Kiel

Working Papers

Kiel Institute for the World Economy

Start-up Complexity and the Thickness of Regional Input Markets

Dirk Dohse,
Andrea Vaona

No. 1842 | June 2013

Web: www.ifw-kiel.de

Kiel Institute for the World Economy, Hindenburgufer 66, 24105 Kiel, Germany

Kiel Working Paper No. 1842 | June 2013

Start-up Complexity and the Thickness of Regional Input Markets

Dirk Dohse and Andrea Vaona

Abstract:
Although there is a large and rapidly growing literature on the determinants of regional variation in new firm formation, relatively little is known about the interrelation between the characteristics of start-up firms and urban structure. It is only recently that scholars of urban economics have suggested a theoretical link between the thickness of regional input markets and the complexity of feasible start-ups. The current paper classifies start-ups in different industry groups according to their complexity and analyzes the impact of regional input market thickness on the frequency of start-ups with different degrees of complexity. We find that thicker input markets do indeed foster more complex start-ups, but that some inputs are more important than others.

Keywords: Entrepreneurship . Start-up complexity . Thick markets . Regional analysis.

JEL classification: L26, D22, R12, M13, O31

Dirk Dohse
Kiel Institute for the World Economy

D-24100 Kiel, Germany
Telephone: ++431-8814-460
E-Mail: dirk.dohse@ifw-kiel.de

Andrea Vaona
Kiel Institute for the World Economy
and University of Verona, Italy

E-Mail: andrea.vaona@univr.it

1. Introduction

Although there is a large and rapidly growing literature on the determinants of regional variation in new firm formation, relatively little attention has been given to differences in the locational needs of different kinds of start-ups. The large majority of research in the field is just concerned with the determinants of total start-up activity, thereby neglecting the fact that, for example, the locational needs of new services firms might differ fundamentally from those of new manufacturing firms or that high tech firms rely on other regional resources than low tech firms.[1]

This neglect is even more surprising since the fact that different kinds of new ventures might have different locationals needs and advantages is known for a long time. The idea that large cities have a higher demand for specialized products and services, making them particularly attractive places for new firms that can satisfy this demand can be traced back to Adam Smith (Smith 1776/2001, Glaeser et al. 2009). So, obviously, Smith is aware of a link between the characteristics of firms and the characteristics of the cities in which they operate. This also holds true for Alfred Marshall (1920) who described the flow of ideas from person to person as an external economy that enhanced entrepreneurship and innovation in cities. He suggested an entrepreneurial chain where "subsidiary trades grow up in the neighborhood, supplying it with implements and materials, organizing its traffic, and in many ways conducing to the economy of its material." (Marshall 1920, p. 225) An obvious implication of Marshall's logic is that entrepreneurs will congregate next to one another (Glaeser et al. 2009) and that there are "… places that are intrinsically full of new ideas and a spirit of change" (Glaeser 2007:19), which, in turn enable advanced forms of entrepreneurship which are not possible elsewhere. In the early 1960s, both, Vernon (1960) and Chinitz (1961) have emphasized the importance of local inputs to entrepreneurship, and Chinitz (1961) has demonstrated that such inputs facilitating entrepreneurship were much more prevalent in New York than in Pittsburgh.

So, whereas the principal idea that there is a close interrelation between the characteristic features of regions and the characteristic features of new firms that emerge in these regions is not new, there is little theoretical and even less rigorous empirical research on this issue.[2] It is only recently that scholars of urban economics have suggested a theoretical link between the thickness of regional input markets and the complexity of feasible start-up projects (Helsley and Strange 2011).

The current paper starts with a parsimonious version of the model proposed by Robert Helsley and Will Strange (2011) and derives two empirically testable hypotheses from the theoretical model (Section 2). Section 3 lays the foundations for the empirical analysis by classifying start-ups in different industries according to their complexity and by suggesting different measures of regional input market thickness. Section 4 provides an in-depth econometric analysis, investigating the

[1] Among the few exceptions are studies by Audretsch and Dohse (2007) and Audretsch et al. (2010) who distinguish young and new firms according to their knowledge-intensity.

[2] In a recent survey article on entrepreneurship and urban economics Glaeser et al. (2009) state that "… there are still many unanswered questions … There is evidence that entrepreneurship increases with the availability of relevant inputs, but we do not know which inputs are most important. For many specific inputs, we still lack well identified estimates of the impact of input availability on entrepreneurship."(Glaeser et al. 2009: 26)

impact of regional input market thickness on the frequency of start-ups with different degrees of complexity. Section 5 presents and discusses the results and Section 6 concludes.

2. Model and Hypotheses

This section presents a parsimonious version of the Helsley/Strange (HS 2011, for short) model, focusing on the core relationship between input market thickness and start-up complexity and fading out more specific aspects like differences in entrepreneurs' individual skills and entrepreneurial skill balance.[3] At the end of the section the main results are summarized and two empirically testable hypotheses are provided.

Note that start-up projects in the remainder of this paper are understood in a broad sense, including all tasks necessary to put an entrepreneurial idea into practice, to enter the market and to survive the first critical months after entry. It is acknowledged that, in this sense, even the simplest and most conventional start-up project is characterized by some degree of complexity. It is, however, also evident that some start-up projects are much more complex than others, i.e. that there exists a substantial variation in start-up complexity.

2.1 Assumptions and definitions

The basic assumptions and definitions used here are – unless otherwise stated – the same as in HS 2011. It is assumed that potential entrepreneurs are immobile[4] and that start-up projects are spatially indivisible, meaning that all tasks of a given start-up must be performed in the same region. Each task i requires a specialized regional input and some entrepreneurial skills. It is further assumed that potential entrepreneurs are equally skilled at all tasks and that the entrepreneur's skill level is given by $\gamma > 0$ for all i. Potential entrepreneurs start a project if the expected payoff is higher than some exogenous reservation payoff $\overline{U}$.

Each start-up project involves tasks numbered by i = 1,2,…N. The higher the number of tasks that have to be performed, the higher the complexity of the start-up project. For each task i, y_i describes the characteristic of the regional input that is best suited for the completion of this task.[5] The value of y_i is assumed to be unknown when the project is started.

The regional economy contains M specialized, non-congestible inputs. Each regional input has a particular skill or ability x_j, j=1,2,…,M. Formally, x_j is assumed to be an address on the unit circle. Assumed that local inputs are evenly spaced on the unit circle, M characterizes the thickness of the local input market. Whether a start-up project is feasible depends on two closely related parameters: *adaptation distance* and *completion time*.

[3] The interested reader is referred to the original HS (2011)-model which is much richer in detail than the 'slim version' presented here.
[4] Recent empirical work (Michelacci and Silva 2007) suggests that this is not an unrealistic assumption.
[5] Formally, y_i is assumed to be an address on the unit circle.

Following HS (2011), the *adaptation distance* d_i for task i is defined as the distance between the best available local input and the needs of task i (in formal terms $d_i = \text{Min}_j |x_j - y_i|$) and *completion time* t_i for task i is defined as $t_i = d_i/\gamma$ with $\partial t_i/\partial\gamma < 0$ and $\partial t_i/\partial d_i > 0$. Quite obviously, completion time is lower the higher is the skill of the entrepreneur, and the closer is the match between the best available local input and the project needs.

Under these assumptions, completion time for the critical (i.e. most time-consuming) task $t_{\max}$ is determined by the worst of the best matches between task needs and available regional inputs, i.e,:

(1) $t_{\max} = \max_i \{t_i\} = (1/\gamma)\max_i d_i$

It is further assumed that the task needs y_i are independent draws from a uniform distribution on the unit circle. As there are M evenly spaced local inputs the distance between adjacent resources is 1/M

2.2 Expected maximum adaptation distance and expected completion time of start-up projects

As for all tasks i (i = 1, …, N) the value of y_i is unknown in advance, there is uncertainty with respect to the value of the two central model parameters adaptation distance and completion time. In other words: Potential entrepreneurs have to form expectations with respect to (maximum) adaptation distance and completion time in order to assess the profitability (and thus feasibility) of a potential start-up project in advance. Denote the largest of N realizations of d_i (i.e. the maximum adaptation distance for a given project) as d_N. Some simple algebra[6] yields the *expected maximum adaptation distance*:

(2) $$\text{E}[d_N] = \frac{1}{2M}\frac{N}{N+1}$$

Inserting (2) in (1) yields the *expected completion time*:

(3) $$\text{E}[t_{\max}] = \frac{1}{2M\gamma}\frac{N}{N+1}$$

As laid out in Section 2.1 the complexity of a start-up project increases with the number of tasks that have to be performed. Differentiating $\text{E}[d_N]$ and $\text{E}[t_{\max}]$ by the number of project tasks N yields:

(4a) $$\frac{\partial E[d_N]}{\partial N} = \frac{1}{2M}\frac{1}{(N+1)^2} > 0 \text{ and}$$

[6] See Helsley and Stange 2011, p. 553 or appendix A1 for details.

(4b) $$\frac{\partial E[t_{\max}]}{\partial N} = \frac{1}{2M\gamma}\frac{1}{(N+1)^2} > 0$$

It can be seen that both, expected difference between available resources and task needs and expected completion time increase with start-up complexity.

Differentiating E[d_N] and E[$t_{\max}$] by input market thickness M yields

(4c) $$\frac{\partial E[d_N]}{\partial M} = \frac{-1}{2M^2}\frac{N}{(N+1)} < 0 \text{ and}$$

(4d) $$\frac{\partial E[t_{\max}]}{\partial M} = \frac{-1}{2\gamma M^2}\frac{N}{(N+1)} < 0$$

Hence, regional input market thickness reduces both, the expected adaptation distance and expected completion time.

Building on the basic insights yielded so far we now move on to the key questions dealt with in this paper: (i) How do input market thickness and start-up complexity affect the decision whether or not to enter the market?, and: (ii) How are input market thickness and start-up complexity related?

2.3 Complexity, input market thickness and the entry decision

As stated in Section 2.1, a potential entrepreneur will only enter the market if her expected payoff from running the start-up project exceeds some exogenous reservation level $\overline{U}$.

The project value (at time 0) can be written as:

(5) PV = $e^{-rt_{\max}}\left(R - C\right)$

Where R denotes the total project revenue, C = $\sum_{i=1}^{N} C_i$ total project costs, r is the market discount rate and $t_{\max}$ = $\max_i\{t_i\}$ is the completion time for the critical task (i.e the task with the longest duration time).

Assuming logarithmic preferences, the potential entrepreneur’s utility can be written as:

(6) U(PV) = ln PV = ln(R-C)-r $t_{\max}$

As completion time $t_{\max}$ is uncertain at the beginning of the project, a potential entrepreneur will only become active if

(7) E[U(PV)]= ln(R-C)-rE[$t_{\max}$] > $\overline{U}$

Inserting the formular for E[$t_{\max}$] from equation (3) into (7) yields

(8) $$E[U(PV)] = \ln(R\text{-}C) - \frac{r}{2M\gamma}\frac{N}{N+1} > \overline{U}$$

Since r > 0 and the comparative static apart from that is the same as in equations (4b) and (4d), it is obvious that:

(9a) $\frac{\partial E[U(PV)]}{\partial N} < 0$ and (9b) $\frac{\partial E[U(PV)]}{\partial M} > 0$

Hence, the likelihood that a start-up project is initiated decreases, ceteris paribus, with the complexity of the project and increases wih the thickness of the regional input market. Moreover, from the implicit function theorem it follows that:

(10) $dN/dM = N(N+1)/M > 0$ and $d^2N/dM^2 = 2N^2(N+1)/M^2 > 0$

Equation (10) motivates an important part of the empirical analysis in this paper as it implies that the complexity of a marginally feasible project (N) increases with the level of regional input market thickness (M).

2.4 Summary and hypotheses

In a nutshell, regional input market thickness appears to be a key determinant of entrepreneurship, as it impacts potential start-up projects in several (interrelated) ways. A higher regional input market thickness does (ceteris paribus):

- Reduce the expected difference between available resources and task needs (4c)
- Reduce expected completion time (4d)
- Increase the expected pay off of initiating a new start-up project with a given level of complexity (9b)
- Increase the complexity of a marginally feasible start-up project, i.e. allow the initiation of rather complex start-ups which are not feasible in regions with lower input market thickness (10)

Hence, the theoretical results suggest the following hypotheses concerning the impact of input market thickness on regional start-up activity:

H1: *Thick regional input markets have a positive impact on regional start-up activity in general.*

H2: *The positive effect of input market thickness on start-up activity increases with the complexity of the start-ups.*

The following sections analyze the hypothesized relationship between input market thickness and start-up complexity empirically. This requires – in a first step – to conceptualize and measure start-up complexity and input market thickness in an operational way.

3. From Theory to Empirical Testing: The Operationalization of Start-up Complexity and Input Market Thickness

The main aim of the empirical part of this paper is to find out how regional input market thickness does affect start-up projects of different complexity. In pursuing this aim, we are faced with two major challenges: The first challenge is to categorize different (groups of) start-ups according to their complexity. The second major challenge is to operationalize input market thickness.

3.1 Making sense of start-up complexity

In the theoretical model laid out in Section 2 an entrepreneurial project is the more complex, the more tasks it involves. In practice, however, the number of tasks that a start-up project involves is per se very large and hardly observable. The problem is aggravated by the fact that different tasks are hardly comparable since some tasks are more complicated, time-consuming, etc. than others. Hence, the question arises how to operationalize the concept of start-up complexity for the purpose of empirical testing. Our proposal is as follows: In a first step the total observable population of start-ups in a country is subdivided into regions and industries. In a second step all available information concerning the complexity of start-up projects in the different industries is gathered and used to classify start-ups in different industries according to their (average) degree of complexity. This procedure enables us to identify industries with highly complex start-ups and to distinguish them from industries with average- or low-complexity start-ups.

In doing so, we make use of two unique and complementary data bases for Germany, namely the Mannheim Enterprise panel and the KfW / ZEW Start-up panel. The Mannheim Enterprise Panel provided by the Centre for European Economic Research (ZEW Mannheim) is based on the database of Creditreform, Germany's largest credit rating agency. The Mannheim Enterprise Panel provides information about the number of start-ups in Germany differentiated by region and by industry (see Almus et al. 2000 and Metzger and Heger 2005 for a detailed description). It does, however, not include detailed information on the firms' founders and on the complexity (in terms of size, sophistication, international reach, etc.) of the start-ups.

In order to improve the availability of data on newly founded firms in Germany, ZEW Mannheim, KfW Bankengruppe[7] and Creditreform are cooperating to provide the KfW/ZEW Start-up Panel, which draws on the same parent population (the Creditreform database) as the Mannheim Enterprise Panel (Fryges et al. 2010). Each of the yearly panel waves contains data on about 6.000 start-ups from almost all industries (ibid).

Table 1 gives an overview of the industry composition (using NACE revision 1 code) of the KfW/ZEW Start-up Panel. The sample is stratified by ten industries, ranging from cutting-edge technology manufacturing to wholesale and resale trade.

[Insert Table 1 about here]

[7] KfW is Germany's largest state-owned promotional bank.

An important feature of the KfW/ZEW Start-up Panel is its large cross-sectional dimension that allows sound investigations of the characteristics of newly founded firms (Fryges et al. 2010: 124). It contains, inter alia, information about the number of founders, average employment, development of employment over time, investment, R&D activity and some other features of newly established firms in their first years. Evaluations of the first waves (Gottschalk et al. 2008, Fryges et al. 2009) show that these characteristics differ substantially across industries. Hence, the KfW/ZEW Start-up Panel forms an excellent basis to classify start-ups in different industries according to their average degree of complexity.

As it is hardly possible to measure start-up complexity and classify industries by a single indicator we consider a series of different indicators provided by the KfW/ZEW Start-up Panel. These indicators include the average size (in terms of employment) of start-ups in different industries in their 1st year (i.e. the year of market entry), employment size in subsequent years, the average number of founders, the share of team start-ups, the share of firms with own R&D, and the share of firms with market novelties, differentiated by novelties to the national market and novelties to the world market (see Table 2).

The underlying idea is that start-ups are on average more complex, the larger they are (in terms of employment / people involved), the faster they grow, the more heterogeneous human capital they require, the more highly sophisticated activities they include, the more innovative they are and the larger their international reach / their international aspirations are.

[Insert Table 2 about here]

Applying the measures indicated in table 2 to the ten sectors/fields of technology by which the KfW start-up panel is stratified yields the classification of industries by start-up complexity displayed in Table 3. A detailed description of the classification procedure is provided in the appendix.[8]

[Insert Table 3 about here]

Start-ups in cutting-edge manufacturing (CTM), in high tech manufacturing (HTM), in technology intensive services (TIS) and in software supply and consulting (SSC) clearly stand out in terms of their average complexity as can be seen from the analysis in Appendix 2. These four industries all have a significantly higher share of firms with own R&D, of team-start-ups, of firms with international market novelties and a significantly higher number of founders.[9] Start-ups in these four industries (CTM, HTM, TIS and SSC) are therefore grouped together and classified as *highly complex start-ups in a narrow sense.*

Start-ups in non-high tech manufacturing (NHM) – although less outstanding than the four industries named before – are nevertheless significantly above average according to three criteria (see Appendix 2 for details). This is taken into account by definition of a second (broader) group of

[8] The authors thank Sandra Gottschalk and Martin Murrmann, both ZEW Mannheim, for their advice and generous support with the data. All remaining errors are ours.

[9] Moreover, start-ups in CTM, HTM and SSC are also on average larger or grow faster than start-ups in other industries.

industries with complex start-ups, which we call *complex start-ups in a broader sense*. This group contains the group of highly complex start-ups in a narrow sense as a sub-group, plus start-ups in non-high tech manufacturing (NHM). The broader definition (including NHM) will primarily be used to check the robustness of results yielded w.r.t. the narrow definition of complexity.

On the other side of the spectrum, start-ups in customer oriented services (COS), in wholesale and retail trade (WRT) and in construction (CST) appear to be the ones with the lowest average start-up complexity (see Table A2 and Appendix 2 for details). Hence, start-ups in these three industry groups are classified *as low-complexity start-ups*.

Our classification rests on the assumption that the relative complexity of start-ups across industries is stable over time, as data from KfW/ZEW Start-up Panel are only available from 2005 onwards. Robustness checks for different cohorts of start-ups (2006, 2008, 2010) give, however, no hints on notable changes of relative start-up complexity across industries over time.

3.2 Dimensions of input market thickness

The second major challenge is to operationalize input market thickness. HS (2011) refer to input market thickness in a rather general way, stating that they are agnostic about the specific forms of market thickness which create ".. an environment that is most conducive to entrepreneurial adaptation" (Helsley and Strange 2011, pp 552). In a related paper, Glaeser et al. (2009) argue that although there is evidence that the availability of relevant inputs increases entrepreneurship " … we do not know which inputs are most important. For many specific inputs, we still lack well identified estimates of the impact of input availability on entrepreneurship." (Glaeser et al. 2009: 26) In fact, there are various different inputs discussed in the pertinent literature and little is known about their relevance for different kinds of entrepreneurship. Hence, in this paper we will test a variety of alternative measures, as it is crucial to know which kind of input market thickness – if any – does really matter for the setup of complex new ventures.

The most common indicators of input market thickness discussed and used in the empirical literature are indicators of population and labor market density. Bleakley and Lin (2007), for instance, measure labor-market density by the logarithm of population per square mile. A related measure used by Fu (2007) is number of workers per square kilometer. In this paper, we measure *population density* by the number of inhabitants per square kilometer and *employment density* (following Fu 2007) by the number of employees per square kilometer.

Simply focusing on measures of population or employment density might, however, be insufficient, as it is often argued that it is not the density of total employment but the density of highly skilled employment that really matters for knowledge spillovers conducive to entrepreneurship. We take this into account by considering *human capital density* (defined as share of highly qualified employees in total regional employment) as an additional explanatory variable. As formal qualification may be not particularly informative with respect to the knowledge-creating capacity of people we further consider *R&D density* (defined as share of R&D employees in total regional employment) as measure of input market thickness.

Finally, to avoid a bias from merely focusing on labor/human capital inputs we also consider physical capital input by looking at *investment density*, which we define as (manufacturing) investment per square kilometer over a four year period.

As can be seen from Table 4, correlations among our explanatory variables are generally low, with the exception of those among population, employment and investment densities. The high correlation between population, employment and investment density is not surprising as all three variables measure agglomeration economies in a rather broad and unspecific sense. Therefore, we will always specify models including only one of these three variables at a time, checking the robustness of our results across different model specifications.

[Insert Table 4 about here]

4. The Empirical Model

4.1 Data set and variables

Our regional-level data set was compiled from various sources, namely the Mannheim Enterprise panel (providing start-up data differentiated by region and industry), the employment statistics of the German Federal Employment Agency (providing regional-level employment data differentiated by various criteria such as nationality, educational level, industry, firm size etc.) and the Federal Statistical Office (providing data on the size of the region in square kilometers, population, disposable income, manufacturing investment and export share in manufacturing.) Data on the characteristics of newly founded firms were taken from the KfW/ZEW start-up panel and were kindly provided by ZEW Mannheim. The 97 German planning regions form the regional basis of the analysis, which encompasses start-ups of the years 1998–2001 and 2002–2005.

As explicated in Section 3.1 we distinguish different groups of start-ups, according to their (average) complexity, namely highly complex start-ups in a narrow sense, complex start-ups in a broader sense, and low complexity start-ups. We take into account that regions differ in size by analyzing start-up intensity (i.e. start-ups per 10.000 working-age inhabitants) rather than the absolute number of start-ups.

As argued in Section 3.2 input market thickness can take very different forms and it is important to know which specific kinds of local inputs really matter for entrepreneurial projects of different complexity. Therefore, we use different measures characterizing different input markets, namely population density, employment density, density of highly-qualified employees, R&D-density and investment density. The exact definitions of these measures have been given in Section 3.2.

Moreover, we include a number of control variables discussed in earlier work on the regional variation in new firm formation (e.g. Reynolds et al. 1994, Armington and Acs 2002, Audretsch et al. 2010), including the regional unemployment rate, the level and growth rate of income in the region, the level and growth rate of income in neighboring regions, the regional firm size structure (as indicated by the employment share of small companies with less than 20 employees) and the manufacturing share in total employment.

4.2 Econometric model and estimation strategy

Following our research goal, we perform several panel data estimations after Baltagi (2003). Specifically we build on the two-way error component model:

$$Y_{it} = \alpha + X'_{it}\beta + \gamma_i + \delta_t + u_{it} \tag{11}$$

where Y_{it} is the dependent variable, X_{it} is the vector of explanatory variables, γ_i and δ_t are region- and time-specific effects, u_{it} is a stochastic error with zero mean and given variance, α and β are coefficients to be estimated.

First we detect whether γ_i and δ_t can be best represented by either fixed constants - as in the fixed effects estimator - or random variables - as in the random effects estimator - by running a Hausman test. The null of this test is that the two bespoken estimators are close. In this case, the random effects estimator is preferable, as it will be both consistent and efficient. If the null cannot be accepted, the fixed effects estimator should be preferred, because it will still be consistent.

For the fixed effects estimator, it is possible to improve the model specification by dropping insignificant time- and region-specific effects. For the random effects estimator, one can resort to a number of different tests to understand whether either time or region specific effects can be constrained to zero, such as the Breusch and Pagan (1980), the Honda (1985), the King and Wu (1997), and the Standardized Lagrange Multiplier tests.

We implement this procedure for all our dependent variables, namely the intensity of highly complex start-ups, of low complexity start-ups and of all start-ups. We also check whether our results are stable across different definitions of "complexity" (narrow definition versus broad definition) and different models specifications (by including highly correlated regressors one at a time).

Once detected the most suitable model specification, we will also implement some further robustness checks. First we will check whether the inclusion of additional control variables or the dropping of variables changes our principal results. For example, we will introduce measures of ethnic/cultural diversity as additional controls or use as explanatory variables the log of income per head and its spatial lags instead of income growth and its spatial lags. In the second place, we will exclude regions with peculiar developments under the period of observation, namely Hamburg, Braunschweig, Westpfalz, Rheinpfalz, and Oberfranken. Third, we will consider East and West Germany separately and we will also offer a test for this division by running a poolability test (see Baltagi, 2003 and, for an application on Germany, Patuelli et al., 2010).

5. Results and Discussion

5.1 Main findings

Table 5 sets out our main findings regarding total start-up intensities. The Hausman test points to the random effects estimator as the one to be preferred. The Breusch and Pagan (1980), the Honda (1985), the King and Wu (1997), and the Standardized Lagrange Multiplier tests all accept the null that δ_t is equal to zero. The constant apart, significant coefficients are the manufacturing share in total employment which has a negative impact on start-up intensity and the population, investment and employment densities, which all have a positive impact on total start-up activity. According to the model specification, elasticities computed at mean values range from -0.16 to -0.20 for the first variable and from 0.04 to 0.05 for the last three. A similar pattern emerges once explaining low complexity start-up intensity, as can be seen from Table 6.

The bespoken results can offer a benchmark for those concerning highly complex start-ups, that are set out in Table 7. In this case, the Hausman test still prefers the random effects estimator to the fixed effects one. However, the two-way error component model finds support in the data. The unemployment rate has a negative and significant coefficient. However, human capital density, R&D density, the share of small enterprises, population, investment, and employment densities all have positive and significant coefficients.[10] Regarding elasticities computed at mean values and focusing only on the most robust variables, they range, across different model specifications and definitions for the dependent variable, from 0.25 to 0.42 for human capital density, from 0.34 to 0.54 for R&D density, and from 0.05 to 0.07 for population, investment and employment densities.

Hence, the results of the econometric analysis suggest that the locational determinants of highly complex start-up projects differ substantially from those of low complexity start-ups and total start-ups. While there appears to be little difference with respect to the impact of population/employment density and investment density (which have a positive but rather small impact on all kinds of start-ups), human capital density and R&D density have a strong positive impact on highly complex start-ups, but no significant impact on low complexity start-ups and total start-ups.

In sum, the Helsley/Strange – hypothesis that thicker input markets make more complex start-up projects feasible is clearly supported by the data. However, our results also indicate that the relationship might be more complicated than reflected by the current state of theoretical research: While general agglomeration economies (as measured by population, investment or employment density) appear to favor a broad range of start-ups, complex start-up projects (both, narrowly and broadly dedfined) have rather different locational requirements and benefit particularly from a high regional density of R&D and highly qualified employees.

[10] Also the manufacturing share in total employment has a positive and significant effect, but only adopting a broad definition of highly complex start-ups and not in the model including investment density.

5.2 Robustness checks

As can be seen from Table 7, our core results are very robust with respect to different definitions of start-up complexity, i.e. the results for complex start-ups in a broader sense resemble very much the results for highly complex start-ups in a narrow definition.

The log of available income and its spatial lags turn out to have a significant coefficient once inserted in our model in the place of variables on growth rate of income.[11] The log of available income has a positive coefficient, while its spatial lags have a negative one (Table 8). However our results do not change much once focusing on variables that have significant coefficients in Table 7.

As further robustness checks we estimate our models by resorting to a maximum likelihood approach and include an indicator of cultural diversity as an additional control variable following Audretsch et al (2010). First, changing our estimation method does not affect our results. Second, we find that cultural diversity has a positive impact on highly complex start-ups, whereas its impact on low complexity start-ups and total start-ups is either insignificant or negative. More importantly, in most model specifications the inclusion of cultural diversity does neither affect the sign nor the significance of the input market thickness measures of interest.[12]

Once peculiar regions are dropped, results slightly change compared to Table 7. Indeed, Table 9 shows that population, investment and employment density lose significance, but the unemployment rate, human capital density, R&D density and the share of small firms do not.

When exploring possible differences between East and West Germany, it turns out that poolability is not rejected for the model including investment density, but it is for the other models. In the first case, the F-test returns a p-value of 0.46, while in the last ones p-values of 0.02 for the model including population density and 0.01 for the model including employment density. In these last models, then, only West variables are significant. The magnitude and the sign of the coefficients are very similar to those in the pooled model.

There is also some evidence that the effect of density variables is stronger in the West compared to the East of Germany. This might be due to the fact that highly complex start-ups require a certain maturity of the economic system and enterpreneurial culture of a region and the East is still lagging behind the West in these respects.

6. Summary and Conclusions

Although the principal idea that there is a close interrelation between the characteristic features of regions and the characteristic features of new firms that emerge in these regions can be traced back to the classics of economic thinking, it is only recently that scholars of urban economics have suggested a theoretical link between the thickness of regional input markets and the complexity of feasible start-up projects. The current paper is to the best of our knowledge the first one that tries to test this link empirically.

In doing so, we make use of two unique and complementary data bases for Germany, namely the Mannheim Enterprise panel and the KfW / ZEW Start-up panel. We suggest a number of criteria for

[11] For the ease of exposition we focus the subsequent discussion on robustness checks for highly complex start-up intensities. The results of further tests are available from the authors upon request.

[12] We did not include cultural diversity in the basic model because the correlation of this variable with R&D density is relatively high.

measurement of start-up complexity which enables us to identify industries with highly complex start-ups and to distinguish them from industries with average- or low-complexity start-ups. These criteria include employment (in year of start-up and later), heterogeneity of founders' human capital, pursuit of highly sophisticated activities, innovativeness and international scope of newly founded firms. Moreover, as there are different forms of input market thickness and scholars of urban economics have argued that it is crucial to know which kinds of local inputs really matter for start-up projects of different complexity we consider five different indicators of input market thickness representing different inputs in our analysis.

We find that the locational determinants of highly complex start-up projects differ substantially from those of start-ups with low or average complexity. While there appears to be little difference with respect to the impact of population/employment density and investment density (which have a positive but rather small impact on a wide range of start-ups), human capital density and R&D density have a strong positive impact on complex start-ups (both, narrowly and broadly defined), but no impact on low complexity start-ups and total start-ups.

While our findings support the principal hypothesis that thicker input markets foster more complex start-ups, they also indicate that it might be useful to be more specific with respect to the input markets that really matter for highly complex projects. The theoretical model provided by Robert Helsley and Will Strange (2011) marks an important step towards a better understanding of the complex interrelation between urban structure and entrepreneurship. The current paper provides empirical support for the hypothesized link but also suggests that input market thickness per se might be a too general concept to allow practically relevant conclusions as some regional resources appear to be more conducive to highly complex start-up projects than others. A main insight from this study is that it is not so much general agglomeration economies (as reflected by population/employment density or investment density) but rather the thickness of the regional labor market for highly-qualified people and R&D workers that really matters for the feasibility of highly complex start-up projects. Hence, there is a need for theoretical models that allow for more than one regional input and are more specific about the importance and interplay of the different regional inputs that matter for new firm formation.

Apart from providing feedback to some recent developments in urban economics theory, our findings have implications for regional development and start-up promotion policies as well. First, they suggest that highly complex start-ups and low complexity start-ups have quite different locational needs. Hence, policy-makers eager to foster the set up and growth of new ventures in their region should take these differences into account. Second, not all regions dispose of the necessary thickness of relevant inputs to make the formation and successful post entry performance of highly complex start-ups feasible. Input market thickness might change over time, but this is a process that takes place very slowly and can hardly be pushed in the short run. Regions with low density of highly qualified labor and R&D are, therefore, unlikely to attract highly complex start-ups. Policy-makers in such regions might, of course, offer high subsidies to attract founders of highly complex new ventures, but it appears unlikely that this will pay off in the longer run.

References

Almus, M., Engel, D. and Prantl, S. (2000), The „Mannheim Foundation Panels“ of the Centre for European Economic Research *(ZEW)*, ZEW Documentation 00–02. Zentrum für Europäische Wirtschaftsforschung, Mannheim.

Armington, C. and Acs, Z.J. (2002), The Determinants of Regional Variation in New Firm Formation, *Regional Studies* 36: 33-45.

Audretsch, D. B. and Dohse D. (2007), Location: A Neglected Determinant of Firm Growth. *Review of World Economics* 143: 79-107.

Audretsch, D. B. and Dohse D. (2010), Cultural diversity and entrepreneurship: a regional analysis for Germany. Annals of Regional Science 45: 55–85.

Baltagi, B. H. (2003), *Econometric Analysis of Panel Data*, 2nd edition. Wiley: Chichester.

Breusch T. S. and Pagan, A.R. (1980), The Lagrange Mulitplier Test and Its Application to Model Specification in Econometrics. *Review of Economic Studies 47*: 239–253.

Chinitz, B.J. (1961), Contrasts in agglomeration: New York and Pittsburgh. *The American Economic Review* 51: 279–289.

Fryges, H., Gottschalk, S., Gude, H., Kohn, K., Müller, K., Niefert, M. and Ullrich, K. (2009), Gründungspanelreport, Fahrt aufnehmen in stürmischen Gewässern. Chancen und Herausforderungen in der Entwicklung junger Unternehmen. Creditreform, KfW, ZEW (eds.), Mannheim.

Fryges, H., Gottschalk, S. and Kohn, K. (2010), The KfW/ZEW Start-up Panel: Design and Research Potential. *Schmollers Jahrbuch* 130: 117 – 131.

Fu, S. (2007). Smart Café Cities: Testing human capital externalities in the Boston metropolitan area. *Journal of Urban Economics* 61: 86–111.

Glaeser, E.L. (2007), Entrepreneurship and the City. Cambridge: National Bureau of Economic Research. Working Paper 13551

Glaeser, E.L., S. Rosenthal and Strange, W. (2009), Urban Economics and Entrepreneurship. Cambridge: National Bureau of Economic Research. Working Paper 15536.

Gourieroux, C., Holly, A. and Monfort, A. (1982), Likelihood Ratio Test, Wald Test and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters. *Econometrica 50* :63–80.

Gottschalk, S., Gude, H., Kanzen, S., Kohn, K., Licht, G., Müller, K., Niefert, M., and Spengler, H. (2008), KfW /ZEW-Gründungspanel für Deutschland – Beschäftigung, Finanzierung und Markteintrittsstrategien junger Unternehmen. Creditreform, KfW, ZEW (eds.), Mannheim.

Grupp, H. and Legler, H. (2000), Hochtechnologie 2000 – Neudefinition der Hochtechnologie für die Berichterstattung zur technologischen Leistungsfähigkeit Deutschlands. Research Report to the Federal Ministry of Education and Research, Karlsruhe, Hannover.

Helsley, R.W. and Strange, W.C. (2011), Entrepreneurs and cities: Complexity, thickness and balance. *Regional Science and Urban Economics* 41: 550–559.

Honda, Y. (1985), Testing the Error Components Model with Non-normal Disturbances. *Review of Economic Studies 52*: 681–690.

King, M.L. and Wu, P.X. (1997), Locally Optimal One-sided Tests for Multiparameter Hypotheses. *Econometric Reviews 16* :131–156.

Marshall, A. (1920), Principles of Economics. MacMillan, London.

Metzger, G. and Heger, D. (2005), Die Bereitstellung von Standardauswertungen zum Gründungsgeschehen in Deutschland und Österreich für externe Datennutzer, Version 2005-02. Zentrum für Europäische Wirtschaftsforschung (ZEW), Mannheim.

Michelacci, C. and Silva, O. (2007), Why So Many Local Entrepreneurs? *Review of Economics and Statistics* 89: 615-633.

Patuelli, R., Vaona, A. and Grimpe, C. (2010), The German East-West Divide In Knowledge Production: An Application To Nanomaterial Patenting. *Tijdschrift voor Economische en Sociale Geografie* 101(5): 568-582.

Reynolds, P., Storey, D. and Westhead, P. (1994), Regional Variations in New Firm Formation – Special Issue. *Regional Studies 28*: 343-456.

Smith, A. (1776/2001), *An Inquiry into the Nature and Causes of the Wealth of Nations,* Bartleby, New York.

Vernon, R. (1960), Metropolis 1985. Harvard University Press, Cambridge, MA.

Table 1 - Composition of Industry Sectors Covered by the KfW/ZEW Start-up Panel

Industry	NACE (revision 1) code
Cutting-edge **T**echnology **M**anufacturing[13]: **CTM**	23.30, 24.20, 24.41, 24.61, 29.11, 29.60, 30.02, 31.62, 32.10, 32.20, 33.20, 33.30, 35.30
High-**T**echnology **M**anufacturing[14]: **HTM**	22.33, 24.11, 24.12 – 4, 24.17, 24.30, 24.42, 24.62 – 4, 24.66, 29.12 – 4, 29.31 – 2, 29.40, 29.52 – 6, 30.01, 31.10, 31.40, 31.50, 32.30, 33.10, 33.40, 34.10, 34.30, 35.20
Technology-**I**ntensive **S**ervices: **TIS**	64.2, 72 (without 72.2), 73.1, 74.2, 74.3
Software **S**upply and **C**onsultancy: **SSC**	72.2
Non-**H**igh-tech **M**anufacturing: **NHM**	15 – 37 (without those which are CTM/HTM)
Skill-**I**ntensive **Services**: **SIS**	73.2, 74.11 – 4, 74.4
Other **B**usiness-oriented **S**ervices: **OBS**	71.1, 71.2, 71.3, 74.5– 74.8 (without 74.84.7), 90, 64.1, 61, 62, 60.3, 63.1, 63.2, 63.4
Consumer-**O**riented **S**ervices: **COS**	55, 70, 71.4, 92, 93, 80.4, 65 – 67, 60.1, 60.2, 63.3
Con**st**ruction: **CST**	45
Wholesale and **R**etail **T**rade: **WRT**	50 – 52 (without 51.1)

Sources: Fryges et al. (2010); Grupp and Legler (2000)

[13] Manufacturing industries with average R&D expenditure > 8.5% of total sales.
[14] Manufacturing industries with average R&D expenditure 3.5 – 8.5% of total sales.

Table 2 - Indicators of start-up complexity in different industries

Indicator	Measure(s) applied
Size of start-up	Average employment size in year of start-up[15]
Development of firm size over time	Average employment in subsequent years
Heterogeneity of founders' human capital	— Share of team start ups — Average number of founders
Pursuit of highly sophisticated activities	Share of firms with own R&D
Innovativeness / International scope	Share of firms with products new to the market (i) Novelties to the national market (ii) Novelties to the international market

Table 3 - Industry groups, classified by their average start-up complexity

Industry groups with …	
(1a) highly complex start-ups in a narrow sense: CTM + HTM + TIS + SSC **(1b) complex start-ups in a broader sense:** CTM + HTM + TIS + SSC + NHM	**(2) low complexity start-ups** COS + WRT + CST

[15] Note that the KfW/ZEW start-up panel also provides information concerning average investment volume by industry. We have, however, abstained from using investment volume as measure of start-up size as this measure is strongly biased towards manufacturing industries; i.e. manufacturing industries usually have much higher investment volumes than service industries.

Table 4 - Correlation matrix

Variable	Indicators of Input Market Thickness					Control Variables					
	1	2	3	4	5	6	7	8	9	10	11
	Population density	Employment density	Investment density	Human capital density	R&D density	Income growth rate	Spatial lag of income growth rate	Unemploy-ment rate	Export share in manu-facturing	Manufacturing share in total employment	Share of small enterprises
1	1										
2	0.9874	1									
3	0.8672	0.8921	1								
4	0.3506	0.3671	0.3584	1							
5	0.3276	0.3359	0.4972	0.4327	1						
6	-0.2655	-0.2193	-0.1669	0.0964	-0.0001	1					
7	0.1254	0.1356	0.1466	-0.0767	0.1713	0.1836	1				
8	0.0561	0.0344	-0.0533	0.3816	-0.3666	-0.2565	-0.4991	1			
9	0.1513	0.1644	0.3244	0.0635	0.5960	0.0044	0.1730	-0.4731	1		
10	-0.0645	-0.0414	0.1496	-0.2767	0.4544	0.1055	0.3135	-0.6801	0.4453	1	
11	-0.4932	-0.5114	-0.6303	-0.4157	-0.7130	0.0688	-0.0831	0.1324	-0.3606	-0.4048	1

Table 5 - Regional determinants of total start-up intensity in German planning regions, 1998 – 2005
Estimation method: one-way random effects estimator

	Total start-up intensity					
	Coeff.	z-stat.	Coeff.	z-stat.	Coeff.	z-stat.
Income growth rate	-33.22	-0.49	-43.02	-0.65	-25.32	-0.38
Spatial lags in income growth rate	38.11	0.39	37.84	0.39	35.49	0.37
Unemployment rate	-52.90	-1.96	-53.85	-2.02	-48.03	-1.80
Human capital density	42.45	0.88	35.87	0.74	32.84	0.69
R&D density	165.05	1.78	158.64	1.73	178.63	1.95
Export share in manufacturing activities	-9.06	-1.09	-11.50	-1.39	-9.49	-1.17
Manufacturing share in total employment	<u>-81.55</u>	-2.11	<u>-94.15</u>	-2.53	<u>-75.81</u>	-2.01
Share of small enterprises	19.79	0.60	24.63	0.75	32.01	0.98
Population density	<u>0.01</u>	2.52				
Investment density			<u>0.00</u>	2.96		
Employment density					<u>0.01</u>	3.32
Constant	<u>40.07</u>	2.30	<u>41.11</u>	2.44	<u>34.72</u>	2.02
Specification tests						
Hausman test	15.85		11.68		15.89	
Breusch and Pagan test - region effect	37.11*		33.35*		37.47*	
Breusch and Pagan test - time effect	0.43		0.30		0.71	
Honda test - region effect	6.09*		5.77*		6.12*	
Honda test - time effect	-0.65		-0.54		-0.85	
King and Wu test - region effect	6.09*		5.77*		6.12*	
King and Wu test - time effect	-0.65		-0.54		-0.85	
Standardized Lagrange Multiplier test - region effect	6.40*		6.09*		6.43*	
Standardized Lagrange Multiplier test - time effect	-0.01		0.21		-0.40	

The constant was dropped because it was never significantly different from zero in pooled estimates.

*: test rejects the null at the 5% level Underlined coefficients are significantly different from zero at the 5% level.

Table 6 - Regional determinants of low complexity start-up intensity in German planning regions, 1998 - 2005

Estimation method: one-way random effects estimator

	Low complexity start-ups					
	Coeff.	**z-stat.**	**Coeff.**	**z-stat.**	**Coeff.**	**z-stat.**
Income growth rate	0.02	0.00	-6.75	-0.16	1.98	0.04
Spatial lags in income growth rate	64.67	1.03	65.79	1.05	64.54	1.03
Unemployment rate	-11.32	-0.65	-12.28	-0.71	-9.04	-0.52
Human capital density	-3.88	-0.12	-6.95	-0.22	-8.57	-0.28
R&D density	85.86	1.45	82.05	1.39	92.91	1.57
Export share in manufacturing activities	-4.09	-0.77	-5.48	-1.03	-4.29	-0.82
Manufacturing share in total employment	<u>-59.75</u>	-2.43	<u>-67.15</u>	-2.82	<u>-57.90</u>	-2.39
Share of small enterprises	17.65	0.84	19.65	0.94	22.57	1.07
Population density	<u>0.00</u>	2.20	-	-	-	-
Investment density	-	-	<u>0.00</u>	2.49	-	-
Employment density	-	-	-	-	<u>0.01</u>	2.67
Constant	<u>25.50</u>	2.30	<u>26.40</u>	2.44	<u>23.35</u>	2.11
Specification tests						
Hausman test	13.12		12.78		9.72	
Breusch and Pagan test - region effect	32.83*		34.20*		33.58*	
Breusch and Pagan test - time effect	0.00		0.08		0.02	
Honda test - region effect	5.72*		5.85*		5.80*	
Honda test - time effect	0.03		-0.29		-0.15	
King and Wu test - region effect	5.73*		5.85*		5.80*	
King and Wu test - time effect	0.03		-0.29		-0.15	
Standardized Lagrange Multiplier test - region effect	6.04*		6.16*		6.11*	
Standardized Lagrange Multiplier test - time effect	1.36		0.71		0.97	

*: test rejects the null at the 5% level

Underlined coefficients are significantly different from zero at the 5% level

Table 7 - Regional determinants of highly complex start-up intensity in German planning regions, 1998 – 2005 Estimation method: two-way random effects_estimator

	Highly complex start-ups - narrow definition						Highly complex start-ups - broad definition					
	Coeff.	z-stat.	Coeff.	z-stat.	Coeff.	z-stat.	Coeff.	z-stat.	Coeff.	z-stat.	Coeff.	z-stat.
Income growth rate	-12.40	-1.29	-15.66	-1.62	-11.65	-1.22	-18.75	-1.41	-23.22	-1.74	-17.80	-1.36
Spatial lags in income growth rate	-16.62	-1.13	-11.22	-0.76	-18.46	-1.27	-21.15	-1.05	-14.14	-0.70	-24.47	-1.23
Unemployment rate	-18.89	-5.65	-19.26	-5.68	-19.02	-5.88	-21.83	-4.93	-22.40	-4.99	-21.97	-5.13
Human capital density	26.85	3.86	28.16	4.03	25.54	3.77	31.65	3.44	33.39	3.62	29.81	3.32
R&D density	40.57	3.56	40.40	3.50	40.33	3.67	45.81	3.04	45.43	2.97	45.66	3.14
Export share in manufacturing activities	-0.53	-0.47	-0.72	-0.63	-0.62	-0.57	-0.29	-0.20	-0.55	-0.37	-0.42	-0.29
Manufacturing share in total employment	0.61	0.14	-1.40	-0.33	0.50	0.12	12.97	2.32	10.39	1.86	12.93	2.39
Share of small enterprises	9.63	4.90	10.09	4.84	10.16	5.28	17.40	6.61	18.03	6.46	18.11	7.04
Population density	0.00	3.57	-	-	-	-	0.00	3.48	-	-	-	-
Investment density	-	-	0.00	2.88	-	-	-	-	0.00	2.82	-	-
Employment density	-	-	-	-	0.00	4.53	-	-	-	-	0.00	4.44
Specification tests												
Hausman test	9.47		9.74		10.03		10.09		11.10		9.60	
Breusch and Pagan test - region effect	20.73*		18.50*		18.22*		16.10*		13.89*		14.01*	
Breusch and Pagan test - time effect	36.56*		49.10*		36.85*		45.71*		59.10*		45.90*	
Honda test - region effect	4.55*		4.30*		4.27*		4.01*		3.73*		3.74*	
Honda test - time effect	6.04*		7.01*		6.07*		6.76*		7.68*		6.77*	
King and Wu test - region effect	4.55*		4.30*		4.27*		4.01*		3.72*		3.74*	
King and Wu test - time effect	6.04*		7.01*		6.07*		6.76*		7.69*		6.77*	
Standardized Lagrange Multiplier test - region effect	4.88*		4.63*		4.60*		4.35*		4.06*		4.08*	
Standardized Lagrange Multiplier test - time effect	13.36*		15.18*		13.43*		14.80*		16.53*		14.84*	

The constant was dropped because it was never significantly different from zero
*: test rejects the null at the 5% level
Underlined coefficients are significantly different from zero at the 5% level

Table 8 - Regional determinants of highly complex start-up intensity in German planning regions, 1998 - 2005

Estimation method: two-way random effects estimator

	Highly complex start-ups - narrow definition						Highly complex start-ups - broad definition					
	Coeff.	z-stat.	Coeff.	z-stat.	Coeff.	z-stat.	Coeff.	z-stat.	Coeff.	z-stat.	Coeff.	z-stat.
Log of available income	0.49	2.07	0.60	2.66	0.40	1.77	0.68	2.16	0.81	2.72	0.57	1.86
Spatial lags in the log of available income	-0.58	-2.40	-0.64	-2.69	-0.53	-2.27	-0.71	-2.23	-0.79	-2.49	-0.65	-2.09
Unemployment rate	-15.47	-4.60	-16.26	-4.82	-15.28	-4.65	-17.96	-4.00	-18.91	-4.20	-17.64	-4.01
Human capital density	19.11	2.80	18.50	2.72	18.78	2.84	20.05	2.21	19.30	2.13	19.56	2.21
R&D density	47.17	3.71	44.94	3.53	48.42	3.90	50.78	2.99	48.13	2.83	52.52	3.16
Export share in manufacturing activities	-0.88	-0.79	-1.17	-1.05	-0.89	-0.82	-0.76	-0.51	-1.12	-0.75	-0.77	-0.53
Manufacturing share in total employment	1.81	0.38	-0.70	-0.15	2.26	0.49	12.63	1.98	9.62	1.56	13.32	2.16
Share of small enterprises	12.06	3.08	11.35	2.88	13.30	3.49	18.10	3.47	17.20	3.27	19.75	3.88
Population density	0.00	2.88	-	-	-	-	0.00	2.60	-	-	-	-
Investment density			0.00	2.35	-	-	-	-	0.00	2.11	-	-
Employment density	-	-	-	-	0.00	3.84	-	-	-	-	0.00	3.54

The constant was dropped because it was never significantly different from zero
Underlined coefficient are significantly different from zero at the 5% level

Table 9 - Regional determinants of highly complex start-up intensity once dropping peculiar regions, 1998 - 2005
Estimation method: two-way random effects estimator

	HC1						HC2					
	Coeff.	**z-stat.**	**Coeff.**	**z-stat.**	**Coeff.**	**z-stat.**	**Coeff.**	**z-stat.**	**Coeff.**	**z-stat.**	**Coeff.**	**z-stat.**
Income growth rate	-5.65585	-0.68214	-8.81177	-1.06912	-5.31157	-0.64381	-13.0449	-1.07394	-17.2794	-1.43803	-12.0962	-1.00146
Spatial lags in income growth rate	-23.4953	-1.97998	-19.6069	-1.65642	-23.869	-2.02028	-26.6986	-1.55155	-21.6202	-1.26441	-27.8338	-1.62572
Unemployment rate	-18.4302	-6.52766	-18.0073	-6.28596	-18.4848	-6.5885	-21.8816	-5.64415	-21.5047	-5.46676	-22.0142	-5.72028
Human capital density	26.4928	4.46	27.0851	4.53715	26.0412	4.39063	33.2369	4.07464	33.9644	4.15285	32.5006	3.99489
R&D density	39.9825	4.01384	41.8236	4.15619	40.1318	4.0679	42.3562	3.09101	44.2834	3.20305	42.405	3.12786
Export share in manufacturing activities	0.817695	0.848295	0.98198	0.99777	0.718746	0.744144	1.57954	1.1953	1.71307	1.27115	1.41048	1.06652
Manufacturing share in total employment	-0.84494	-0.23731	-1.57177	-0.44184	-0.92828	-0.26267	10.6992	2.19582	9.61794	1.97982	10.6362	2.20185
Share of small enterprises	9.35622	5.63348	8.90523	5.07041	9.53879	5.72623	17.1843	7.42049	16.8026	6.86101	17.5162	7.54831
Population density	0.000398	1.61049					0.000561	1.66359				
Investment density			5.41E-05	0.409185	-	-			0.000125	0.687468	-	-
Employment density	-	-	-	-	0.000962	1.85464	-	-	-	-	0.00143	2.02265

The constant was dropped because it was never significantly different from zero.
Underlined coefficient are significantly different from zero at the 5% level.

Dropped regions include Hamburg, Braunschweig, Westpfalz, Rheinpfalz, and Oberfranken.

Appendix A1

Derivation of expected maximum adaptation distance (following Helsley and Strange 2011: 553)

From the assumptions made in Section 2.1 it follows that d_i < 1/(2M), i.e. the adaptation distance d_i must be smaller than the midpoint of the arc between any two local inputs. For d < 1/2, there are two values of y on the unit circle satisfying d_i = d, such that the probability density of d_i equals 2 for 0 < d < 1/2 and 0 otherwise. The density of d_i = d, conditional on y ∈ (x-1/(2M), x+(1/2M)), is Pr{ d_i = d, y ∈ (x-1/(2M), x+(1/2M)}/Pr{y ∈ (x-1/(2M), x+(1/2M)}= 2/(1/M) = 2M. Hence, the density function of di is f(d) = 2M for 0 < d < 1/(2M) and 0 otherwise and the associated distribution function is F(d) = 2Md for 0 < d < 1/(2M) and 0 otherwise.

Obviously, the probability that the largest of N realizations of d_i takes on a value not larger than d is $F(d)^N$. This is the distribution of d_N, the largest order statistic of d_i. The density of d_N is g(d) ≡ N $F(d)^{N-1}$ f(d) = $N(2Md)^{N-1}2M$ = $2^N M^N N d^{N-1}$, for 0 < d < 1/(2M), and 0 otherwise.

The expected value of d_N is thus

$$\mathrm{E}[\mathrm{d_N}] = 2^{\mathrm{N}}\mathrm{M}^{\mathrm{N}}\mathrm{N}\int_0^{1/(2\mathrm{M})}\mathrm{z}^{\mathrm{N}}\mathrm{dz} = \frac{1}{2M}\frac{N}{N+1}.$$

Appendix A2

Classification of start-ups in different industries according to their complexity

We make use of the information provided by the KfW/ZEW start-up panel to assess average start-up complexity in 10 different industries. These industries, by which the KfW/ZEW start-up panel is stratified, include:[16]

- Cutting-edge Technology Manufacturing : CTM
- High-Technology Manufacturing : HTM
- Technology-Intensive Services: TIS
- Software Supply and Consultancy: SSC
- Non-High-tech Manufacturing: NHM
- Skill-Intensive Services: SIS
- Other Business-oriented Services: OBS
- Consumer-Oriented Services: COS
- Construction: CST
- Wholesale and Retail Trade: WRT

As argued in Section 3.1, start-ups are on average the more complex the larger they are (in terms of employment / people involved), the faster they grow, the more heterogeneous human capital (as indicated by the share of team start-ups or the average number of founders) they require, the more highly sophisticated activities (like R&D) they include and the more innovative outputs (products new to the market) they create. The large cross-sectional dimension of the KfW/ZEW start-up panel allows us to consider a whole bunch of complexity indicators such as average employment in year of start-up and in subsequent years (reflecting the average size of the start-ups in a given industry and its development over time), the share of firms with own R&D (reflecting highly sophisticated activities), the share of team-start-ups and the average number of founders (reflecting the input of heterogeneous human capital into the start-up project) and the share of start-up firms which provide market novelties to the international and national markets (reflecting the innovativeness, the international scope and, again, the sophistication of start-ups in a given industry).

The necessary data for the classification have been kindly provided by ZEW Mannheim. Table A1 displays the values for the start-up cohort 2008. The values for the other cohorts – as far as they are available – and their variation across industries are very similar to the 2008 cohort (i.e. rather stable over time) and are therefore not displayed here: They are, however, available from the authors upon request.

[16] See table 1 for details.

Table A1: Some structural characteristics of the start-up cohort 2008

Industries	Share of firms with own R&D	Share of team start-ups*	Share of firms with market novelties		Firm size (average employm.)		Average number of founders*
			Internat.	national	1st year	4th year	
CTM	0,491	0,337	0,074	0,064	2,6	4,6	1,5
HTM	0,550	0,430	0,083	0,080	4,1	6,8	1,7
TIS	0,261	0,329	0,034	0,056	2,2	4,8	1,5
SSC	0,471	0,441	0,077	0,128	2,3	5,3	1,6
NHM	0,240	0,315	0,021	0,058	3,2	5,4	1,4
SIS	0,183	0,400	0,044	0,072	1,8	2,4	1,5
OBS	0,089	0,286	0,014	0,043	2,1	7,7	1,4
COS	0,073	0,226	0,006	0,037	2,3	2,8	1,3
CST	0,038	0,195	0,000	0,020	2,2	4,2	1,3
WRT	0,127	0,237	0,028	0,053	2,0	3,4	1,3
All Industries	**0,115**	**0,260**	**0,017**	**0,045**	**2,2**	**4,0**	1,4

* cohorts 2008-2010

Source: ZEW Mannheim; own calculations

Table A1 gives a first impression of how different (on average) start-up projects in different industries are. Table A2 displays in a more systematic way which industries deviate significantly from the average of the whole sample. What is striking about table 2 is that for most industries there is little ambiguity, i.e. the various complexity indicators (as far as they are significant) point into the same direction.[17] Start-ups in cutting-edge manufacturing (CTM), for instance, are on average significantly above the all-industry mean w.r.t. to five criteria and insignificant w.r.t. to two criteria. Likewise, start-ups in customer oriented services (COS) are on average significantly below the all-industry mean w.r.t. to five criteria and insignificant w.r.t. to two criteria.

From table A2 it is quite obvious that start-ups in cutting-edge manufacturing (CTM), in high tech manufacturing (HTM), in technology intensive services (TIS) and software supply and consulting (SSC) stand out in terms of their average complexity. These four industries all have a significantly higher share of firms with own R&D, of team-start-ups, of firms with international market novelties and a significantly higher number of founders.[18] Start-ups in these four industries (CTM, HTM, TIS and SSC) are therefore grouped together and classified as *highly complex start-ups in a narrow sense.*

[17] The only exception is start-ups in skill-intensive services which are on average small and slow-growing, but have an-above average number of founders and share of tem start-ups.
[18] Note that start-ups in CTM, HTM and SSC are also on average larger or grow faster than start-ups in other industries.

Table A2: Significant differences from the all-industry mean

Industries	Share of firms with own R&D	Share of team start-ups*	Share of firms with market novelties		Firm size (average employm.)		Average number of founders*
			Internat.	national	1st year	4th year	
CTM	+	+	+		+		+
HTM	+	+	+		+	+	+
TIS	+	+	+				+
SSC	+	+	+	+		+	+
NHM	+	+			+		
SIS		+			–	–	+
OBS						+	
COS	–	–	–			–	–
CST	–	–	–	–			–
WRT		–			–	–	–

* cohorts 2008-2010
+ Industry mean significantly (α=0,05) above mean of the whole sample
– Industry mean significantly (α=0,05) below mean of the whole sample
Source: ZEW Mannheim; own calculations

Non-high tech manufacturing (NHM) is clearly above average in terms of average start-up size (1st year), the share of team start-ups and the share of firms which perform own R&D, but figures less prominent w.r.t. all other criteria. We take this into account by defining a second (broader) group of industries with complex start-ups, which we call *complex start-ups in a broader sense.* This group contains the group of highly complex start-ups in a narrow sense as a sub-group, plus start-ups in non-high tech manufacturing. The broader definition (including NHM) will primarily be used to check the robustness of results yielded w.r.t. the narrow definition of complexity.

On the other side of the spectrum, start-ups in customer oriented services (COS), in wholesale and retail trade (WRT) and in construction (CST) appear to be the ones with the lowest average start-up complexity (Table A2). Hence, start-ups in these three industry groups are classified *as low-complexity start-ups*, and it will be tested whether highly complex start-ups and low-complexity start-ups are differently affected by input market thickness.

Table A3 summarizes the classification of industry groups by their average start-up complexity that will be used in the empirical part of the paper.

Table A3: Industry groups, classified by their average start-up complexity

Industry groups with …	
(1a) highly complex start-ups in a narrow sense:	**(2) low complexity start-ups**
CTM + HTM + TIS + SSC	COS + WRT + CST
(1b) complex start-ups in a broader sense:	
CTM + HTM + TIS + SSC + NHM	