Sources of persistence in regional start-up rates—evidence from Sweden

Andersson, Martin; Koster, Sierdjan

Published in:
Journal of Economic Geography

DOI:
10.1093/jeg/lbp069

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2011

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 17-09-2023
Sources of persistence in regional start-up rates—evidence from Sweden

Martin Andersson* and Sierdjan Koster†

*Jönköping International Business School (JIBS), Jönköping, and the Centre of Excellence for Science and Innovation Studies (CESIS), Royal Institute of Technology, Stockholm.

(email) martin.andersson@jibs.hj.se

†Urban and Regional Studies Institute, Faculty of Spatial Science, University of Groningen.

(email) sierdjan.koster@rug.nl

Abstract

This article assesses the spatial–temporal persistence of a dynamic phenomenon: start-ups. Two mechanisms that explain persistence in start-up rates are distinguished: (i) determinants of start-ups that are spatially ‘sticky’ and durable and (ii) path dependence in start-up activities in the form of a response mechanism between previous and current start-up activities. A dynamic panel analysis of the relationship between lagged start-up rates and current start-up rates confirms that both effects are important in explaining persistence. The second mechanism implies a regional dimension in persistence, such that regions with high levels of start-up rates will exhibit stronger persistence. We find empirical evidence of this using quantile regression techniques.

Keywords: Entrepreneurship, persistence, path dependence, regional start-up dynamics

JEL classifications: L26, R11, R12, O18

Date submitted: 5 June 2009   Date accepted: 17 December 2009

1. Introduction

The relationship between start-up rates and regional economic development has been studied extensively in recent years [see Van Praag and Versloot (2007) for a review of recent studies]. Based on the ideas of Schumpeter, these studies contend that start-ups are a necessary condition for long-term regional economic development. Accordingly, there is a large interest in the geography of start-ups and its determinants among both academics and policy makers. An area that has received considerably less attention, however, is the dynamics of start-up activities across regions over time. The extent to which the spatial distribution of start-up activities is persistent or changes over time and the reasons for any such changes (or the lack thereof) remain largely unexplored in the literature.

The current article adds to the limited literature on the dynamics of start-up activities across regions over time. We provide new empirical evidence of persistence in start-up activities and try to untangle the forces behind this persistence. The latter aspect constitutes the main contribution of the article. We introduce and test two non-exclusive mechanisms that could explain persistence in start-up rates. First, start-up
rates may be influenced by regional characteristics, such as incomes, educational level and population density. Since these factors are spatially ‘sticky’ and change only slowly, start-up rates would also be expected to be persistent over time. Second, high start-up rates in a region over time could generate demonstration effects and spur the development of institutions that stimulate an ‘entrepreneurial climate’. Such a climate may sustain start-up activity in subsequent periods. The first mechanism argues that persistent differences in start-up rates across regions can be explained by persistent heterogeneity across regions. The second suggests path dependency in start-up activities themselves, meaning that current start-up activities are partly a response to similar activities in a region’s history. Such responses may be motivated by spatially bounded externality phenomena and local institutions associated with start-up activities (Martin and Sunley, 2006). Path dependence in start-up activities also implies a regional dimension in persistence because the described response mechanisms are expected to help sustain a high level of start-up activity, such that persistence is stronger in regions where start-up activities are more pronounced.

Although persistence in start-up activities has been documented in previous studies, few studies have focused on persistence as a phenomenon. The only study that deals directly with persistence in start-up rates is by Fritsch and Mueller (2007), who found strong persistence in start-up rates across German regions.¹ They focused on the policy implications and concluded that persistence in start-up rates indicates that policies to promote entrepreneurship can only be successful if they adopt a long time horizon. In relation to previous studies, the novelty of our article is that it assesses the sources of persistence and tries to conceptually embed persistence in start-up activities within the general discussion about patterns of change and path dependence in evolutionary economic geography [see, for example, Martin and Sunley (2006) and Boschma and Frenken (2006)]. Although path dependence and regional change are important topics in evolutionary economics and geography, there is little evidence of regional and temporal factors within these processes (Martin and Sunley, 2006). The spatial–temporal aspect of start-up activities is an interesting one since a start-up activity is by definition a change-phenomenon. In contrast to many static regional indicators, such as unemployment rate or skill intensity, there are no obvious reasons to expect persistence in start-up activity across regions over time.² When it comes to unemployment rates, for instance, persistence can be explained by the fact that the same individuals are unemployed over time. The same applies to persistence in skill intensity. Start-ups are different and their persistence in space over time is an example of persistence in a change-phenomenon.

In the general discussion about path dependence and regional development, there are two basic aspects of path dependence that are related to start-up dynamics. On the one hand, start-ups can initiate new paths of economic development (Garud and Karnøe, 2001). Start-ups thus introduce variety into a regional economy, which may induce long-term economic growth (Boschma and Frenken, 2006). On the other hand, start-up behaviour may itself be path dependent. Fritsch and Mueller (2007) and Van Stel and

¹ This finding has been further empirically substantiated for the UK, Portugal and the Netherlands in a special issue of Small Business Economics (2008, issue 1) on the relationship between new firm formation and employment generation.

² One could, for example, argue that a high level of start-up activity in one period would imply less start-up activity in a subsequent period because of competition effects.
Suddle (2008) interpret persistence in start-up rates as a sign of path dependence. They do not, however, address the mechanisms that potentially drive persistence in start-up rates, or conceptually relate persistence to path dependence.

Our empirical analysis consists of two parts. First, we empirically assess the two proposed mechanisms behind persistence in start-up activities, i.e. persistent regional characteristics that influence start-ups and path dependence in start-up activities. For this purpose, we employ a dynamic panel model which includes lagged start-up rates as regressors. In line with our argument, this model includes distinct reasons for correlations in start-up rates over time. One is directly through start-up rates in previous periods, i.e. state dependence. The other is through observable and unobservable heterogeneity across regions. Hence, the model includes the two mechanisms that we hypothesize may drive persistence in start-up activities. The general idea is that a positive effect of lagged start-up rates on current ones, while controlling for other determinants of regional start-ups, will be consistent with the hypothesis of path dependence in start-up activity. In the second part, we address regional differences in the strength of persistence. We apply transition probability analysis and quantile regressions. The transition probability analysis examines whether the likelihood of switching ranks, in terms of the regional level of start-up rates in a given period, is related to the previous rank. The quantile regression technique has a semi-parametric nature and allows us to test whether the estimated marginal effect of lagged start-ups rates on current start-up rates differs with different levels of start-up rates across regions.

The remainder of the article is organized in the following fashion: Section 2 presents the theoretical framework. It starts by discussing the link between persistence and path dependence and then goes on to discuss reasons for path dependence in start-up rates. Section 3 describes the data and illustrates persistence in start-up rates across Swedish municipalities using data spanning a decade. Section 4 presents the results of the empirical analysis in which regional differences in persistence are explained. Section 5 draws conclusions and suggests avenues for future research that emerge from the findings.

2. Path dependence and persistence in start-up rates

2.1. Path dependence and persistence

Studies that find persistence in the regional distribution of start-up rates generally interpret this as an indication of path dependence in regional start-up dynamics (see e.g. Fritsch and Mueller 2007; Van Stel and Suddle, 2008). The relationship may, however, be more subtle as becomes apparent when comparing the definitions of path dependence and persistence. Path dependence is the process in which later conditions are dependent on previous ones (following Martin and Sunley, 2006), such that development trajectories depend on initial conditions. Persistence, in contrast, is a statistical measure that addresses the lack of change in a phenomenon or in its distribution. Interpreting persistence in start-ups to be a sign of a path dependency in start-up activities is somewhat of an oversimplification. Rather, the relationship between persistence and path dependence should be seen as more indirect: persistence in start-up rates may be the result of path-dependent processes that underlie regional start-up behavior. Regarding the relationship between persistence and path dependence in
this way allows one to seek out and identify various sources of persistence. In the subsequent two subsections, we elaborate on two sources of persistence in start-up rates: ‘sticky’ regional characteristics that influence start-ups and path dependency in start-up activity.

2.2. ‘Sticky’ regional characteristics and persistence

In a system of differential equations in which variables change on different time scales, it can be shown that the evolution and development trajectory of the fast-adjusting variables is governed by the slower variables (Haken, 1983). In a similar way, regional characteristics that are durable and change only slowly play an important role in shaping the production possibilities and development trajectories of regions. Such characteristics can be a fact of nature or may be created by various types of investments over time. A typical example of the latter is material infrastructure in the form of buildings, roads, airports and other investments in durable interaction capacity. Johansson and Wigren (1996) introduce the concept of the ‘production milieu’, as a comprehensive term for such durable and spatially ‘sticky’ regional attributes. This concept can also be linked to production factors that are, at least in principle, mobile.

Many of the factors that influence start-up activities are rather fixed in time and fit within the concept of a slowly changing production milieu. There is plenty of evidence that regional characteristics that reflect local demand- and supply-side conditions, such as education level, innovation activity, market-size, industry structure and agglomeration economies, influence the rate of new firm formation in a region (Verheul et al., 2001). The processes that typically change such characteristics are slow moving and therefore these characteristics can be seen as part of a region’s production milieu. Educated and skilled workers are clear examples of, in principle, mobile production factors. However, research has consistently shown that highly educated and skilled workers are typically concentrated in human capital-intensive locations [see, for example, Glaeser et al. (2003)]. As such, the experience, knowledge and competence of the labour force in a region can be considered as a durable attribute from an aggregate perspective. Moreover, the production milieu of a region is also determined by agglomeration economies. Such place-specific economies of scale typically evolve into self-reinforcing and self-organized processes over time and, once materialized, constitute a durable attribute (Krugman, 1996). The observed persistence in the size distribution of cities and spatial hierarchies can be viewed as an illustration of this durability.

Start-up rates across regions are in this sense governed by durable and spatially sticky variables. Given this, one would expect to find persistence in start-up rates because of the durability of their determinants. This does not mean that start-up rates will not change from year to year. Exogenous shocks may change the opportunities for start-ups leading to year-on-year differences. However, despite possible temporal fluctuations, the regional distribution of start-up rates is expected to persist over time.

2.3. Path dependence in start-ups and persistence

The second source of persistence in start-up rates is path dependence in the start-up process itself, such that start-up activities in one period are partly a response to the
same phenomenon in previous periods. Our arguments for path dependence in start-up activities are in line with two of the general reasons for path dependence discussed by Martin and Sunley (2006): institutional hysteresis and dynamic increasing returns. Institutional hysteresis refers to formal and informal institutions being both the products and the determinants of economic exchange and behaviour (cf., North, 1990). The concept of dynamic increasing returns refers to positive feedback mechanisms including learning and the establishment of traded and untraded externalities (Arthur, 1994). The general argument for path dependency in start-ups is thus the existence of a stable institutional context that is more or less conducive to start-ups and the existence of a feedback mechanism in start-up dynamics.

Since businesses are started by people, the described path dependence should be located in the conceptualization of entrepreneurship as a result of individual action in response to perceived economic opportunities [following Shane (2003)]. Shane (2003) stresses that the decision to become an entrepreneur is influenced not only by the personal preferences of the potential entrepreneur (including risk assessment, previous experiences and locus of control), but also by the availability and perception of business opportunities in a sector or a region. The existence of business opportunities is considered as an exogenous factor, but the recognition of the opportunities is endogenous and depends on the personal characteristics, experience and attitudes of the would-be entrepreneur. Shane (2000) illustrates this point in the development of the 3D printing industry. He shows that although the new technology was generally available, only a small group of people recognized its relevance and spotted a business opportunity, and that industry experience seemed particularly important in this.

In Shane’s framework, there are at least three aspects of entrepreneurship that are relevant in explaining the existence of response mechanisms in start-up dynamics. First, the availability of business opportunities is important: this is exogenous to the entrepreneurs and defines the scope for entrepreneurship in a given region or sector. Second, the recognition and grasping of the opportunities by the potential entrepreneurs needs to be taken into account. Third, a personal preference for entrepreneurship plays a role as this influences the decision whether to turn an opportunity into a new business. Both the latter aspects are endogenous to the entrepreneurs. In relation to the two general reasons for path dependence discussed above, the first two aspects pertain to dynamic increasing returns whereas the third pertains to institutional hysteresis.

A high level of start-up activities generates new entrepreneurial opportunities (Holcombe, 2003). When new ideas are acted upon by entrepreneurs in a region by developing new products or services, which are delivered by new firms, these can generate further new entrepreneurial opportunities (Audretsch and Keilbach, 2004). Frenken and Boschma (2007) formulate a theoretical dynamic model of rank-size distributions on this basis. The model is built on a branching process with product divisions as the units of analysis. A new product division is started when either a firm or an employee decides to commercialize an innovation. In line with Schumpeter (1934), an innovation is defined as a recombination of existing resources. This means that with each new product division, the number of possible new combinations increases non-linearly. The growth in the number of new combinations (and the related product divisions) can be interpreted as path dependence in the process itself. On this basis, one can argue that a high frequency of start-ups creates economic diversity and opportunities, which can lead to new start-ups thus resulting in persistence in
start-up rates. The opportunity generating effect of start-ups has been described in the development of specific regions and sectors, such as Silicon Valley in the USA (Saxenian, 1996), the automotive industry around Detroit (Klepper, 2007) and also in the Gnosjö-region in Sweden (Johannisson 1983, 1984). Indirect empirical support is also given by, as in this article, the persistence in start-up rates, although it cannot be attributed to higher opportunities per se.

Recognizing and acting upon business opportunities are inherently processes at the individual level. However, the context in which these processes manifest themselves is important in shaping individual responses (Verheul et al., 2001). Therefore, regional differences can be anticipated in the individual assessment of opportunities and the propensity of people to actually pursue a business opportunity and start a new firm. An important argument in explaining regional differences in such ‘entrepreneurial talent’ is entrepreneurial learning, and the role of imitation and entrepreneurial role models in this. Guiso and Schivardi (2005) argue that entrepreneurial talent is not innate and maintain that when more entrepreneurs are active in a region, people will have greater opportunities to acquire entrepreneurial skills. In their framework, the individual accumulation of entrepreneurial skills is thus to an extent a function of the characteristics of the regional environment in which they operate. A region’s history of start-up activities is an important characteristic here as entrepreneurial learning depends on the regional intensity of entrepreneurs. Entrepreneurial role models have indeed been shown to have a positive impact on the propensity of people to start new firms [see, for example, Aldrich (1999), Blanchflower and Oswald (1998), Arenius and Minniti (2005)]. Knowing an entrepreneur and having an entrepreneur in the family are good estimators of entrepreneurship. Role models not only assist in developing entrepreneurial skills, but they are also a sign of the social acceptability of entrepreneurship. In addition, existing entrepreneurs may serve as bellwethers of certain business opportunities that imitative entrepreneurs may follow (Baumol, 1993). As such, this means that the recognition of opportunities is also influenced by role models. At the regional level, a wide availability of role models may thus generate ‘demonstration effects’ such that potential entrepreneurs are stimulated to develop an idea in the form of a new firm (Henrekson and Stenkula, 2007). Johannisson (1983, 1984) discusses and illustrates such an effect in the Gnosjö region of Sweden.

Apart from learning and demonstration effects, previous start-up activities may also reinforce existing formal and informal institutions (North, 1990) that support entrepreneurship, i.e. the institutional hysteresis argument. In this respect, Kenney and Patton (2005) discuss supporting institutions for entrepreneurship and stress their relevance for start-ups of law firms. Westlund and Bolton (2003) discuss the link between social capital and entrepreneurship. Etzioni (1987) points to the role of societal legitimation in entrepreneurship. Davidsson and Wiklund (1997) show that local ‘values’ and ‘beliefs’ have an impact on start-up activities. Moreover, Michelacci and Silva (2007) argue that there may be region-specific collateral that facilitates local entrepreneurs’ access to finance from local banks. Generally, one can expect a region’s history of start-up activities to influence such factors. Regions that, for whatever reason, have high start-up rates over an extended period may develop a positive climate towards entrepreneurship involving both formal and informal institutions. Such an entrepreneurial climate will exert a positive influence on the subsequent start-up of new firms (Wagner and Sternberg, 2002; Fölster, 2000).
Figure 1 summarizes the arguments regarding the sources of persistence in start-up rates within regions. The first argument is that the regional attributes that explain start-up rates change through slow processes and, consequently, start-up rates change only slowly over time. The second factor is a response mechanism, such that current start-up activities are partly a response to those in previous periods. Any change process that is driven by slowly changing regional characteristics and invariant response mechanisms will be persistent.

Accepting the existence of a response mechanism between current and past start-up activities in a region involves accepting that, on top of durable and spatially sticky characteristics, there is an enduring advantage for regions that have seen high start-up rates in the past. The path dependency in start-up activities should help to sustain a high level of start-up activity. This leads to the expectation that the strength of persistence in start-up rates is dependent on the ‘level’ of start-up rates. Regions with high start-up rates over time are assumed to have a strong entrepreneurial culture, or a well-developed entrepreneurial capital, which sustains start-up rates.

The article continues along the lines outlined in Figure 1. First, we investigate persistence in start-up rates in Swedish regions (Section 3). The second step of the analysis searches for differences in persistence related to the level of regional start-ups (Section 4).

3. Illustrating persistence in start-up rates

3.1. Data and setting the scene

For this study, we use data obtained from Statistics Sweden (SCB) that provide information on the number of start-ups in the private sector for each municipality in Sweden from 1994 until 2004. In total, this amounts to 286 municipalities for the period considered. In the data, a start-up is defined as a new establishment. The data contain FAD information distinguishing between truly new establishments and new establishments that are a result of reorganizations or changes in the ownership structure [see also, Persson (2004)]. Unfortunately, however, we are not able to discriminate between

---

3 Data have been corrected for changes in municipality classifications between 1994 and 2004.
4 FAD = Företagens och Arbetsställens Dynamik, i.e. Dynamics of Firms and Establishments. In the Business Register, each establishment is assigned a unique identity number (CFAR). For all new identity numbers, the FAD database identifies whether an establishment is truly new or a result of mergers and
new establishments started by entrepreneurs (new firms) or new establishments started
by incumbent firms. At the aggregate level, the latter type of start-up constitutes about
12% of all new establishments.

Figures 2 and 3 provide an overview of aggregate start-up dynamics in Sweden over
the study period. These figures describe the context in which the analysis must be
placed. Figure 2 presents the co-variation between the total number of start-ups, the
start-up rate (total number of start-ups divided by labour market population) and GDP
per capita over time. Both the start-up rate and the GDP per capita in each year are
expressed as the ratio between the period’s value and the average value over the whole
period, such that they are displayed on the same scale. The first conclusion is that start-
up activity has fallen over the study period while, at the same time, GDP per capita
has increased steadily. One reason for the decrease in start-up activity since 1994 could
be that 1994 marked the end of a recession. This period was associated with
high unemployment which could have forced people to become entrepreneurial.

Figure 2. Start-ups and GDP per capita in Sweden 1994–2004 (ratio between the level in each

Figure 3. Distribution of start-ups across three industry aggregates 1994–2004.
Improving economic conditions and a consequent recovery in the labour market from 1994 onwards may have discouraged start-up activities.

Figure 3 divides the start-ups per year into three industry aggregates: (i) manufacturing, (ii) low-end services; and (iii) high-end services. It is evident from the figure that most of the start-ups have been in service industries with a declining share in manufacturing industries. During the period 1994–2004, high-end services have seen the strongest growth in the number of start-ups. This is in line with developments in other European countries.

3.2. Persistence of start-up rates in Swedish municipalities

This section addresses the persistence in start-up rates across Swedish municipalities. The analysis is akin to one by Fritsch and Mueller (2007) enabling a comparison with their results based on German data. Following the labour market approach to the measurement of start-up rates (Audretsch and Fritsch, 1994), the start-up rate in each municipality is calculated by dividing the total number of start-ups in a municipality by its labour market population. As a first look at persistence in start-up rates, we investigated whether the observed change in overall Swedish start-up rates during the period 1994–2004 was associated with any change in the distribution of start-up rates across Swedish municipalities. Figure 4 presents the estimated kernel density for start-up rates across Swedish municipalities in 1994 and 2004. It is evident from the figure that the distribution of start-up rates among municipalities has a similar shape in 1994 and 2004. In line with the overall decline in Swedish start-up rates, the curve describing

---

5 Low-end services are defined by NACE code 50-64 and include retail, wholesale, hotels, restaurants and repair shops. High-end services are defined by NACE code 65-99 and include advanced producer services and R&D institutions.

---

the distribution in 2004 is to the left of the equivalent curve for 1994, but the shape of the curve remains unchanged.

A distribution can however remain stable over time even if individual locations change relative positions over the time interval. Table 1 offers a more comprehensive overview of persistence in start-up rates over time. It shows the results of 10 separate ordinary least squares (OLS) regressions with the start-up rate in 2004 as the dependent variable. Each of the regressions has a lagged start-up rate as the explanatory variable. The lag length of the independent variable varies from 1 to 10 years. The results support the idea that there is large degree of persistence in the regional distribution of start-up rates. Over a decade, the coefficients only drop slowly, and the start-up rate with the greatest lag \((t-10, 1994)\) still explains about half of the variance in the start-up rate in the base year (2004). These findings concur with the results of Fritsch and Mueller (2007) who found a similar pattern, although the explanatory power of lagged start-up rates seemed to be somewhat stronger across regions in Germany. It is difficult to directly compare both sets of results as Fritsch and Mueller used a higher level of spatial aggregation (labour market areas). A wider spatial aggregation will tend to average out possible local differences resulting in higher overall levels of persistence. Having said this, the still high levels of persistence in our study, on the lower spatial scale of the municipality, further emphasizes the significance of persistence in the regional distribution of start-up rates. Not only do static indicators (such as population and employment) show strong persistence, also the dynamic indicator of ‘start-up rates’ is very persistent over time.

Table 2 looks for a possible temporal dimension in persistence by showing the correlations between the start-up rate in one year and in previous years. Again, persistence is clearly visible in the data and, further, the table suggests that persistence is rather time invariant. Regardless of the base-year, there are strong and only gradually declining, correlations between start-up rates and their lagged counterparts.

The main conclusion from the results presented above is that the regional distribution of start-up rates is persistent over time. This can indeed be interpreted as support for the

**Table 1.** OLS regressions of start-up rate across Swedish municipalities at time \(t\) (2004) with start-up rates with various lag lengths as the independent variable

<table>
<thead>
<tr>
<th>Start-up rate ((t-\text{lag}))</th>
<th>(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start-up rate ((t-1))</td>
<td>0.81***</td>
</tr>
<tr>
<td>Start-up rate ((t-2))</td>
<td>0.84***</td>
</tr>
<tr>
<td>Start-up rate ((t-3))</td>
<td>0.93***</td>
</tr>
<tr>
<td>Start-up rate ((t-4))</td>
<td>0.83***</td>
</tr>
<tr>
<td>Start-up rate ((t-5))</td>
<td>0.83***</td>
</tr>
<tr>
<td>Start-up rate ((t-6))</td>
<td>0.79***</td>
</tr>
<tr>
<td>Start-up rate ((t-7))</td>
<td>0.85***</td>
</tr>
<tr>
<td>Start-up rate ((t-8))</td>
<td>0.76***</td>
</tr>
<tr>
<td>Start-up rate ((t-9))</td>
<td>0.67***</td>
</tr>
<tr>
<td>Start-up rate ((t-10))</td>
<td>0.73***</td>
</tr>
</tbody>
</table>

Estimated parameters of start-up rate for different lag lengths. A separate regression analysis was carried out for each lag length.

***P<0.01.
argument that policy measures geared towards promoting new company start-ups should have a long time-horizon. It also provides an interesting empirical example of how relatively static spatial distributions in economic phenomena can be explained by dynamic indicators [see also, Frenken and Boschma (2007)]. If a spatial distribution is to remain stable over time, the distribution of related dynamic indicators (such as start-up rates) needs to mimic the existing distribution. The persistence of start-up rates is a good example of such a phenomenon.6

4. Sources of persistence and variation among regions

The previous section was focused on illustrating the existence of persistence in start-up rates across Swedish regions. This section assesses (i) the sources of this persistence and (ii) whether there are any regional differences in the levels of persistence. Section 4.1 addresses the first of these issues and Section 4.2 the latter.

4.1. Sources of persistence in regional start-up activity—an empirical analysis

The conceptual framework introduced in Section 2 suggests a two-pronged explanation for persistence in start-up rates. First, the presence of slowly changing regional characteristics that are pertinent to explaining start-ups suggests there will be persistence in start-up rates within a region. Second, persistence in start-up activities may also be due to path dependency in the start-up process itself, such that start-up activities in one period are partly a response to the same phenomenon in previous periods. These two explanations are not mutually exclusive; they are complementary.

6 Fritsch and Mueller (2007) find that both the levels of start-up rates and changes in these levels are explained by the same variables, which adds to this idea.
In order to test both effects we estimate a dynamic panel data model:

\[ S_{it} = \gamma_1 S_{it-1} + \cdots + \gamma_n S_{it-n} + \mathbf{x}'_i \mathbf{\beta} + \alpha_i + \mu_t + \epsilon_{it} \]  

(1)

where \( S_{it} \) is the start-up rate in municipality \( i \) in year \( t \), and \( t-n \) denotes lag length. \( \mathbf{x}_i \) is a matrix of regional characteristics assumed to influence start-up activities. \( \alpha_i \) denotes time-invariant municipality-specific effects, \( \mu_t \) time-effects and \( \epsilon_{it} \) is an error term. In line with our theoretical framework, the dynamic panel model in Equation (1) provides various reasons for correlations in start-up rates over time (Cameron and Trivedi, 2009), namely:

(i) through start-up rates in previous periods. Current start-ups are to an extent a response to the municipality’s start-up rates in recent history, suggesting path dependency in start-up activities. Such a response mechanism is motivated by the arguments in Section 2.3.

(ii) through observed heterogeneity in the form of factors in \( \mathbf{X}_i \) assumed to influence start-up rates in a region.

(iii) through unobserved time-invariant heterogeneity captured by municipality-specific effects, \( \alpha_i \).

The dynamic panel model (1) enables us to test the empirical relevance of the arguments outlined in Section 2 and summarized in Figure 1. The model incorporates both regional characteristics that influence start-up rates and a response mechanism linking previous to current start-up rates. In addition, we control for unobserved heterogeneity.

Our choice of variables in \( \mathbf{X}_i \) is based on previous literature on the determinants of regional start-up rates. Verheul et al. (2001) provide a set of variable types that are pertinent determinants of start-ups rates within regions. They distinguish between supply, demand and institutional effects. In our analysis, the supply-side is represented by the educational level of the population, measured as the share of the population with higher education (≥3 years of university education). In addition, the regional share of all service-industry firms is entered as an indicator of the economic structure, recalling that the bulk of start-ups are affiliated to the service industry. Both variables are expected to positively correlate with start-up rates. The demand side plays a role in the regression analysis through paying particular attention to possible spatial patterns in demand. We apply an accessibility measure that describes each municipality’s access to GRP. The employed measure is based on exponential distance decay, which satisfies criteria of consistency and meaningfulness (Weibull, 1976). Specifically, demand-side conditions in municipality \( i \) at time \( t \) is approximated by \( \sum_j GRP_{jt} \exp\{-\lambda t_{ij}\} \), where \( GRP \) denotes the gross regional product of municipality \( j \), \( \lambda \) is a distance friction parameter and \( t_{ij} \) is the time distance between \( i \) and \( j \) in terms of the travelling time by car. In addition to the composite measure of demand, income is also included as it is has been shown to be an important variable in explaining variation in start-up rates (Reynolds et al., 1995). The interpretation, however, is ambiguous. While income can be seen as a demand variable reflecting the market potential of a region, it can also be interpreted as a supply-side variable indicating the availability of start-up capital. Finally, it can negatively influence start-up rates as high incomes equate to high opportunity costs in becoming an entrepreneur. The median income level is used because the distribution of income in regions is skewed. We also include the
employment rate as a further control variable. The effect of the employment rate is ambiguous [see, for example, Storey (1991), Tervo and Niittykangas (1994), Ritsilä and Tervo (2004)]. On the one hand, a high employment rate suggests a vibrant local economy with many opportunities for new businesses, suggesting a positive association with start-up rate. On the other hand, it could indicate the availability of jobs, which reduces the number of people starting a business out of necessity. The net effect, whether positive or negative, can depend on the region and period studied (Tervo and Niittykangas, 1994). Therefore, we have no firm expectations on this variable.

The explanatory variables included in the model are all stable over time, which reflects the argument of start-up activities having slowly changing determinants. We also include a dummy variable to account for metropolitan areas, namely Sweden’s three major cities (Stockholm, Göteborg and Malmö). 7

To estimate the model proposed in Equation (1), we employ the two-step system GMM estimator. In our empirical context, with regional characteristics that change slowly over time, the system GMM estimator is particularly appropriate. The literature show that pooled OLS and fixed effects models produce biased estimates of the lagged dependent variables. The system GMM estimator accounts for this bias and has some attractive properties relative to other estimators (Blundell and Bond, 1998; Blundell et al., 2000), especially the Arellano–Bond estimator (Arellano and Bond, 1991). The Arellano–Bond estimator is based on first differencing the data and then instrumenting all potentially endogenous variables with their own levels. Blundell and Bond (1998) show that the first-difference GMM estimator behaves poorly when the variance in the individual-specific fixed effects, \( \alpha_i \), is large compared to the variance in \( \epsilon_{it} \). When time series are persistent, such that variables change slowly over time, lagged levels are weak instruments for the first differenced variables. As proposed by Blundell and Bond (1998), the system GMM estimator adds moment conditions and combines first-differences and levels as it is based on a system of first-differences instrumented on lagged levels and of levels instrumented on first differences. Persistent series are one area where the system GMM estimator has superior performance to the first-difference GMM estimator (Blundell and Bond, 1998).

Table 3 presents the results from the estimation exercise. 8 The table shows results obtained with pooled OLS, the within (fixed effects) panel estimator and the two-step

---

7 Another variable that is often included as a determinant of start-up activity in a region is the average size of establishments or, alternatively, small establishments as a fraction of all establishments. Studies show that small establishments are more likely to spawn new entrepreneurs (see, for example, Hyytinen and Maliranta (2008), Rosenthal and Strange (2009)). The argument is that employees in small establishments are more exposed to the entire business process and may therefore acquire more relevant experience for starting a new business. We did not include this variable in our final empirical model because indicators of the size structure of establishments are highly correlated with some of our other regressors, primarily the share of services in the local economy. Service industries tend to be characterized by small-scale businesses. We have carried out robustness checks for our empirical model. When the fraction of small and medium-sized establishments (1–49 employees) is included in the model it is found to be an insignificant factor and does not alter our results. However, if we exclude the variable reflecting the presence of service industries, the effect is significant and has the expected sign.

8 An issue when empirically analysing spatial data is problems associated with spatial dependence (Anselin, 1988; Anselin and Florax, 1995). In our study, start-up rates across municipalities are spatially dependent. Appendix A presents a Moran’s scatter plot for start-up rates in 2004. The estimations in Table 3 do not directly address issues associated with spatial dependence. Although the market-size variable could be argued to partly capture spatial effects across municipalities (cf. Andersson and Gräsjö, 2009), we know of no formal test for assessing this issue in a dynamic panel data context. As a crude check, we have
system GMM estimator. The estimated parameters of the lagged start-up rates obtained using the system GMM estimator pass the ‘bounds test’ (Roodman, 2006) and lie between the estimates obtained using the pooled OLS (upward bias) and the within (downward bias) alternatives. In the system GMM estimator, standard errors are calculated using the Windmeijer (2005) correction and are robust to heteroscedasticity. Here, all the variables, except the time dummies, the dummy for metropolitan areas and external market size, are specified as endogenous. The estimations are made using three lags of the dependent variable and the test statistics regarding autocorrelation and validity of the instruments are satisfactory. First, the tests for first- and second-order autocorrelations, AR(1) and AR(2), indicate no problem. As one hopes, the null hypothesis in the AR(1) test is rejected, and the null hypothesis in the AR(2) test is not rejected. Second, the Hansen test for determining overidentifying restrictions is a satisfactory check on the validity of the instruments. Here, the null hypothesis is not rejected. While the Hansen test is robust, it can be weakened by having too many instruments. Here, the number of instruments (260) is less than the number of groups (286).

Table 3. Dynamic panel estimations. Estimated parameters for the explanatory variables in a regression equation with start-up rates in Swedish municipalities as the dependent variable

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pooled OLS</th>
<th>Within (fixed effects)</th>
<th>Two-step system GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lagged start-up rate ( t-1 )</td>
<td>0.307*** (0.028)</td>
<td>-0.059** (0.028)</td>
<td>0.264*** (0.028)</td>
</tr>
<tr>
<td>Lagged start-up rate ( t-2 )</td>
<td>0.222*** (0.026)</td>
<td>-0.069** (0.028)</td>
<td>0.187*** (0.035)</td>
</tr>
<tr>
<td>Lagged start-up rate ( t-3 )</td>
<td>0.269*** (0.024)</td>
<td>-0.030 (0.028)</td>
<td>0.227*** (0.026)</td>
</tr>
<tr>
<td>Education intensity</td>
<td>0.017*** (0.005)</td>
<td>0.013 (0.017)</td>
<td>0.018** (0.008)</td>
</tr>
<tr>
<td>Services (share)</td>
<td>0.431*** (0.131)</td>
<td>0.675 (0.758)</td>
<td>0.814** (0.321)</td>
</tr>
<tr>
<td>Market-size</td>
<td>0.678*** (0.259)</td>
<td>4.053 (2.472)</td>
<td>1.461*** (0.498)</td>
</tr>
<tr>
<td>Income (log)</td>
<td>0.192 (0.411)</td>
<td>-1.736*** (0.467)</td>
<td>-0.909 (0.853)</td>
</tr>
<tr>
<td>Employment rate</td>
<td>-1.051** (0.481)</td>
<td>-2.285 (1.555)</td>
<td>-0.435 (0.906)</td>
</tr>
<tr>
<td>Metropolitan dummy</td>
<td>0.185* (0.105)</td>
<td>-</td>
<td>0.450 (0.309)</td>
</tr>
<tr>
<td>First-order autocorrelation AR(1) ( P)-value</td>
<td>-</td>
<td>-</td>
<td>0.00</td>
</tr>
<tr>
<td>Second-order autocorrelation AR(2) ( P)-value</td>
<td>-</td>
<td>-</td>
<td>0.98</td>
</tr>
<tr>
<td>Hansen test for overid. restrictions ( P)-value</td>
<td>-</td>
<td>-</td>
<td>0.21</td>
</tr>
<tr>
<td>Number of instruments</td>
<td>-</td>
<td>-</td>
<td>260</td>
</tr>
<tr>
<td>Number of groups</td>
<td>286</td>
<td>286</td>
<td>286</td>
</tr>
<tr>
<td>Number of observations</td>
<td>2288</td>
<td>2288</td>
<td>2288</td>
</tr>
</tbody>
</table>

The two-step system GMM estimation is performed with the `xtabond2` (Roodman 2006) command in the STATA 9.2 package. Standard errors are estimated using the Windmeijer (2005) correction. The two-step system GMM estimation is specified with the time dummies, the metropolitan dummy and the regional market size as exogenous variables (IV-style) and the remaining variables as endogenous ones (GMM-style). The null hypothesis in the test for first-order autocorrelation, AR(1), is that autocorrelation is present. The null hypothesis in the test for second-order autocorrelation, AR(2), is that there is no autocorrelation. The test statistics for AR(1) and AR(2) are satisfactory. The test statistic resulting from the Hansen test for overidentifying restrictions (validity of the instruments) is satisfactory. The null hypothesis is not rejected. While the Hansen test is robust, it can be weakened by having too many instruments. Here, the number of instruments (260) is less than the number of groups (286).

*** \( P<0.01 \), ** \( P<0.05 \), * \( P<0.1 \); Robust standard errors in parentheses.
not rejected. Although the Hansen test is robust, it can be weakened by having too many instruments. However, in our model there are 260 instruments which is less than the number of cross-section units (286), and so the number of instruments should not be an issue (Roodman, 2006). The estimated parameters obtained with the two-step system GMM estimator are conditional on municipality-specific fixed effects and take account of endogeneity associated with the regressors. As such, this approach enables a more causative interpretation than pooled OLS and fixed effects models. Consequently, we focus on the results obtained using the two-step system GMM estimator.

The results in the table are supportive of the idea that start-up rates over time in a municipality can be explained by both regional characteristics and previous start-up rates. The lagged start-up rates are statistically significant and the results suggest that previous start-up activities do have an effect on current start-up activity in a municipality after controlling for other determinants of start-ups. This is consistent with there being path dependency in start-up activities. Moreover, the regional characteristics are significant and support the findings in previous literature. The relationships between education level, market-size, share of services in the local industry and start-up activities are positive and statistically significant. Although the level of income and the employment rate are found to be negatively associated with start-ups, these relationships are not statistically significant.

In Table A1, we present the estimated parameters of these variables when lagged start-up rates are excluded, and this does not change the pattern seen in Table 3. We conclude from the results in Table 3 that the two sources of persistence in start-up rates are confirmed. In addition to the influence of slowly changing regional characteristics, the results suggest a response mechanism linking previous to current start-up rates, indicating path dependence in start-up activities. These two sources of persistence operate alongside each other and have individual effects. In the next subsection, we turn to the regional dimension in persistence.

### 4.2. Explaining regional persistence levels

A response mechanism linking current to past start-up activities in a region implies that, in addition to durable and spatially sticky characteristics, there is an extra enduring advantage for regions that have shown high start-up rates in the past. As argued previously, this leads to the expectation that the strength of persistence in start-up rates will be dependent on the level of start-up rates. Regions with high start-up rates over time are assumed to have developed a strong entrepreneurial culture, or well-developed entrepreneurial capital, which induces persistence in start-up rates. The analysis so far has empirically verified the relevance of slowly changing regional attributes and response mechanisms from previous to current start-up rates. To further substantiate the analysis, this section first presents results from a transition probability analysis and then considers regression quantiles.

Figure 5 presents a graphical representation of the strength in persistence related to the level of start-up rates. It shows the likelihood that a region will retain, over a ten-year period, its rank (defined using 10 groups of equal size) when regions are sorted according to start-up level. A high probability therefore indicates that most regions assigned to a particular rank-group in 1994 are still in the same one 10 years later.

As is evident from the figure, persistence is particularly strong at the extremes of the distribution. This indicates that municipalities with either very low or very high levels of
start-up rates are the most likely to remain in the same group over the decade analysed (1994–2004). For high levels of start-up rates, this is in line with the theoretical framework outlined earlier. However, a general problem with rank analyses is the fact that the probability of shifting groups is, by definition, lower in the extreme rank-groups, simply because members of these groups can only move in one direction: regions in the group with the highest (lowest) start-up levels can only move downwards (upwards). In the middle of the distribution, regions can move in either direction. This systemic bias may be responsible for inflating the level of persistence at the extremes. However, it does not undermine the overall analysis because persistence at the high end of the distribution is much stronger than at the lower end. This trend cannot be the result of the above-discussed bias, but it is in line with the argument that dynamic increasing returns makes a region less vulnerable to external shocks and that this results in persistently high start-up rates.

To substantiate the transition probability analysis, we also apply quantile regression analysis [see Koenker and Hallock (2001), Buchinsky (1998), Rogers (1992)]. The systemic bias related to rank analyses is not an issue with this analysis. The quantile regression technique is semi-parametric, since the parameter estimates for the marginal effects of the explanatory variables are allowed to differ across the quantiles of the dependent variable. The quantile regression technique thus allows us to test whether the estimated marginal effect of lagged start-up rates on current start-up rates differs across
the distribution of start-up rates across regions. If the relationship found between current and previous start-up rates is stronger in municipalities with higher levels of start-up rates, this will support our hypothesis.

We apply the quantile regression technique to the relationship between start-up rates in 1994 and 2004, i.e. over a decade, across municipalities in a cross-sectional setting. First, we estimate the relationship between start-up rates in 2004 and 1994 for each quantile without including any additional control variables. We then perform a similar analysis but included all the control variables measured in 2004, i.e. education intensity, market-size, share of services, income, employment share and the metropolitan dummy. Further, because heteroscedastic data tend to underestimate standard errors, we apply a bootstrapping procedure.

Figure 6a and b presents the estimation results, respectively with and without the control variables. In each figure, the horizontal axis covers the different quantiles, and the vertical axis the magnitude of the estimated coefficients of the independent variable in question, i.e. start-up rates in 1994. The shaded area around the line represents the 5% confidence interval of the coefficient estimate.

The results of the quantile regression analysis illustrated in Figure 6a and b supports the conclusions drawn from Figure 5: that the strength in persistence is related to the level of start-up activity in a region. The magnitude of the estimated coefficient for the marginal effect of start-up rates in 1994 on start-up rates in 2004 steadily increases for successive quantiles. That is, the relationship between current and previous start-up rates is stronger in regions with higher start-up rates. This relationship cannot be explained by sticky regional characteristics alone. Although these characteristics can explain persistence as such, we see the relationship found between the level of start-up rate and the level of persistence as an indication of the existence of a response mechanism that works in such a way that previous start-up activities influence later start-up activities. The results are consistent with start-up activity being associated with institutional hysteresis and dynamic increasing returns, which explains why persistence is stronger in regions with high levels of start-up rates.

5. Conclusions

Persistence is a widespread phenomenon in the economy. Many economic indicators only change slowly over time and have a stable regional distribution. This article ties in with the general discussions about persistence by assessing persistence in a dynamic phenomenon: the start-up of new firms. Persistence in start-up activities across space over time is an example of persistence in a change-phenomenon.

The article distinguishes between two sources of persistence in regional start-up rates. The first being that the processes that change the determinants of regional start-up activity are slow moving and, as a result, changes in start-up rates will also be slow processes and can be expected to be persistent over time. The second source is path dependence in the start-up process itself. Localized learning and demonstration effects can colour the entrepreneurial climate, and so stimulate future start-up activities. As argued in the article, this suggests a response mechanism from previous to current start-up activity.

The empirical analyses show strong persistence over time in Swedish start-up rates. This is in line with previous results for Germany (Fritsch and Mueller, 2007).
By estimating a dynamic panel model, we find support for the view that both sources of persistence are significant and operate simultaneously. In addition to regional characteristics that reflect demand- and supply-side conditions, previous start-up rates have a significant effect on current start-up rates. The results are thus consistent with the view that persistent change-processes, in the form of start-up activities across space over time, can be explained by slowly changing regional characteristics and response mechanisms operating at the regional level. We also show that while

![Graph](https://example.com/graph.png)
persistence is strong in overall terms, there is a regional dimension to the level of persistence: regions with high start-up rates demonstrate greater persistence than regions with low start-up rates. This is consistent with the hypothesis that an entrepreneurial climate may develop in regions with high levels of start-up activity, and that this climate will help to sustain a high level of start-up activity over time such that persistence is stronger in such areas. The article emphasizes two components of such an entrepreneurial climate: localized externalities and formal and informal institutions conducive for start-up activity.

The article adds to the empirical evidence on persistence in start-up rates. It has also tried to link the phenomenon of persistence in start-up rates to the more general literature on path dependence in regional economic development. In this light, persistence in dynamic phenomena, such as start-up rates, is an interesting case. It may for instance be related to static spatial distributions, such as the city size distribution. If a static distribution remains stable over time then the dynamic processes that maintain that distribution in theory should have a similar distribution. A further scrutiny of persistence in dynamic processes, and its drivers, appears a fruitful line of research in order to further understand the different types of path dependency and their roles in economic development. One particularly relevant issue in this respect is assessing when the response mechanism is strong enough in order to sustain a stable entrepreneurial climate. This would involve further scrutinizing these processes, not only at the high end of the distribution, but also at the low end.

We end with a final remark on the policy implications of our findings. With regard to the question which regions to aim for with start-up policies, there are in principle two basic contrasting perspectives. On the one hand it could be argued that policy efforts should be concentrated to regions with already established entrepreneurial climate, as the effects of a policy may be higher in these regions, e.g. more people willing to opt for starting new businesses. On the other hand, one could argue that policy efforts should instead be concentrated on the lagging regions as the leading regions will be fine anyway. For regions with low start-up activity, our findings may raise questions about the scope for policy makers to make effective interventions. The results indeed demonstrate that raising start-ups in regions with persistently low start-up rates is a difficult task, with durable processes working against the objectives of policy. Given this, policy aiming at a real influence on start-up activity and the long-term development in these regions most likely need to be catalytic in nature, which are able to alter pertinent slowly changing features of the regions (cf. Andersson and Johansson, 2010). Such catalytic policy measures could, for example, comprise measures to increase in-migration of people with entrepreneurial skills and competencies through e.g. novel housing policies, it may comprise of the establishment of new R&D centres, or it could focus on stressing entrepreneurial skills in education. Although the specific policy measures may differ from region to region, the main implication seems to be that policy should focus on influencing the structural elements of a regional economy. This in turn can then influence the entrepreneurial activity in the long run. It is clear that these change processes are inherently slow and that policy measures should have a long time horizon. Such a policy strategy appear as having higher potential than start-up policies that focus on small adjustments of the conditions for starting new firms, e.g. start-up subsidies in lagging regions.
Funding
EU sixth framework programme (MICRODYN project) and Forum for Reforms, Entrepreneurship and Sustainability (FORES).

Acknowledgements
We are grateful for comments and remarks from two anonymous referees as well as by Börje Johansson, Pontus Braunerhjelm, Andreas Stephan, Aleid Brouwer and by seminar participants at the ERSA Congress in Liverpool 2008 and the EMAEE conference in Jena 2009. We also thank Jan Andersson at Statistics Sweden for support with data issues.

References
Sources of persistence in regional start-up rates


**Appendix**

A. Moran’s scatterplot for start-up rates in 2004.

**Figure A1.** Moran’s scatterplot for start-up rates across municipalities 2004. Each element $w_{ij}$ in the applied spatial weight matrix is the inverse of the time distance by car between municipality $i$ and $j$. $w_{ij} = 0$ if the time distance between $i$ and $j$ exceeds 120 min.
Table A1 Static panel estimations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pooled OLS</th>
<th>FE (within)</th>
<th>FEVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education intensity</td>
<td>0.068***</td>
<td>0.078***</td>
<td>0.078***</td>
</tr>
<tr>
<td>Services (share)</td>
<td>2.565***</td>
<td>1.315**</td>
<td>1.315***</td>
</tr>
<tr>
<td>Market-size</td>
<td>1.739***</td>
<td>0.728</td>
<td>4.330***</td>
</tr>
<tr>
<td>Income (log)</td>
<td>−0.304</td>
<td>−3.652***</td>
<td>−3.652***</td>
</tr>
<tr>
<td>Employment rate</td>
<td>−1.673***</td>
<td>1.529</td>
<td>−0.463</td>
</tr>
<tr>
<td>Metropolitan dummy</td>
<td>1.416***</td>
<td>–</td>
<td>1.531***</td>
</tr>
<tr>
<td>R-square</td>
<td>0.414</td>
<td>0.176</td>
<td>0.750</td>
</tr>
<tr>
<td>Number of observations</td>
<td>3146</td>
<td>3146</td>
<td>3146</td>
</tr>
</tbody>
</table>

Estimated parameters for the explanatory variables in a regression equation with start-up rates in Swedish municipalities as the dependent variable.

***P<0.01, **P<0.05, *P<0.1; robust standard errors in parentheses.

FEVD refers to the fixed effects vector decomposition model developed by Plümper and Troeger (2007). This estimates the effect of time-invariant and almost time-invariant variables in a fixed-effect setting. In the calculation, income, market-size and the metropolitan dummy variable are specified as time-invariant.