-scores from our raw data. Although **X** represents $n = 50$ states/rows by $K = 143$ driver/column panel, we suppress individual metric subscripts below for clarity. Analytically, a z-score provides the deviation of a raw metric variable within the panel X from its mean and is then normalized by its standard deviation. This yields mean zero standard variance variables.

$$x^z = \left(X - \overline{X}\right)/\sigma_X \tag{1}$$

$\overline{X}$ is the across state mean driver value for a column in X. The use of a z-score is a common method for normalizing data but may not be appropriate if the source data do not satisfy normality requirements. To account for this possibility, we also employ a version of a median score that does not place any restrictions on the distributional properties of the raw data, Rousseuw and Croux (1993). The maximum efficiency variant used is,

$$x = \left(X - median(X)\right)/MAD \tag{2}$$

$$MAD = median(\eta \cdot |\mathrm{X} - \mathrm{X}'|) \tag{3}$$

where |X-X'| represents the absolute difference of each states' driver value X from the value for all other states X' and η = 1.1926 is an efficiency enhancing constant. Taking the median of these differences intuitively allows MAD to perform a role similar to the standard deviation normalization in Eq. (1).

4.2 Principal components analysis

The usefulness of principal components can be demonstrated by considering how it deals with the problem created by multicollinearity. Assume a matrix of normalized driver metric values **x** which has *K* columns (different variables & lags) and *n* observations, but that **x** may have fewer than *K* truly independent sources of variation. The use of principal components is an attempt to extract from **x** the smallest set of independent variables that account for the majority of the variation in **x,** Green (2000). This process is a balancing act where one wishes to throw out as many independent variables (columns of **x**) as possible in order to eliminate the collinearity, while at the same time keeping enough drivers so that one can keep a reasonable amount of the variation present in **x**.

Importantly, it may be the case that there are interaction effects within normalized predictors **x** that are relevant for our dependant variable **y**. To account for this possibility the normalized driver dataset **x** is augmented to include a number of interaction terms as motivated by the extant literature described in Section 2. This yields a modified panel of driver variables denoted $\widetilde{x}$.

Algebraically, we can think of the principal component problem as finding the linear combination of the columns of $\widetilde{\mathrm{x}}$ that provides the best fit to all columns of $\widetilde{\mathrm{x}}$. This linear combination will be,

$$\mathrm{c_i} = \mathrm{w_i} \cdot \widetilde{\mathrm{x}} \tag{4}$$

where *i* ranges from 1 to K components, i.e. potentially as many components as there are columns (variables) in $\widetilde{\mathrm{x}}$. Also, w_i is a vector or list of *K* component weights from the matrix of component weights, which transform the data in $\tilde{x}$ into a particular principal component c. Of course, we hope to obtain fewer components *L* than the number of variables in $\tilde{x}$, otherwise we may not eliminate the collinearity problem. Notice that for each component *c*, all *K* driver metrics are potentially present, albeit adjusted by the factor loading/weight *w* that could be zero. If $L<K$ is chosen, in order to make this decision one typically considers the 'scree plot' which is a representation of the eigenvalues or the amount of variance each component explains. In a scree plot one looks for a point at which the additional or marginal variance explained by additional components starts to sharply fall off; i.e. one looks for the 'elbow'. An additional consideration is that for a cross section of $n = 50$ states one may select a set of principal components so that $L < n$ allowing sufficient statistical power to draw inferences from regression parameters relating *y* to the multiple components in *c*.

An important limitation of the PC method is that the data reduction / ortogonalization calculation is carried out without regard to any potential dependant variable of interest. This can result in one or a few non-causality related variables being retained because they are highly correlated with a larger number of actually causal variables. This occurs, as there is a tendency of the PCA method to try to reduce the number of variables in the dataset. The implication for our method is that a variable may pass a Granger causality test and appear causally related to our dependant variable because it is an instrument for multiple underlying causal drivers. Due to this well-known limitation, we additionally take a conservative position and suggest that at the very least this method, when combined with the other techniques described later, can be helpful for predicting variables. Such variables could be included in a standard regression analysis relating the normalized dependant variable to a small set of normalized metrics without employing the principal component composite predictors. However, first we describe how we might relate dependant variables to principal components.

4.3 Regression analysis of principal components

The third step in our statistical methodology is to quantify the extent to which the principal components $c(\tilde{x}, w)$ based on the normalized metrics and PC weights, is associated with the normalized dependant variable y. By substituting the PCs **c** for the normalized metric data $\tilde{x}$, we have eliminated any potential collinearity among predictors, by employing normalized data we have mitigated the effects of non-normally distributed variables and outlier values, and can directly compare regression parameters to determine important components since the components are based upon mean zero and standardized variance data. Most importantly, our analysis employing $L < n$ predictors allows sufficient statistical power, which is absent from Eq. (1) for performing inference tests. In contrast to Eq. (1) the regression is now,

$$y = b \cdot c(\tilde{x}, w) + \varepsilon \tag{7}$$

4.4 Calculating metric driver importance

The most unique aspect of our methodology is that we have multiple ways to infer the importance of the underlying drivers in $\tilde{x}$ despite the fact that they have been transformed into components **c**. Specifically, to deal with the limitation of principal component analysis described above, we construct two 'importance' measures, relative importance I and absolute importance $\tilde{I}$ for each metric driver within the panel $\tilde{x}$.

Consider that after identifying the principal components weights matrix **w** and obtaining the regression coefficients **b**, the relationship among drivers $\tilde{x}$, component weights **w**, regression coefficients **b**, and **y** is,

$$\hat{y} = \sum_{i=1}^{L} b_i \cdot \left(w_{i1} \cdot \tilde{x}_1 + w_{i2} \cdot \tilde{x}_2 + \cdots + w_{iK} \cdot \tilde{x}_K \right) \quad (8)$$

Since each metric potentially appears in each component, i.e. across *i* in the above summation, to determine the importance of any metric in the overall regression we need to sum the product of the regression coefficient for the PC and the PC weight for the metric for all principal components included in the regression and for which the regression parameter was statistically significant ($t \geq 1.96$). Specifically, the importance of any individual driver $k \subset K$ within $\tilde{x}$ is I_k,

$$I_k = \sum_{i=1}^{L} b_i \cdot w_{ik} \cdot \left(| t_{bi} | \geq 1.96 \right) \quad (9)$$

Importantly, since principal component weights and regression coefficients can be positive and negative, using this method identifies only the net effect of potentially important drives. The difficulty with this approach is that we may have important drivers that have both positive and negative effects, and by summing, we erroneously exclude them from the list of important drivers. To control for this possibility we also consider the sum of the product of absolute values of component weights and regression coefficients. This provides our absolute importance measure,

$$\tilde{I}_k = \sum_{i=1}^{L} | b_i | \cdot | w_{ik} | \cdot \left(| t_{bi} | \geq 1.96 \right) \quad (10)$$

Eq. (10) tells us about drivers that are important in an absolute sense and are statistically significant, but this measure will not tell us the direction of the net influence as in the former case. By considering both of these importance measures, we can more robustly identify the most statistically associated drivers $\tilde{x}$ with our dependant variable **y**. In addition, any difference in these measures indicates that effects may be canceling out in the across components summation process, and our net measure may underestimate their importance.

4.5 Granger causality analysis

The previous methodological steps correct for many statistical problems not accounted for in many earlier studies and provide a very robust list of drivers that have a strong association with our dependant variable of interest. However, without further analysis we can only make statements of association rather than causality. To make stronger statements about causality we employ the Granger causality test, Granger (1969), Sims (1972) on our standardized data to test for causality from an individual driver column in $\tilde{x}$ to y.

4.6 Robustness checks

In order to determine the robustness of our results regarding causality we perform eight types of robustness checks. Our baseline analysis calculates net importance by employing the median score normalized dependant and independent variables with principal component analysis, linear regression, net importance measures, and Granger causality tests with a target year for the dependant variable of 2007.

Our first robustness check calculates the absolute importance measure employing all other baseline conditions (Eq. 10). This tells us about the absolute importance of a driver, independent of the direction of effect. Any discrepancy among this and the baseline results informs us about the extent to which a driver may not appear important due to its impacts on the dependant variable switching signs and canceling out across time or predictor lags. Our second robustness check employs the z-score normalization method rather than the modified median score method, uses the net importance measure, and retains all previous baseline conditions. This can mitigate problems associated with binary data or data with large numbers of the same value, which would result in medians scores of zero after the normalization process. Any discrepancy among the baseline and z-score robustness check results indicates the baseline may have omitted important drivers due to lack of cross sectional variation due to common values among drivers. Our third check employs the non-normalized raw metric values when calculating the principal components, using the net importance measures. This test is used to investigate whether scale effects may have been responsible for results in other studies, which deviate from our results. Using raw values is most likely an inappropriate method given the sensitivity of principal components techniques to scale effects.

Our four, fifth, sixth and seventh robustness checks change the target year for the dependant variable to 2006 and conducts the baseline and checks one through three on the new 2006 dependant variable and preceding predictor data; this represents four tests because the baseline median score analysis on a previous year is now a robustness check. This allows us to test the stability of the identified relations across time, and may be better described as an element of a dynamic analysis.

The eighth and final check takes an altogether different approach as described at the end of Section 4.2. We perform a basic regression analysis relating the normalized dependant variables to a small set of normalized predictor variables ($\tilde{x}^{Basic}$) which satisfied one of three criteria. They either: 1) passed all seven previous robustness checks, passed the Granger causality test, and had importance scores lying outside two standard deviations of zero; or 2) they were variables that satisfied the previous criteria 1) for an additional analysis using one of the three presumed instrumental variables described in section 5.5 as dependant variables; or 3) they had importance scores just within one standard deviation of zero.

$$y = b^{\text{Basic}} \cdot \tilde{x}^{Basic} + \varepsilon \qquad (11)$$

This most strict robustness check allows us to distinguish variables that have a direct effect on the dependant variable from those who may be instruments of other variables. If the second set of variables (group 2) supplants the presumably instrumental variables in group 1 for which they appeared as important predictors, this is an indication of instrumental / indirect effects. The last set of variables included is expected to be insignificant in this regression test. This informs us about the extent to which our two standard deviation criteria for importance scores is a reasonable approximation of a standard error that can be used to identify drivers with an effect significantly different from zero. Integrating the results of this last test with earlier evidence allows us to distinguish direct effects from variables with indirect or instrumental effects, yields an approximation of statistical significance, and may provide results that are more intuitive. The final set of drivers identified after performing these tests are considered to be robust.

5. Results

Our results are based on an analysis of over 150 policy relevant metrics across 50 states available for the years 1997 to 2009; this provides a panel dataset with over 105,000 observations that can be used to address our research questions. The object of our analysis is to determine which groups of metrics were best able to account for cross state variation in the prevalence of high growth firms as defined by their sales and employment.[13]

5.1 Geographic and policy correlations for high-growth establishments

State groupings. Our first research question asked about commonalities across states in terms of innovation and entrepreneurship drivers associated with the share of HGEs. Three broad sets of performance indicators are taken from GrowthEconomics' state benchmarking data (each is a composite of six or more of the relevant metrics): Capital Formation, Research and Innovation, and Business Costs. A first-level uni-variate examination using the scatterplots below shows a weak relationship between each of these parameters and HGEs. The relationship is more definitive for HGEs defined by sales growth.

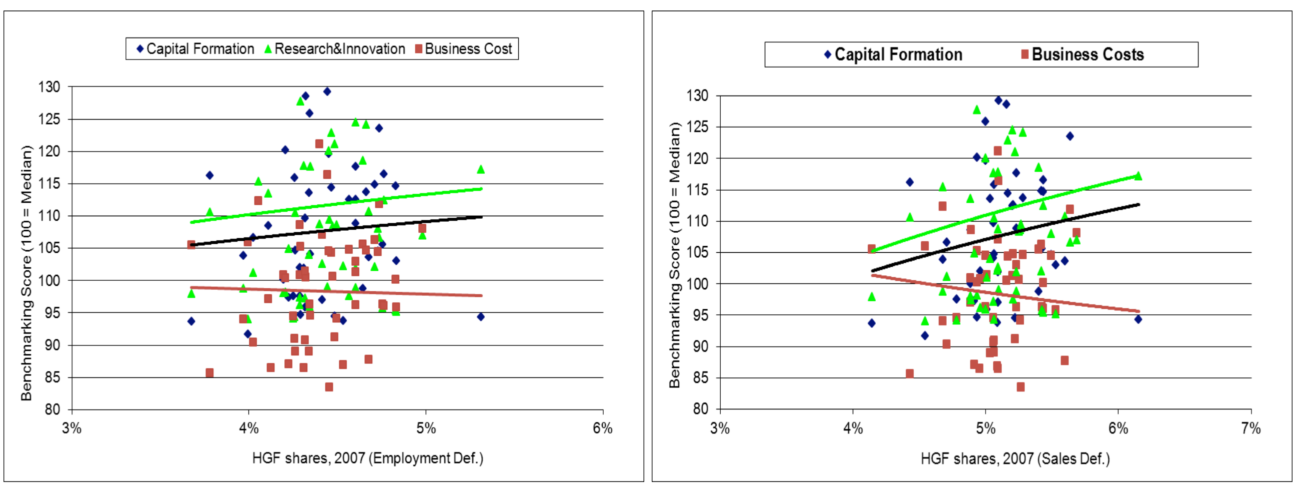

Figure 1 & 2. HGE Shares and Benchmarking Performance (Employment & Sales)

[13] This version of the paper applied stricter robustness tests compared to earlier drafts and results might therefore differ slightly.

As a rule, states scoring higher on capital formation and R&D, and lower on business costs, present higher HGE shares. This is congruent with mainstream U.S. economic development research suggesting that better capital and R&D conditions make a difference to firm growth. Further, the negatively sloping business costs line is congruent with the 'new economy' paradigm, which argues that business costs and financial location incentives matter far less for today's state growth strategy. Nevertheless, sophisticated causal analysis in this study finds that these factors are less important than is embraced by mainstream development practice.

Geographic clusters. If these broad policy parameters have only mild influence on HGEs, another avenue of investigation is geographic proximity. Our second research question asks if states with similar shares of HGEs occur in any specific geographic groupings. A simple single variable examination is displayed by the Figures below. The U.S. economy is a composite of meta-regions held together by common geographic attributes, history, culture and neighborly dialogue and competition. For businesses with exceptional job and sales growth performance between 2004 and 2007, Figures 3 and 4 show that the West, Mountain and Southeast Regions of the U.S. show to be generally more vibrant. These similarities may have much to do with related industrial structures, which are shown later to be a significant causal factor.

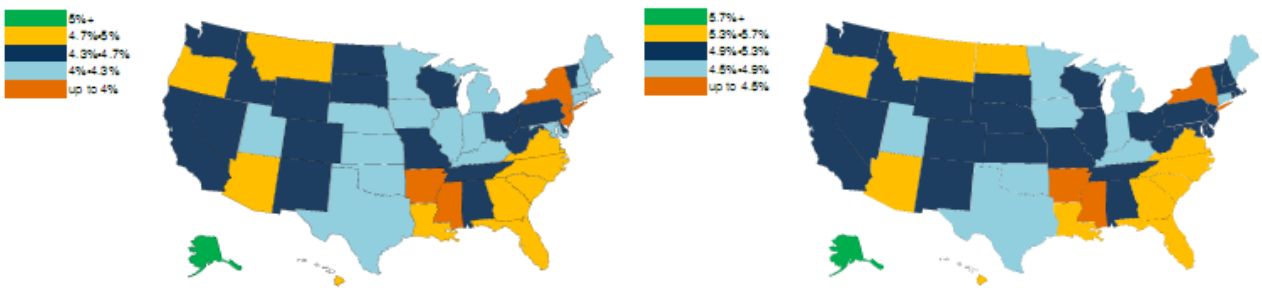

Figure 3 & 4. HGE Shares by Employment & Sales 2007

5.2 Principal component results

The results of the PCA consist of sets of factor loading weights *w*, see Eq. (4). Based upon visual inspection of the PCA scree plot which describes the amount of variance explained by each component *c*, and given the limitation that the number of components should be less that the cross section size of our dataset n = 50 states, L=30 components were selected. Together, these 30 components explain 86% of the total variation in the dataset $\tilde{x}$ for both the Sales and Employment defined dependant variables. Additionally, utilizing the results of regressions relating PCs to the two dependant variables in Table 1 we observe that less than all 30 PCs are statistically significant. For HGE(Empl) there were five significant components, which together represent 17% of the total variation in the predictor dataset $\tilde{x}$, for HGE(Sales) there were six significant components representing 22% of $\tilde{x}$ variation.

5.3 Regression analysis of principal components results

The third stage of our analysis involves regressing the normalized dependent variables of interest, HGE(Empl) and HGE(Sales), against the 30 principal components identified in Section 5.2. Table 1 gives the results for the baseline year 2007.

	HGE(Empl)		***HGE(Sales)***	
Component	Coefficient	t-stat	Coefficient	t-stat
1	0.00	0.66	0.01*	1.83
2	-0.03**	-3.70	-0.03**	-3.52
3	0.01	1.24	0.02**	2.39
4	-0.04**	-3.76	-0.03**	-3.25
12	-0.03**	-1.98	-0.02	-1.35
14	-0.06**	-4.00	-0.06**	-4.51
18	-0.04**	-2.43	-0.04**	-2.48
20	0.03*	1.87	0.03**	2.20
R^2	0.78		0.80	

Table 1 Regression results for statistically significant components
Note: **p<0.05, *p<0.10.

We first observe that only 8(7) of the 30 components considered are statistically significantly associated with either one or both of the dependant variables of interest at the 10%(5%) significance level. Importantly, 5 of the 8 components were significantly associated with both definitions of HGEs; these include components 2, 4, 14, 18 and 20. This indicates that these measures are important for explaining multiple HGE attributes. Alternatively, component 12 was only significantly associated with HGE(Empl) while components 3, and marginally 1, were only associated with HGE(Sales).

Additionally, R^2 model fits for both dependant variables are similar and high. Further, for both definitions of HGEs the signs and magnitudes component associations (regression parameters) are very similar. These similarities indicate that the same processes drive HGE shares for both definition, and the direction and magnitudes of the effect are quite similar. Because these components are based on normalized data, their relative magnitudes tell us directly about their associative importance. Interestingly, most components have a depressing effect on HGE share with the exception of the positive effects of C20 and to a lesser extent C3. Component 14 is observed to have the consistently largest negative effect, followed by the negative effects of components 4 and 18. Although these results tell us about the importance of components composed of multiple drivers, they do not directly give us an indication of the importance of individual drivers that are of interest to policy makers.

5.4 Calculated importance of metrics

To obtain information about the importance of individual underlying drivers, we combine the regression parameters from Table 1 with the factor loadings (unreported) using Eqs. (9) and (10). Tables 2a and 2b provide our main results regarding the importance of underlying and causally relevant drivers of HGE shares by employment or sales. Only drivers with at least one lag passing the Granger causality (GC) test and passing 7 of the 8 robustness checks are included. Column 3 gives the net importance score based upon Eq. (9). Summing across lags, column 4 gives the absolute importance score based upon Eq. (10).

Although the drivers are cardinally ranked based upon a combination of the number of robustness tests passed and by the absolute value of the net importance magnitude, this ranking information is less relevant than the fact that the drivers have passed an extensive battery of tests and have been selected out of over 150 individual metrics. Additionally, these importance rankings may be interpreted in at least two intuitive ways. Either, a high score indicates the direct influence of a driver upon the share of HGEs or alternatively, it may identify instruments that reflect an indirect influence upon the share of HGEs. For metrics with counterintuitive and strong results, we presume these are instruments with an indirect influence and we conduct the entire causality analysis and perform the eight robustness checks with the presumed high importance instrument as the dependent variable.

The mean importance scores and 2 times the standard deviations across identified drivers and for net and absolute importance measures displayed at the bottom of Table 2a and 2b. They can give us insights about the overall sign and magnitude of the influence of the drivers on HGE shares, about the variability across drivers of these contributions, and a sense about whether these contributions are significant, i.e. more than two standard deviations from zero. Intuitively, it should always be the case that for an individual driver, the absolute importance exceeds or equals in absolute magnitude the net importance since there would be no cancellation effect when summing importance scores across components; and this regularity is observed. We observe in Table 2a and 2b below that the overall mean net importance of drivers for HGE(Empl) is 0.0004 with a standard deviation of 0.012, and for absolute importance, the values are 0.003 and 0.009, respectively. For HGE(Sales) the mean net importance is 0.001 with a standard deviation of 0.012, and the values for absolute importance are 0.008 and 0.011, respectively. These positive averages tell us that the underlying drivers, once recovered from components with mostly negative regression parameters but including negative factor loadings, contribute on average positively to HGE shares for both definitions. However, the standard deviations indicate that the variability of this contribution across individual drivers is quite high. In other words, some drivers are consistently more important than others, in that their contributions are larger than two standard deviations from other drivers' values and they are more than two standard deviations from zero.

Individual drivers are grouped into two broad definitions, policy influenceable variables (on top), and control variables (below). For exposition, we include in the tables only metrics with particularly high importance scores and robust effects. For HGE(Empl), Table 2a below, five policy drivers stand out. First, the presumed indirect or instrumental effect of parkland is the largest and positive; additional results for this measure are provided below. Next, we find a presumed direct negative effect of establishment turnover, and the positive effects of broadband infrastructure, net migration rate, and skilled immigrants (see Appendix Table 1 for a description of all variables). Among the control variables positive effects include, lagged growth in the HGE share, job shares in other services, administrative services, and real estate, and the negative effects of the median HGE age and health sector job share.

Metric	*Lags Passed*	*Net Importance*	*Absolute Importance*	*Robustness Tests Passed*
Parkland	3	0.024	0.026	8
EstabTurnover	2	-0.006	0.010	8
BroadbInfra	3	0.003	0.005	8
NetMigrRate	3	0.012	0.014	7
SkillImmigr	3	0.010	0.010	7
Control Variables				
GroHiGroEstabShare	3	0.012	0.016	8
HiGroEstabAge	3	-0.019	0.019	7
OthServJobShare	3	0.015	0.018	7
AdminJobShare	3	0.014	0.014	7
HealthJobShare	3	-0.012	0.014	7
RealEstJobShare	3	0.010	0.010	7
Median		0.0004	0.007	
2*St.Dev.		0.012	0.009	

Table 2a Calculated importance of underlying metrics for HGE defined by employment[14]

For HGE(Sales), Table 2b below, all but two of the employment policy and control variables are present. Absent variables with respect to employment-defined growth include broadband infrastructure and net migration rates. Below we focus on notable differences. Overall, 10 policy variables stand out, of which three are common with HGE(Empl). New variables include the presumably instrumental positive effects of energy and urban housing costs, although these variables importance scores lie within two standard deviations of zero. Next, we find a the presumed direct negative effect of state and local tax burden and startup rates and the positive effects of air access to technology market hubs, state own source tax revenue and entrepreneurial generation. Among the 10 control variables identified, all six identified for HGE(Empl) are present and four additional drivers appear important. These include the negative effect of the manufacturing sector job share, and positive effects of arts job share, accommodation job share, and the lagged dependent variable.

[14] Previous versions of the paper multiplied all importance scores by a factor of 100 for ease of exposition. However, to maintain consistency with comparable studies, the unmodified score are reported in the tables of this paper.

Metric	*Lags Passed*	*Net Importance*	*Absolute Importance*	*Robustness Tests Passed*
StartupRate	3	-0.007	0.016	8
EstabTurnover	2	-0.007	0.011	8
SkillImmigr	3	0.005	0.014	8
EnergyCost	2	0.003	0.011	8
UrbHousCost	3	0.002	0.006	8
Parkland	3	0.027	0.030	7
TechMarkets	3	0.013	0.014	7
StateLocalTax	3	-0.010	0.014	7
OwnSourceTax	3	0.008	0.014	7
EntreprGen	3	0.007	0.009	7
Control Variables				
AdminJobShare	3	0.013	0.013	8
ArtsJobShare	3	0.011	0.011	8
GroHiGroEstabShare	3	0.010	0.016	8
AccomJobShare	3	0.019	0.022	7
OthServJobShare	3	0.015	0.022	7
HiGroEstabAge	3	-0.014	0.017	7
HiGroEstabShare	3	0.011	0.019	7
HealthJobShare	3	-0.011	0.017	7
RealEstJobShare	3	0.010	0.011	7
MfgJobShare	3	-0.007	0.009	7
Median		0.001	0.008	
2*St.Dev.		0.012	0.011	

Table 2b Calculated importance of underlying metrics for HGE defined by sales

Interestingly, there are a number of commonalities across these two definitions of HGEs, and all variables present for both measures have the same signs of effect and similar magnitudes. This suggests identification of robust and precise effects. In terms of policy variables, common positive effects include parkland and skilled immigrants, and establishment turnover rate as the only common negative effect. For control variables common positive effects include administrative job share, other services job share, and real estate job share, and the lagged level and growth of the dependent variable. Common negative effects include median HGE age and the health sector job share.

Preliminary investigations into the share of high growth establishments active during the recessionary period 2007-2009 show some commonalities. Variables for HGE(Empl) common with the 2006/7 analysis (and passing the robustness checks 1 to 3 as well as the basic regression), include the positive effects of the net migration rate and the control variable administrative job share, and the negative effect of the median establishment age.

Though not passing the basic regression test but with an importance score of more than two standard deviations from zero, are establishment turnover, although now with a positive effect compared to 2006/7. We also find among the control variables a positive effect of the growth in the lagged dependent variable and the negative effect of the health sector job share. In contrast, the robust effects in the 2006/7 periods of parkland, skilled immigrants and broadband infrastructure did not persist.

For HGE(Sales), strong time consistent effects included only the two control variables administrative job share and the lagged dependent variable, with the latter showing an importance score of almost four times the standard deviation. The share of establishments able to grow over such a short time period during a global economic crisis, especially in terms of sales, was therefore strongly path dependent. Commonalities with 2006/7 that passed all robustness checks but the basic regression and were more than two standard deviations above the median include the positive effects of air access to technology markets, establishment turnover, though again with a switch in sign from 2006/7, and the entrepreneurial generation variable. The strong effects of skilled immigrants, the two tax variables and the assumed instrumented variables of parkland, energy and urban housing costs did not persist.

Noticeably, two new strong effects, which did not pass causality and/or robustness checks in the pre-recession dataset, emerged for 2007/9 that were common across the employment and sales definition. A positive effect of patent productivity emerged, as well as the negative (likely across lags nonlinear) effect of the five-year establishment survival rate (e.g. establishments that survived in the first lag the period 2002-2006). In addition, a negative impact of the rural-urban employment and wage divergence emerged as significant.

Finally, two dozen driver interactions were tested for the baseline 2007 analysis. These were motivated by earlier literature and they reflected combinations of business cost, educational attainment and finance measures. Overall, only a few stood out as showing a larger impact in combination compared to their individual contribution. Some of the largest effects include a negative effect on HGE(Empl) of interaction 1 (associate degree attainment combined with business tax burden), which passes two robustness checks; and a positive effect of interaction 2 (business turnover – birth plus deaths relative to total establishments – combined with a 5-year survival rate), which passes only one robustness check. For HGE(Sales), a positive effect of interaction 2, which passes three robustness checks ; also, interaction 3 (BA attainment combined with technology-oriented venture capital), which passes one robustness check; and a negative effect of interaction 1 which passes one robustness check. Interactions 1 and 2 have similar signs and magnitudes and are common to both definitions of HGE, however, interaction 2 reflecting business churn and survival rate has an initially (lag 3 and 2) small negative effect but eventually (lag 1) a large and more than offsetting positive effect.

5.5 Basic regression analysis results

Key variables that may be playing a role as instruments and which require additional investigation include the positive effects of parkland or amenities, urban housing costs, and energy costs. Of course, a complete hierarchical analysis of all drivers would be preferred but is beyond the scope of this paper. However, the strongest counterintuitive results can be investigated more generally. We presume there may be a hierarchical relation among HGEs through these instruments down to the underlying drivers. As observed above, the strong effects of parkland and counter intuitive positive effects of urban housing and energy cost suggest these variables may be instrumenting for other more directly relevant variables. This may be possible due to the commonly known limitation of principal components analysis as reflected by its tendency to reduce the dataset to a smaller set of drivers, which may be highly correlated, with a larger set of variables. Additionally, a more general criticism of the Eq. (9) and (10) importance scores might be that they cannot be used to identify the role of individual drivers since once integrated into a component it is no longer possible to extract out the role of an individual variable. This is despite the fact that the mathematical relations described at Eqs. (8) - (10) reflects the exact algebraic relation among components, regression parameters, and drivers.

To investigate this possibility and the potential that instrumental variables are present we take the most conservative position and argue that at the very least these importance scores identify underlying metrics that one could analyze using more traditional/basic regression techniques. We estimate Eq. (11) with the metrics identified in Tables 2a and 2b. We add drivers identified via the same causal analysis applied for the HGE dependent variables as highly important for the three presumed instrumental variables (parkland, urban housing, and energy costs), and we add as many measures within 2 standard deviations of an importance score of zero that a positive degree of freedom allows.

These results can validate the use of our importance score as a means to identify important underlying drivers, indicates the relevance of our two standard deviation cutoff for identifying importance scores different from zero, and could be considered as a conservative substitute for the importance technique presented here. This test also performs the role of an eighth and final robustness check applied to our drivers for the main analysis. Limitations of this test include the fact that our cross section is limited to the 50 states and therefore given the numerous predictor variables satisfying the above criteria, only the years 2007 for the dependant variable and 2006 for predictor variables can be included. Therefore, the results of this test *can serve to confirm, but not to refute the results of earlier analysis* due to the limited dataset and potential for results highly specific to the 2006/07 period. Only variables identified to be significant are included in the Table 3 summary of these results below.

We gain a number of important insights from this analysis. First, comparing Table 3 to Tables 2a and 2b identifies a number of important commonalities. In terms of validating important drivers for HGE(Empl), variables that remain include the positive effects of broadband infrastructure and growth in HGE share and negative effects of establishment turnover and skilled immigration. The signs are consistent, except the influence of skilled immigrants, which switches to a negative effect. However, it can be shown that this effect is due to the small time sample (2006/07) employed in this test. Noticeably absent are the three instrumental variables that are discussed in Section 5.6 below.

HGE(Sales)	*Coeff. Est*	*T-stat*	*HGE(Empl)*	*Coeff. Est*	*T-stat*
EnergyCost	-0.27**	-2.06	BroadbInfra	0.51**	3.04
IPODeal	-0.19*	-1.92	EstabTurnover	-0.63*	-1.81
UrbHousCost	0.83**	2.03	SkillImmigr	-0.42**	-3.19
StartupRate	0.70**	2.25	GroHGEShare	0.64**	3.04
EstabTurnover	-0.69**	-2.32			
SkillImmigr	-0.50**	-3.29			
FDIValue	0.16**	1.97			
AdminJobShare	0.50**	2.86			
ArtsJobShare	0.42**	2.18			
MfgJobShare	-0.25*	-1.94			
GroHiGroEstabShare	0.92**	4.73			
R^2	0.96		R^2	0.93	
Durban Watson	2.08		Durban Watson	2.27	

Table 3 – Statistically Significant Basic Regression Results 2006/2007
Note: **p<0.05, *p<0.10.

For HGE(Sales) a number of variables remain and retain the same sign as for the importance score. These include the positive effects of administrative and arts job shares, growth in the dependant variable, and urban housing cost, and the negative effects of establishment turnover and manufacturing sector job share. Variables remaining but switching signs include again a counterintuitive negative effect of skilled immigrants, as well as a negative effect of energy costs and a positive effect of startup rates. New variables identified as significant for the regression but with importance scores within two standard deviations of zero for importance scores include a negative IPO deals effect and positive FDI value effect. Although two instruments are absent (parkland and energy cost), the positive urban housing cost effect remains.

In summary, these results lend support to the relevance of broadband infrastructure, establishment turnover rates, and growth in HGE shares for causally influencing the HGE share levels. They also support the importance, significance, and direction of effect of industrial sector control variables; in particular, the positive effects of administrative and arts sector job shares, and the negative effect of manufacturing sector job share. This information is reflected in Tables 2a and 2b for variables that have passed all eight robustness checks. However, the results also suggest that the influence of skilled immigrants and startup rates may be transitory or otherwise highly time specific given the switch in signs to negative and positive respectively. Fortunately, these changes can be related to distinctiveness in the 2006 predictor sample relative to 2007. Referring back to the PC regression results and inspecting the signs on lagged values for these variables (unreported) we observe that both variables have opposite signs on the 2006 lag compared to earlier lags and the cumulated net importance scores. Because the net importance measure sums effects across lags, the positive (negative) effects of lags prior to 2006 dominate the negative (positive) 2006 effects leading to positive (negative) net importance scores for startup rates (skilled immigrants). This contrasts with the basic regression results which only includes the 2006 predictor variables indicating contradicting negative (positive) effects for skilled immigrants (startup rates), respectively. Evidence of this effect can be inferred by comparing the net and absolute importance scores reported at Tables 2a and 2b. Net values (Eq. (9)) that are much smaller in absolute magnitude than the absolute importance score Eq. (10)) indicates variables where the direction of effect changes across lags; and we observe that skilled immigrants and start up rates are two variables for which this effect is most pronounced. The net importance scores are 0.005 and -0.007 but absolute scores are 0.014 and 0.016 respectively. This is evidence of cancellation of their effects across lags and therefore accounts for the sign switches when only the 2006 data is employed.[15] Overall, this indicates that the importance scores from the first seven robustness checks provide a more robust and broadly relevant account for the direction of effect of these variables.

The second insight from this analysis is that of the 15 marginally important variables, i.e. those within two standard deviations of zero for importance scores, 13 were not statistically significant within the regression. This suggests our two standard deviation cut off point is a reasonable approximation that can be used to identify important terms; until a more theoretical statistical measure of importance score standard error can be derived. Third, in all cases the R^2 and Durban Watson values indicate that the analysis explains a large amount of HGE share variance, and that the subset of terms included does not suffer from residual serial correlation.

5.6 Instrumental variable basic regression analysis

Regarding the instrumental variables, some insights about which drivers parkland, urban housing and energy costs may instrument for is provided by additional analysis. Each of these three variables was subjected to the same 8 robustness check causality analysis applied to the definitions of HGEs, making each instrument the dependant variable. Again the importance score calculations provided a set of drivers for each presumed instrument, and we focus on those passing all 8 robustness checks and which had importance scores beyond two standard deviations of zero. For parkland, important drivers included small business growth, growth in graduate degrees and growth in associate degrees. For urban housing cost these included energy cost (itself and instrument), homeownership rates, leisure employment, productivity diffusion, and sustainable income population. For energy costs, these included FDI and government interest payments. If our hypothesis about these three instruments is true, i.e. they are indeed instruments, these variables should be replaced or have their counter-intuitive influence attenuated by the underlying drivers they represent.

As a final check, we conducted a supplementary basic regression analysis relating both definitions of HGEs to important drivers previously identified, including the three presumed instruments, and including the aforementioned drivers of these instruments. This produced a table (unreported) similar to Table 3 but included new variables and the important drivers of the instruments.

---

[15] In addition, the variance of the data on the dependent variable is notably higher for 2006, and summary statistics of the median age across the 50 states indicates a sharp drop from a median of 8 years to a median of 5 years from 2006 to 2007, reflecting some instability in the characteristics of the firms represented in each year.

For HGE(Empl) the results are dramatic, all of the drivers of the three instruments are significant and the significance of the instruments themselves remains, although the magnitude of parkland is much reduced, and the sings of urban housing and energy costs become negative. Overall, this indicates that parkland as a representation of amenity is still importance once its correlated / underlying drivers are extracted out. In addition, the anomalous signs of the costs variables are resolved. The Durban Watson statistic for this regression is 2.3. For HGE(Sales) the results are less clear. All predictors are at best marginally significant, with parkland having the largest marginally significant positive effect. The estimates for the instruments themselves for the most part retain larger magnitudes of effect than their underlying drivers. The energy cost variable has a negative but insignificant sign, but urban housing costs retains its positive albeit insignificant sign. The Durban Watson statistic for this regression is 1.91.

Overall, there is some evidence that the large magnitude of parkland and anomalous positive signs for energy and urban housing costs are due to the fact that they are instruments for other variables. When we include the important drivers underlying the three presumably instrumental variables (parkland, urban housing cost, and energy cost) we observe a reduction of the predicted magnitude of effect (parkland) or inclusion with the intuitive negative signs (energy and urban housing costs).

6. Discussion and policy implications

We set out to address five principal research questions with our multifaceted statistical analysis of drivers of HGE shares that achieve an average growth of at least 20% per year in terms of employment or sales. Our focus was on identifying drivers of the share of all establishments, whether standalone or branches, to identify what creates or attracts the most vibrant businesses rather than a more limited focus on aggregate firm growth.

First, in terms of general results regarding Research Question 1, state policy commonalities, several of today's most widely held policy paradigms focused at fostering high growth businesses such as capital access and business cost seem to have strong associations with the share of high-growth businesses. However, in our analysis few emerge as primary causal candidates. The investigation for Research Question 2 into geographic groupings did show some regional commonalities for the share of HGEs in 2007 for both sales and employment definition that deserve further, more spatially oriented analysis. In terms of Research Question 5, most interaction terms pass fewer of the robustness checks and/or have smaller magnitude importance measures and are therefore not reported in Tables 2a and 2b. Future research will investigate a substantially larger set of interactions.

More specifically, in terms of Research Question 3, the most direct and broadly important influences on HGE shares include labor market factors reflecting net domestic in-migration as well as international skilled immigrants. Less influencable but highly important and robust influences include the share of employment across particular sectors. In particular, the job share in the administrative, other services, and real estate sectors and to a lesser extent leisure and accommodation sectors have positive effects and the health and manufacturing sector job shares are observed to have negative effects. This yields insights into which sector endowments could enhance or degrade the potential for a locality's HGEs. In particular, it is apparent that value added services including traded sector services dominate HGEs and an industry structure characterized by value added services, with manufacturing playing a lesser role, might positively contribute to the presence of high growth businesses.[16] Additionally, the negative effect of the median age of high growth establishments stands out in line with previous literature, challenging economic developers to consider both size and age when pursuing high growth business strategies. Other important drivers of HGEs include the positive effects of broadband infrastructure and a moderated degree of establishment turnover or churn and startup rates for HGE(Empl), the latter possible reflecting a nonlinear impact on the dependent variable, as well as the benefits of technology market air access. In addition, for HGE(Sales) we found evidence of the benefits of a low state and local tax burden and a high share of locally generated taxes. Some of the top variables from the analysis may be playing the role of instruments rather than direct influences. Additional causal analysis of the presumed instrumental variables identified underlying drivers of parkland, energy cost and urban housing cost, including positive effects from educational attainment, productivity diversity across sectors, venture capital funding, and small business growth features, and negative effects of high degrees of government employment. The more intuitive direction of effect from these drivers to HGE through the instruments may account for sometimes anomalous positive influences from the cost based instruments.

In terms of Research Question 4, any metric passing all robustness tests displays high stability across time, such as establishment turnover for both definitions of the dependent variable, broadband infrastructure and parkland for HGE(Empl) as well as skilled immigrants and startups for HGE(Sales). Time consistent effects for control variables include the growth in the lagged dependent variable for both definitions as well as the share of arts and accommodation job shares for HGE(Sales). For the stricter comparison of the pre-recession 2007 to the shorter recessionary 2009 period, there is still surprising consistency among a small set of drivers common to 2007 and 2009 and both HGE definitions, and a large degree of consistency across HGE definitions within 2009.

Our results with respect to research questions 3-5 have some key policy implications. Few entrepreneurial attributes of state economies matter to home-grown fast growth businesses, especially during non-recessionary periods. The exception seems to be 'entrepreneurial generation', a labor supply measure of the share of the population 25-44 year old that is considered the prime age group for entrepreneurs; and the ratio of skilled immigrants (with at least a Bachelor degree) to the overall population.

Testing for the direction of causality is a key contribution of this analysis. For example, our 2007 results for sales and employment without tests for the direction of causality and additional robustness might have

[16] An investigation of the share of high growth businesses across major industry sectors (2-digit NAICS codes), shows that close to 40% of HGE by either sales or employment definition in 2007 and close to 50% in 2009 were present in just three service sectors: professional, scientific, and technical services, administrative and waste/remediation management, and other services that include activities, such as equipment and machinery repairing, grant making, advocacy, personal care services, photofinishing services, temporary parking services.

erroneously highlighted for example the role of business incubators. However, subsequent statistical tests ruled these variables out as causal drivers of the share of high-growth businesses.

A few key causal drivers identified in our analysis can be shaped by public policy: Migration: Migration, especially international, has grown out of favor as a growth determinant in maturing U.S. states. This analysis finds otherwise. Apparently, in-migration, both domestic and international (skilled migrants), sets up both dynamic local consumer and labor markets in addition to a climate conducive for innovation. Notably, the ALEC-Laffer State Economic Competitiveness Index is built upon the assumption of dual drivers: competitive taxation and net migration. Business Churn and Startup Rates. The negative sign on establishment turnover consistently appearing for both HGE definitions and in both years most likely reflects a non-linear relationship. Some churn as well as a healthy startup rate in the economy is necessary in the spirit of creative destruction but a too high rate will result in the inefficient allocation of resources. Broadband: Broadband policies have played a relatively minor role in U.S. state growth strategy although a few states took it seriously nearly a decade ago, such as North Carolina. Over the past two decades several developed countries have deliberately ramped up broadband efforts – South Korea's broadband policy initiatives in the mid 90's; the e-Japan Strategy, 2001; and Digital Britain, 2009. In 2010, The Federal Communications Commission released The National Broadband Plan for the U.S. These findings serve to validate the importance of such efforts. Taxes: Our analysis has highlighted the importance for sales oriented high growth firms of low state and local tax burden and a high fiscal independence. If state recovery from the 'Great Recession" is any indication, it appears that fiscally health states such as North Dakota have been advantaged over the fiscally distressed like Illinois.

Parkland/Amenity The economics of amenity remains a less developed field partly because of definitional, measurement and modeling issues. Nevertheless, development professionals in the U.S. are being confronted more frequently with apparent linkages between community development and economic development policies. Although, this study finds a large instrumental effect of this variable, once accounting for its drivers, parkland remains a significant driver, most likely representing the combined effect of amenities. In recent years, a few U.S. states, e.g. Michigan, have been taking the challenge of linking entrepreneurial development and quality of life/sense of place seriously. Consequently, this finding is no surprise. Interestingly, key drivers of this instrument include educational attainment and small business growth. This suggests a policy focus on these elements can potentially have a relative large impact on the share of high growth businesses. Energy and Urban Housing Cost. The U.S. states have been quite active in both areas of public policy for at least two decades. However, direct linkages between such policy initiatives and economic growth have remained elusive. This analysis seems to support that these variables at best act as instruments and suggests a closer examination might be worthwhile.

It is no surprise that many of the structural or control variables appeared throughout definitions and years as an important impact in the HGE share across states. These variables rather reflect the historic structure of a state and could be considered long-term intervention variables. The three main industry findings of this analysis are that the employment share in most service sectors except healthcare has a positive causation while the manufacturing share has a negative impact. The implications will be of particular interest to ‘cluster strategy’ proponents. Nearly all U.S. states have an industry cluster strategy/plan. While their prevailing approach is to identify industry clusters as a modern-day variant of ‘targeted industries’, this research points to a focus less on particular industries and more to striking a balanced industry mix with a heavy dose of advanced/value added services in general and traded services in particular. Another structural issue is the focus on a particular size or age of businesses. Empirical studies suggest that smaller and younger firms grow faster (though with more volatility). The median age of high-growth establishments determined at the beginning of the growth period in our analysis came through consistently as a negative impact on the share of high growth firms (though not indicating its impact on subsequent firm growth) consistent with the literature on firm growth.

7. Conclusion

Our results indicate that governments do have a constructive role to play in fostering HGEs. More definitive findings for 2007-2010 will be possible once 2010 data is available in late 2012. What is already apparent is that causal factors vary over the business cycle especially one as tumultuous as the 2004-2009. That being the case, the method developed in this paper lends itself to routine reruns as new annual data becomes available, providing a time series tracking of causality.

Given the complex and dynamic nature of high growth firms, there is a consensus that effective, enabling policies need to be flexible in their approach, adaptable over time, broad-based across policy areas, and tuned to the different stages of firm growth. They furthermore have to be implemented in coordination with existing policies, and especially SME support policies have to be carefully examined to prevent conflicting incentives (Buss, 2001; Autio and Hoelzl, 2008). Current policy initiatives relating to high performance firms still focus substantially on SMEs (OECD, 2010; Autio et al., 2007) and there is still no consistent evidence of best practice, indicating the need for further policy analysis and empirical guidance. Strategic interventions must move away from a bias of helping start-ups to accelerating growth in young firms (Hoffman, 2011). Selection of strategic focus will vary from jurisdiction to jurisdiction and may take policy-makers outside conventional wisdom as this paper does and as do Estrin et al. (2011) who find corruption and weak property rights matter.

This paper finds that a selection of tested legal climate and regulatory climate measures are not causal, corroborating recent survey findings in Britain (NESTA, 2011). Rather, these firms grow most in economies that are dynamic/‘on the move’ evidenced in the paper by the impact of positive net migration. We furthermore find that causalities may shift over the business cycle. The policy implication, also noted by McFarland et al. (2010), is that what works well in one stage of a business cycle may not be as effective in another. This argues for proactive policy practice involving careful and routine analytical tracking with dashboards to alert the need to make policy adjustments, something governments are seldom set up to do.

This research adds to a small body of recent empirical policy research pointing policy makers in the following general direction: **1.** Growing jobs from entrepreneurial businesses calls for much more than today’s overdone focus on startups. In particular, greater attention needs to be paid to small young existing HGEs. **2.** By their nature HGEs are unpredictable with regards to growth spurts, particular sectors, technologies or market reach (Mason and Brown, 2010). However, this research finds that overall industry structure matters greatly. Higher HGE growth is found in advanced services state economies in the U.S -- less so in state economies heavily weighted toward production manufacturing. **3.** Even types of HGF firms are difficult to predict – their size, age, corporate structure, production organization. However, age does matter. Results from this research, for example, presents a generally lower share of HGEs in states with higher HGF median age. **4.** Even conventional wisdom about what matters to HGEs may not be at play. Specifically, this study points policy-makers to devise an agile system of measuring, tracking and modeling HGE progress. This must be a continuous process with annual or at least biennial updates with intent to capture shifts in causality due to changing markets, technologies and business cycle conditions. A one-size, one-time policy set will not work. As great a challenge lies in designing an appropriate structure of policy oversight to manage such strategic interventions as in determining what those interventions should be at any point in time.

Limitations and next steps. A caveat to our results is that all results have to be interpreted relative to the denominator used, e.g. per capita, per GDP, etc. Crain and Lee (1999) point out that statistical results can be sensitive to how variables are transformed and future analysis should investigate alternative scaling. Some of the anomalous results that persist deserve further investigation regarding issues with data definitions and data quality. Further, the primary use of a dataset using an absolute measure of HGEs, i.e. they must achieve at least 20% growth in employment or sales per year, results in a highly volatile dataset especially during a recessionary period. A more relative measure, the top 5% of firms, would produce less variation and could affect the magnitude of our results, i.e. the importance score or predicted effects might be smaller. Using a combination of relative and absolute change in the growth definition will also be part of the next phase of this research and adding further outcome variables such as net worth, might provide a more comprehensive picture of what drives the presence of high growth businesses.

The preliminary investigation of geographic elements suggests an explicit spatial modeling approach is warranted. In addition, most empirical studies on firm growth have focused on cross-sectional analysis. The current study also introduced a time element by checking the consistencies of results through time but an important step would be to track the path of impact of key drivers over a longer time series to identify, for example, an optimal sequence of policies. In addition, focusing at the establishment level allowed us to bypass the more complex issue of ownership structure, legal form and finance structure. However, these factors might have important impacts on business growth, and deserve further investigation. Similarly, firms tend to grow along a diverse range of paths and some type of growth paths, with less volatility, might be preferred as policy targets. Local independent businesses tend to form where the entrepreneur happened to be located, while startups and branches might be more influenced by state business climate and market opportunities. Separating out firm branches and standalone businesses in further research might highlight important policy differences.

Appendix

VARIABLE NAME	DESCRIPTION
1-Year Establishment Survival Rate	1-Year Establishment Survival Rate
Expansion VentureCapital	Expansion and Later Stage VC per $100,000 private GDP
5 Year Establishment Survival Rate	5 Year Establishment Survival Rate
Accommodation Job Share	Employment Share of Accommodation and Food Sector
Administrative Job Share	Employment share of administrative, support, waste management and remediation services sector
Advanced Producer Services	Share of advanced producer services industries that are highly concentrated
Airport Performance	Percent of departing flights delayed
Arts Job Share	Employment share of Arts, Entertainment, and Recreation sector
Associate Degree Attainment	Percent of 16-and-older labor force with an Associate degree or equivalent
ATP Financing	ATP financing per $1 million GDP
Bachelor's Degree Attainment	Percent of 16-and-older labor force holding a bachelor's degree or higher
Bank Commercial and Industrial Lending	Total bank lending to commercial and industrial customers per $1,000 private gross domestic product
Bridge Quality	Percent of bridges characterized as "obsolete" or "deficient
Broadband Connection	Number of broadband Internet lines per 1,000 residents
Broadband Coverage	Premises with High Speed Internet Access Score
Business Incubators	U.S Technology Administration, State Science and Technology Indicators; Business Incubators per 100,000 firms
Business Liability Costs	Average business-liability direct premiums paid per $100,000 of private gross domestic product
Business Taxes	State and local business taxes per dollar of private economic activity
Capital Investment in Manufacturing per Job	Capital expenditures per production employee
Construction Job Share	Employment Share of Construction sector
Cultural Institutions	Number of establishments NAICS 711 and 712 per 100000 residents
Education Job Share	Employment share of Educational Services sector
Energy Costs	Average industrial and commercial energy price per kilowatt-hour
Establishment Formation Rate	Establishment births relative to all establishments
Establishment Termination Rate	Establishment deaths relative to all establishments
Establishment Turnover	Establishment births plus deaths realtive to all establishments
Export Intensity	Merchandise Export dollars per private GSP dollar
Export-related Jobs	Percent of private industry jobs that are merchandise export-related
Federal OJT	Percent of labor force that participated in federal on-the-job training
Federal R&D	Federal funded research and development expenditure estimates per $100,000 GDP
Finance Job Share	Employment share of Finance and Insurance sector
FDI	Employment by foreign majority-owned U.S. affiliates as percent of all employment
FDI Value	Gross book value of non-bank property, plants, equipment as percent of private GSP
Fortune 500	Total number of Fortune 500 headquarters
Gender Equity	Percent of the female workforce in "top jobs" (managerial, business, financial, professional and related occupations)
Generational Creative Class	Percent of labor force age 16-34 and 55+ years old with a Bachelor's degree or higher
Government Assistance	Per capita assistance and subsidies expenditures
Government Gross State Product	State and local government GDP as a percentage of total GDP
Government Interest Payments	State and local government interest payments on general debt as share of total expenditures
Government Job Share	Percent of total jobs held by state and local government
Green Industries	Percent of private establishments in green industries
Gross Domestic Product Per Capita	GDP per capita
Gross Domestic Product per Job	Gross domestic product per Job
Health Job Share	Employment share of Health Care and Social Assistance sector
Health Mandates	Number of mandated health insurance benefits in each state
Health Spending	Per capita health expenditures by state of residence
HGE Age	Median Age of HGE
HGE Size	Median Employment Size of HE
High Income Tax Rate	Maximum tax rate for an additional $1000 of income on an initial $1,500,000 of wage income
High School Attainment Only	Percent of 16-and-older labor force holding a high-school diploma
Hightech Manufacturing Employment	Percent of total covered manufacturing employment in high-tech manufacturing industries

Hightech Services Employment	Percent of total covered service-providing employment in high-tech service industries
Highway Quality	Miles graded "rough" or worse per 1,000 miles of highway
Income Strength of Traded Sector	Traded sector earnings (in dollars) per worker
Industrial Diversity	Hachman Index of industrial concentration
Industry Job Diffusion	Percent of industries with job gains
Industry R&D Performance	Industry research and development expenditures per $100,000 GDP
Industry-Funded University R&D	Industry-financed R&D expenditures at universities and colleges per private $100,000 GDP
Industry-University R&D Performance	Ratio of Industry R&D to University R&D expenditures
Information Job Share	Employment share of Information sector
Inventor Patents	Inventor Patents per 1,000 Innovation Worker
IPO Awards	IPO deals received over most recent 3 years per 1,000 firms
IPO Financing	Amount of initial public offerings over most recent 3 years per $100,000 private gross domestic product
Laborforce Participation	Labor force participation rate
Limited Activity Days	Average number of days in the past 30 days that a person could not perform work or household tasks due to physical or mental illness.
Local Phone Competition	Percent of phone lines controlled by competitive local exchange carriers
Malpractice Costs	Index of medical malpractice insurance rates across three disciplines (internal medicine, general surgery, OB/Gyn)
Management Job Share	Employment share of Management of Companies and Enterprises sector
Manufacturing GSP per Job	Manufacturing GDP per Manufacturing total employment
Manufacturing Job Share	Employment share of manufacturing sector
Manufacturing Value Added per hour	Manufacturing Value Added per hour
Manufacturing Value Added per Job	Value added per 1,000 production worker
Metro Office Rents	State population-weighted Metro Office Rents Index average
Mining Employment	Employment share of mining sector
Net Domestic Migration Rate	Net domestic migration to the state relative to population in start year
Net Establishment Entrants	Establishment births less deaths relative to all establishments
Next Generation Internet	Number of Abilene network primary participants and connectors per 100,000 establishments
Nonfarm Self-Employment	Number of non-farm proprietors per 1,000 jobs
Nonfederal Government R&D to Universities	Nonfederal Government R&D to Universities and Colleges per $100,000 GDP
Non-Labor Income per Capita	Dividends, rents, and interest income per resident
NSF Funding Rate	Share of National Science Foundation proposals funded
Occupational Fatalaties	Occupational fatalities per 100,000 workers;
Other Innovation Degrees	Percent of degrees earned in quasi-science and quasi-technical fields
Other Innovation Workers	Workforce Share of quasi-science and quasi-technical occupations
Other Services Job Share	Employment share of Other Services sector
Own Source Tax Revenue	Share of total state and local tax revenue from own sources
Parkland	Acres of state and national parkland per 10 square miles of land
Patents	Number of patents per 10,000 innovation workers
Patents Productivity	Number of patents per $100 million research and development investment
Per Capita Disposable Income	Per capita disposable income
Physical Science and Engineering Degrees	Percent of bachelor's and graduate degrees earned in physical sciences and engineering
Physical Science and Engineering Workers	Workforce Share of physical sciences and engineering occupations
Population Density	Population Density
Prime Working Age Population	Percent of Population 25-44 years old
Private Lending to Small Businesses	Private loans to small businesses per 1,000 small firms
Productivity Diffusion	Share of industries with above average productivity
Professional Job Share	Employment share of Professional, Scientific, and Technical Services sector
Racial Equity	Percent of the non-white workforce in "top jobs" (managerial, business, financial, professional and related occupations)
Real Estate Job Share	Employment share of Real Estate and Rental and Leasing sector
Renewable Energy Use	Percent of Net Generation from Renewable Energy Sources
Research Institutions Inventions	AUTM, Invention disclosures per $100,000,000 AUTM R&D dollar
Research Institutions Licenses to Small Businesses	Licenses and options to small businesses per 100,000 small firms
Research Institutions New Patent Applications Filed	AUTM, patents per $100,000,000 AUTM R&D dollar
Research Institutions Royalty/License	Gross royalty and license income per $1 million gross domestic product

Income	
Research Institutions Spinoffs	Spinout businesses per $1 billion R&D funding
Retail Trade Job Share	Employment share of Retail Trade sector
Rural Internet Access	Percentage of farms with internet access
Rural-Urban Disparity	Index of rural-urban economic performance differences
SBIC Awards	SBIC awards per 10,000 small businesses
SBIC Financing	SBIC financing per $100,000 of small business payroll
SBIR Awards	SBIR Awards per 10,000 small businesses
SBIR Financing	SBIR Financing per $100,000 small business payroll
Seed Venture Capital	Seed/Early Stage VC per $100,000 private GDP
Service Sector Productivity	Services GDP per Services Jobs
Skilled Immigrants	Percent of population that are naturalized citizens with a bachelor's degree or higher
Small Business Health Care Premiums	Average of mean single and family premiums for small firms per enrolled worker
SME Export Share	Percent of small firms that export
Some College Attainment	Percent of 16-and-older labor force with some college attainment
State and Local Tax Burden	State and local taxes as a percent of private income
State Business Tax Structure	Tax Foundation Corporate Tax Index
State R&D Tax Credit	State Effective R&D Tax Credit Rate
STTR Awards	STTR Awards per 10,000 small businesses
STTR Financing	STTR Financing per $100,000 small business payroll
Tech Venture Capital	Tech-oriented venture capital per $100,000 of private gross domestic product
Technologist and Technician Workers	Workforce Share of technologists and technicians
Technology and Technician Degrees	Percent of associate's degrees and post-secondary vocational awards earned in technology and technician fields
Technology Market Access	Enplanements to largest U.S. commercial and emerging technology markets per 1,000 residents
Time to Work	Average Travel Time to Work of Workers 16 Years and Over Who Did Not Work at Home
Top Ranked Graduate Programs	Total top graduate programs per 100 educational institutions
Transportation Job Share	Employment share of Transportation and Warehousing sector
Unemployment Insurance Costs	Average employer unemployment insurance rate paid on all taxable wages
Unemployment Insurance Structure	Tax Foundation Unemployment Insurance Tax Index
Unemployment Rate	Unemployment Rate
Union Membership	Percent of workers represented by unions
Unit Labor Costs	Unit labor cost index, U.S.=100
University R&D Performance	Research and development expenditures by universities per $100,000 gross domestic product
Urban Cost of Living	Cost of Living Index in largest metro area
Urban Housing Cost	Hourly wage needed to afford two-bedroom housing at fair-market rent
Utilities Job Share	Employment share of Utilities sector
Water Systems	Percent of population served by water systems with reported health violations for treatment technique and maximum contaminant level
Wholesale Trade Job Share	Employment share of Wholesale Trade sector
Workers' Compensation Benefits	Workers' Compensation Benefits Per $100 of Covered Wages
Workers Compensation Premiums	Average workers' compensation rate paid per $100 of payroll

Appendix Table 1. List of Explanatory Variables[17]

[17] More detail about sources for each variable can be obtained by request from the corresponding author.

References

Achtenhagen, L., Naldi, L., & Melin, L. (2010). Business growth. Do practitioners and scholars really talk about the same thing? Entrepreneurship: Theory &Practice, 34(2), 289–316.

Acs, Z. J. and Mueller, P. (2008): "Employment effects of business dynamics: Mice, Gazelles and Elephants", Small Business Economics, 30: 85-100.ArmAcs 2002

Audretsch, D.B. and M. Fritsch (2002), 'Growth Regimes over Time and Space', Regional Studies, 36: 113-124.

Audretsch, D.B. and M.C. Keilbach (2004), Entrepreneurship Capital and Economic Performance, Regional Studies, 38 (8), 949-959.

Autio, E. and W. Hölzl, (2008). Addressing Challenges for High-Growth Companies: Summary and Conclusions of the Europa Innova Gazelle.

Autio, E., M. Kronlund and A. Kovalainen (2007). High-growth SME support initiatives in nine countries: Analysis, categorization, and recommendations. MTI Publications 1/2007, Helsinki: Ministry of Trade and Industry

Barringer, Bruce R., Jones, Foard F. and Neubaum, Donald O. (2005), "A Quantitative Content Analysis of the Characteristics of Rapid-growth Firms and their Founders." Journal of Business Venturing 20(5), 663–687.

Birch, D.L., 1979. The Job Generation Process. MIT program on neighborhood and regional change, Massachusetts Institute of Technology, Cambridge, MA: Birch 1999

Birch, D.L., Medoff, J., 1994. Gazelles. In: Solmon L.C., Levenson A. R. (Eds.), Labor Markets, Employment Policy and Job Creation. Westview, Boulder, CO, pp. 159–167.

Bruce, D. and Deskins, J. (2010). Can state tax policies be used to promote entrepreneurial activity? Small Business Economics. 1-23.

Buss, Terry F., "The Effect of State Tax Incentives on Economic Growth and Firm Location Decisions: An Overview of the Literature," Economic Development Quarterly 15:1 (2001), 90-105.

Coad, A. & Rao, R. (2008). Innovation and Firm Growth in 'Complex Technology' Sectors: A Quantile Regression Approach. Research Policy, 37, 633-648.

Coad, A. (2009). The Growth of Firms: A Survey of Theories and Empirical Evidence. Cheltenham: Edward Elgar.

Coad, A., & Hoelzl, W. (2009). On the autocorrelation of growth rates. Journal of Industry, Competition and Trade, 9(2),139–166.

Cullen, Julie Berry, Roger H. Gordon. Taxes and Entrepreneurial Activity: Theory and Evidence for the U.S.," NBER Working Paper No. 9015, 2002.

Daunfeldt, S.-O., N. Elert and D. Johansson (2010), 'The economic contribution of high-growth firms: do definitions matter?', Ratio Working Papers no. 151, May.

Davidsson, P. and Gartner, W. B. (2003): "Arriving at the high-growth firm", Journal of Business Venturing, 18: 189-216.

Davidsson, P., & Henrekson, M. (2002). Determinants of the prevalence of start-ups and high-growth firms. Small Business Economics, 19(2), 81–104.

Davidsson, P., 2006. Nascent Entrepreneurship: Empirical Studies and Developments. Foundations and Trends in Entrepreneurship 2.

Davidsson, P., Delmar, F., & Wiklund, J. (2006). Entrepreneurship and the growth of firms. Cheltenham: Edward Elgar.

Davis, S.J., Haltiwanger, J., & Schuh, S. (1996). Small business and job creation: Dissecting the myth and reassessing the facts. Small Business Economics, 8(4), 297-315.

Delmar, F. 1997. Measuring growth: methodological considerations and empirical results. In R. Donckels & A. Miettinen (Eds.), Entrepreneurship and SME Research: On its Way to the Next Millennium (pp. 199-216). Aldershot, England: Ashgate.

Delmar, Frédéric and Per Davidsson (1998), "A Taxonomy of High-growth Firms." In Paul D. Reynolds et al., eds., Frontiers of Entrepreneurship Research, Wellesley, MA: Babson College, 399–413.Delmar, F. F.,

Djankov, Simeon, Rafael La Porta, Florencio Lopez deSilanes, Andrei Shleifer, "Regulation of Entry", NBER Working Paper No. W7892.

Estrin S., Korostelva, J., Mickiewicz, T., 2011. Which Institutions Encourage Entrepreneurs to Create Larger Firms? Discussion Paper 8247, CEPR.

Friedman, J. 1995. The effects of industrial structure and resources upon the distribution of fast-growing small firms among US urbanised areas. Urban Studies v32, n6.

Fritsch, M., & Mueller, P. (2007). The effect of new business formation on regional development over time: The case of Germany. Small Business Economics.

Goedhuys, M., and L. Sleuwaegen (2010). "High-Growth Entrepreneurial Firms in Africa: A Quintile Regression Approach," Small Business Economics. 34, 1: 31 -51.

Goetz, S.J. and A. Rupasingha 2009, "Determinants and Implications of Growth in Non-Farm Proprietorship Densities: 1990-2000," Small Business Economics. 32, 4: 425-38.

Goetz, S.J., and D. Freshwater, "State-Level Measures of Entrepreneurship and a Preliminary Measure of Entrepreneurial Climate," Economic Development Quarterly 15, 1 (February 2001):58-70.Gordon 1998

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica, **37**, 424–438.

Green, W. 2000. Econometric Analysis. New York, NY: McMillan

Grilo, I. and A.R. Thurik (2005), 'Latent and actual entrepreneurship in Europe and the US: some recent developments', International Entrepreneurship and Management Journal, **1** (4), 441–59.

Hall, J. and R. Sobel (2008) 'Freedom, Entrepreneurship, and Economic Growth', in Lessons from the Poor: Triumph of the Entrepreneurial Spirit, Oakland, CA: Independent Institute.

Haltiwanger, J., R. Jarmin and J. Miranda (2010), 'Who creates jobs? Small vs. large vs. young', Discussion Paper 101910, Center for Economic Studies (CES), US Census Bureau,

Henrekson, M. and Johansson, D. (2009): "Competencies and institutions fostering high-growth firms", Foundations and Trends in Entrepreneurship, 5: 1-80.

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

Henrekson, Magnus (2005), "Entrepreneurship: A Weak Link in the Welfare State." Industrial and Corporate Change 14(3), 447–467.

Hoffman, A.N. (2011). Promoting Entrepeneurship: What are the real policy challenges for the EU? In E. Phelps, H. Sinn (eds), Perspectives and performance of the continental economies. MIT Press.

Hoffman, A.N., & Junge, M. (2006). Documenting data on high-growth firms and entrepreneurs across 17 Countries. Fora, Copenhagen: Mimeo.

Hölzl, W. (2009): "Is the R&D behaviour of fast-growing SMEs different? Evidence from CIS III data for 16 countries", Small Business Economics, 33: 59-75.

Klapper, L., L. Laeven, and R. Rajan. 2006. Entry regulation as a barrier to entrepreneurship. Journal of Financial Economics 82 (3): 591– 629.

Kolko, Jed, and David Neumark, Business Location Decisions and Employment Dynamics in California (San Francisco: Public Policy Institute of California, 2007).

Kreft, S.F., Sobel, R.S., 2005. Public policy, entrepreneurship, and economic growth. Cato Journal 25 (3), 595–616.

Laffer, Arthur B. and Stephen Moore. 2008. "Rich States Poor States: ALEC-Laffer State Economic Competitiveness Index." American Legislative Exchange Council. Washington. <http://www.alec.org/fileadmin/newPDF/ALEC_Competitiveness_Index.pdf>.

Lee, N. (2011). Free to grow? Assessing the obstacles faced by actual and potential high growth firms. NESTA. Working Paper, November 2011.

Levie, J. (1997). Patterns of growth and performance: An empirical study of young, growing ventures in France, Ireland and Scotland. In P.D. Reynolds, W. Bygrave, N.M. Carter, P. Davidsson, W.B. Gartner, C. Mason, et al. (Eds.), Frontiers of entrepreneurship 1997 (pp. 419–443). Wellesley, MA: Babson College.

Mason, C. and Brown, R. (2010) , High Growth Firms in Scotland. Final Report for Scottish Enterprise, Glasgow.

McFarland, C., McConnell, K., Geary, C. 2010. SMALL BUSINESS GROWTH DURING A RECESSION: LOCAL POLICY IMPLICATIONS. Conference Paper, Federal Reserve Bank Atlanta, October 26-27, 2010.

Neumark, D., B. Wall, and J. Zhang (2011): .Do Small Businesses Create More Jobs? New Evidence for the United States from the National Establishment Time Series. The Review of Economics and Statistics, 93(1), 16.29.

Nooteboom, B., & Stam, E. (2008). Micro-Foundations for Innovation Policy, Amsterdam,

Nordic Entrepreneurship Monitor (2010). Nordic Council of Ministers, Copenhagen 2010. http://www.norden.org/en/publications/publikationer/2010-748

OECD, 2010. SMEs, Entrepreneurship, and Innovation, OECD Policy Report, Paris.

OECD. (2002). High-growth SMEs and employment. Paris: OED

Parker, S.C., Storey, D.J., & van Witteloostuijn, A. (2005). What happens to gazelles? The importance of a dynamic management strategy. Research Report, Entrepreneurship Research Series, Centre for Entrepreneurship, Durham Business School, Durham University, UK.

Parker, S.C., Storey, D.J., van Witteloostuijn. (2010). What happens to gazelles? The importance of dynamic management strategy. Small Business Economics 35(2).

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

Reynolds, P.D., N.M. Carter, and S.B. White (1997), The Entrepreneurial Process: Economic Growth, Men, Women and Minorities. Westport, CT: Quorum Books.

Rousseeuw, Peter J. and Christophe Croux. 1993. "Alternatives to the Median Absolute Deviation." Journal of the American Statistical Association 88 (December): 1273–1283.

Schreyer, P., 2000. High-growth firms and employment. Working Papers, 2000/3, OECD Science, Technology and Industry, Paris.

Sims, C. (1972). Money, income, and causality. American Economic Review 62: 540–552.
Stam, E. (2005). The geography of gazelles in the Netherlands. Tijdschrift voor Economischeen Sociale Geografie, 96(1), 121–127.
Stam, E. and Wennberg, K. (2009): "The roles of R&D in new firm growth", Small Business Economics, 33: 77-89. Stangler and Kedrosky (2009).
Storey, D. J. (1994): Understanding the small business sector. London: Routledge.
Sutton, J. (1997). Gibrat's legacy. Journal of Economic Literature, 35(1), 40–59.
Teruel, M. and G. deWit (2011). Determinants of high-growth firms. Why have some countries more high-growth firms than others? EIM Research Reports, H201107.
Van Praag, Miriam C. and Versloot, Peter, H. (2008), "The Economic Benefits and Costs of Entrepreneurship: A Review of the Research." Foundations and Trends in Entrepreneurship Research 4(2), 65–154.
Van Stel, A.J. and V. Stunnenberg, 2006, Linking business ownership and perceived administrative complexity, Journal of Small Business and Enterprise Development 13 (1), 7-22.
van Stel, Andre, David J. Storey, and A. Roy Thurik, 2007. "The Effect of Business Regulations on Nascent and Young Business Entrepreneurship." Small Business Economics 28, 171-186.
Weinzimmer, L. G., Nystron, P. C., & Freeman, S. J. 1998. Measuring organizational growth: Issues, consequences and guidelines. Journal of Management, 24(2): 235-262.
Wiklund, J. 1998. Small Firm Growth and Performance. Entrepreneurship and Beyond. Jonkoping, Sweden: Jonkoping International Business School (diss.).
Wiklund, J. and Shepherd, D. (2003): "Aspiring for, and Achieving Growth: The Moderating Role of Resources and Opportunities", Journal of Management Studies, 40: 1919-1941.
Wiklund, J., & Davidsson, P. 1999. A resource-based view on organic and acquired growth. Paper presented at the Academy of Management Conference, Chicago.