

Bauernschuster, Stefan; Falck, Oliver; Heblich, Stephan

Working Paper

Occupational choice and social contacts across regions

Jena Economic Research Papers, No. 2008,079

Provided in Cooperation with:

Max Planck Institute of Economics

Suggested Citation: Bauernschuster, Stefan; Falck, Oliver; Heblich, Stephan (2008) : Occupational choice and social contacts across regions, Jena Economic Research Papers, No. 2008,079, Friedrich Schiller University Jena and Max Planck Institute of Economics, Jena

This Version is available at:

<https://hdl.handle.net/10419/31741>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



JENA ECONOMIC RESEARCH PAPERS



2008 – 079

Occupational Choice and Social Contacts Across Regions

by

**Stefan Bauernschuster
Oliver Falck
Stephan Heblich**

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Max Planck Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

© by the author.

Occupational Choice and Social Contacts Across Regions

October 2008

Stefan Bauernschuster
University of Jena

Oliver Falck
Ifo Institute for Economic Research

Stephan Heblich
Max Planck Institute of Economics

Abstract

This paper tests the importance of social contacts on entrepreneurship. To measure differences in the interconnectedness of social contacts, we compare rural and agglomerated areas. A smaller community size in rural areas generates greater network closure. Agents' neighborhoods are more likely to overlap, which intensifies social contacts and thus facilitates resource mobilization. Analyzing the impact of social contacts across regions, we find that greater network closure increases the likelihood of being an entrepreneur by 1.9 to 14.2 percentage points, depending on the number of underlying social contacts. These results remain robust after applying matching techniques and concentrating on young entrepreneurs.

JEL-classification: J 24; Z13

Keywords: Occupational Choice, Entrepreneurship, Social Contacts

Acknowledgement: We are indebted to David Audretsch, Thomas Astebro, Olav Soerenson, Marc Piopiunik, Guido Schwerdt, and Diemo Urbig for helpful comments on an earlier working version of this paper.

Contact: Stefan Bauernschuster, University of Jena, Graduate College "The Economics of Innovative Change", Carl-Zeiss-Str. 3, D-07743 Jena (Germany), Phone: +49 3641 9 43270, Email: stefan.bauernschuster@uni-jena.de.

1. Introduction

Economic insight into transactions among agents is becoming increasingly influenced by network considerations originating from sociology.¹ For example, it is now widely understood that social networks link individuals and that they are thus reciprocally granted access to each others' resources. Initially, these resources comprise information about past interactions and future developments, which eventually have an impact on an individual's material investment today in the hope of cooperative payoffs in the future. The individual's benefit from social contacts in the form of an increased scale of exchange varies with the community's size that eventually determines the network's closure (Coleman 1988; Mobius and Szeidl 2007). Networks have a high closure if neighborhoods of connected agents are likely to overlap, which is more likely in smaller communities with less potential contacts (Allcott *et al.* 2007).

In this paper, we proxy social contacts by an individual's club memberships (cf. Putnam 2000) and focus on the variation in the information function of social contacts stemming from differences in their interconnectedness. We argue that an individual's social contacts influence his or her decision about becoming and staying an entrepreneur by helping overcome potential resource constraints. The underlying mechanism that supports cooperation is determined by the network's closure. As highly interconnected social contacts make the future outcome more predictable by publishing past actions, they provide better access to resources. By contrast, less interconnected social networks require formal institutions in certain situations (e.g., a contract for supplies or a loan), rather than simply trust between parties who know each other well and also know each others' connections, which reinforces their own knowledge of each other (Knack and Keefer 1997). Assuming that social contacts become more interconnected with frequent interaction and regular contact (Granovetter 1985), it seems plausible to assume that rural areas are characterized by more interconnected social contacts than are agglomerated areas. In rural areas where the population is sparse, there are fewer arenas in which people can meet and interact, i.e., supermarkets, recreation areas, and, especially, clubs, and so it is more likely that network neighbors of two individual overlap, thereby leading to greater network closure (Allcott *et al.* 2007). Eventually, this situation will result in the formation of more interconnected social contacts and trust, thus facilitating

¹ Recent publications include, e.g., research on informal hiring networks (Bayer *et al.* 2008), entrepreneurial location decisions (Michelacci and Silva 2007), worker's occupational choice and productive advantage

resource mobilization in rural areas. By contrast, in agglomerated areas where social networks are composed of less interconnected social contacts, formal institutions take over the role trust plays in the rural areas. Based on this line of reasoning, we assume that social contacts have an impact on occupational choice and that this impact should vary with a region's network closure.

Our empirical identification strategy draws on the differing impact of social contacts on becoming and staying an entrepreneur *across* regions—rural and agglomerated areas. Thereby, the region-specific impact of social contacts is estimated from exploiting the variation between individuals *within* regions. Unobserved heterogeneity between individuals that influences both the likelihood of being a social network member as well as the likelihood of being an entrepreneur may cause endogeneity problems (Durlauf 2002). However, if we assume that the self selection process is the same in both rural and agglomerated areas, unobserved heterogeneity cancels out when only looking at differences in the impact of social contacts between regions. Our results show that social contacts have a significantly higher impact in rural areas, i.e. more interconnected social contacts increase the likelihood of being an entrepreneur. We find that intensifying social contacts increases the likelihood of being an entrepreneur by 1.9 to 14.2 percentage points, depending on the number of underlying social contacts. These results remain robust when applying various approaches, including matching techniques. In the matched sample, individuals from rural and agglomerated areas do, on average, not differ in their number of club membership. These results also remain robust when concentrating on young entrepreneurs who have started their business within the last twelve months preceding the survey.

The remainder of the paper is organized as follows. Section 2 develops our assumption of differences between agglomerations and periphery in more detail. Section 3 introduces our method; Section 4 our data. Then, in Section 5 we present our results and the robustness checks that show them to be reliable. Section 6 concludes with some suggestions for further research.

2. Regional Variation of Occupational Choice

Occupational Choice

The occupational choice literature defines the entrepreneur as an individual who starts and runs his or her own business (Kihlstrom and Laffont 1979). In the simplest form of this model, potential entrepreneurs choose between starting an own business, thereby earning an uncertain profit, and being dependently employed and earning a given wage. Assuming that the individual possesses the resources necessary to start up and given his or her individual risk aversion, the potential entrepreneur will start and run a business as long as the expected discounted future profits are larger than the discounted sum of future earnings from dependent employment.² An entrepreneur will continue running a business as long as this condition holds.

The expected entrepreneurial income is determined by a set of personal characteristics that either enhance or detract from the individual's productivity. Lazear (2005) finds that a balanced portfolio of skills is most conducive to becoming an entrepreneur. Individuals who have acquired outstanding skills in one particular field tend to specialize in that particular field, reaping the benefits of the division of labor. By contrast, entrepreneurs are frequently jacks-of-all-trades who have developed a basic understanding of several different fields, thus giving them the skills to manage the entire Schumpeterian innovation process from the birth of a new idea (invention) to its introduction on the market (innovation). During a person's lifecycle, his or her skill set most likely changes. For example, while formal qualifications from secondary education can degrade with age (due, in part, to new technology that renders previously acquired skills obsolete), knowledge gained from practical experience increases with age. Personal status, financial obligations, and risk attitude also change with age. The findings of several empirical studies suggest that there is a positive—usually an inverse u-shaped—relationship between an individual's age and the individual's decision to start a business (Parker 2004, p. 106).³

² The occupational choice as described here considers wage differences between industries and regions, e.g. due to higher (lower) competition for qualified workers in agglomerated (rural) areas,

³ The results of most empirical studies based on individual data suggest an inverse u-shaped effect of age on the probability of becoming self-employed (e.g., Evans and Leighton 1989b; Blanchflower and Meyer 1994; Blanchflower 2000). However, there are a few studies (Blau 1987; Evans and Jovanovic 1989; Evans and Leighton 1989a) that do not find any significant effect of age on self-employment.

Occupational Choice and Social Contacts

In addition to personal characteristics, social characteristics such as an individual's charisma or status that contribute to his or her prestige, also play an important role in occupational choice. This type of characteristic determines an individual's access to (and success in using) the embedded resources of a social network that can have a positive impact on his or her economic performance (Glaeser *et al.* 2002). Sanders and Nee (1996) point out three mechanisms by which social contacts generate resources that are especially supportive of entrepreneurship. First, they provide instrumental support by facilitating access to resources such as capital or labor that directly influence performance. Second, they give access to productive information about opportunities or risks, which can have an indirect influence on the entrepreneur's performance. Being socially embedded decreases transaction costs and reduces the uncertainty surrounding future developments (Granovetter 1985; Hayek 1937). Sanders and Nee's third mechanism, in contrast to being of an economic nature like the first two, consists of psychological aid. Social contacts can alleviate emotional stress that might affect the entrepreneur's success and help sustain what Schumpeter rather romantically describes as "the will to conquer," "the dream and the will to found a private kingdom," and "the joy of creating, of getting things done" (1912, p. 93). In fact, psychological well-being is regarded as an important driver of individual productivity (Darity and Goldsmith 1996).

The decision on whether to start or continue a business implies the preexistence of productive information and appropriate motivation, but a resource-constrained individual encounters hurdles to success. In the absence of informative past experience, potential cooperation partners can evaluate the entrepreneur only conditionally, thereby causing a suboptimal supply of resources. However, through repeated interaction and successful experiences, individuals learn about their partners' trustworthiness, which then increases the scale of exchange (Chwe 1999; Coleman 1988; Sobel 1985; Watson 1999). Trust developed from frequent interaction facilitates more accurate evaluation of the entrepreneur's success chances and thus increases the entrepreneur's possibilities for mobilizing resources. Accordingly, trust is a way to overcome potential information asymmetries regarding whether the entrepreneur is purposely exaggerating his or her prospects so as, for example, to obtain financing (Amit *et al.* 1990) and to detect cases where the entrepreneur is simply being too optimistic of success (Camerer and Lovo 1999; Wu and Knott 2006).

Network Structure Across Regions

Whether social interactions produce a high quantity of information that facilitates opportunity recognition or, instead, high quality information leading to trust depends on the community size and thus the network structure (Mobius and Szeidl 2007). To clarify this, let us consider some differences that distinguish rural areas from agglomerated areas. A smaller community size in rural areas basically creates an environment where everyone knows everyone else. Given that people are more likely to interact with those who are around them (Jackson and Roger 2006; Tabellini 2008) a smaller community means that agents' neighborhoods are more likely to overlap, thereby increasing network closure. In this context, clubs are a well-known arena for social interaction (Glaeser *et al.* 2002; Knack and Keefer 1997; Putnam 2000), and it seems safe to assume that rural areas will usually have only a few of these, perhaps, for example, one sport club, several church communities, and a music society. Accordingly, individuals are more likely to be linked via one or more of these groups, which will eventually contribute to the formation of more interconnected social contacts and trust (Allcott *et al.* 2007; Ellickson 1991; Mobius and Szeidl 2007).⁴ This tendency is enhanced in rural areas by the fact that there is generally less competition for people's recreational time, that is, there are not very many bars, clubs, restaurants, or theaters (cf. Glaeser *et al.* 2001) to frequent instead of joining a club or church group.

In contrast to rural areas, people living in urban areas are less likely to develop more closed networks. There are, of course, overlapping social contacts in cities, but they are not likely to cover the entire city-community; rather, they involve smaller parts of the city, for example, a neighborhood or a workplace (Jacobs 1961).⁵ These social networks will have weak links to other networks, for example, a person's club may include one or two of his business contacts, thus loosely linking the two networks (cf. Burt 1992; Putnam 2000). Thus the city-community is somewhat heterogeneous and there is a great number of sparsely interconnected social contacts that act as gateways into social networks. On the one hand, this type of social network system leads to less trust within the city-community because no one *really* knows anyone else very well (with a few exceptions, of course). However, on the other hand, it is

⁴ "Experiments, with both children and adults, show that subjects are more likely to give rewards to those assigned the same label than to others, even when the assignment is random, recipients are anonymous, and there is no impact on own payoffs." (Akerlof and Kranton (2002).

⁵ Jacobs (1961) also assumes that people living in agglomerated areas have less friends such that social networks in agglomerations exhibit a low network closure per se. However, Mobius and Szeidl (2007, p.19) find that, "even when controlling for the number of friends, large communities have less dense social networks".

exactly the existence of this great number of sparsely interconnected social contacts, i.e., connecting factors, that facilitates knowledge flows and obtaining information.

Occupational Choice Across Regions

The individual's occupational choice is greatly influenced by a combination of personal characteristics along with the surrounding social network. As social contacts develop gradually from personal interaction they tend to be regionally bound. Eventually, this leads Feldman (2001) to the conclusion that entrepreneurship is a regional event. Considering further that network structures vary with community size, it seems plausible to expect that there will also be a connection between the network structure and its impact on occupational choice (Michelacci and Silva 2007). In rural areas, social contacts overlap more frequently such that individuals have better access to resources. There is a clear implication that follows from this assumption. The entrepreneur's success depends on a set of factors that are observable only to a certain extent; however, because of the dense networks characteristic of rural areas, the entrepreneur's success (or lack thereof) will be more predictable because his or her history of previous action (or inaction) is fairly well known and bad behavior more easily punished or, better yet, avoided ahead of time (Bowles and Gintis 2002; Fehr and Gächtner 2000). Thus, trust facilitates the process of resource mobilization in less densely populated rural areas.

However, none of this means that rural areas are superior to urban areas as regards resources, only that the way of accessing them might be different. In an environment of frequently overlapping social contacts, trust serves as collateral. In the absence of trust, i.e., readily known information of an individual's past experience, a different type of information is required and this generally involves transaction costs (Dixit 2004; Williamson 1979; Zak and Knack 2001; North 1990). One common way to reduce these transaction costs is to use intermediaries, for example, venture capital firms, that specialize in condensing and evaluating a variety of information. Thus, in congested areas where social networks are composed of sparsely interconnected social contacts, formal institutions assume the role played by trust in more rural areas.

Following this line of reasoning, we assume that more interconnected social contacts have an impact on occupational choice and that this impact should vary with a region's social network structure, i.e., between urban and rural areas.

3. Method

When estimating the effect of social contacts on an individual's occupational choice, we are primarily concerned with omitted variables bias. If social contacts are measured through memberships in clubs and associations, it might well be that there is unobserved heterogeneity between individuals that influences both the likelihood of being a member of an association or club and the likelihood of being an entrepreneur. Thus, these omitted variables cause correlations between the error term and the independent social capital variable, which biases the estimates. To illustrate this point, imagine there are some people who are outgoing, energetic, active, and adventuresome. These people might more often join clubs than do others; however, they might also be more likely to be an entrepreneur due to the very same character traits. If we do not control for these traits in a multivariate analysis, we might mistakenly attribute the fact that someone is an entrepreneur to her membership in clubs (i.e. social contacts) although, in reality, it is not social contacts that accounts for differences in occupational choice but unobserved individual characteristics. This means that the effect of social contacts on a person's propensity to be an entrepreneur might be overestimated. Due to this kind of endogeneity, a positive coefficient of the associational activity variable cannot be interpreted as a causal effect of social contacts on the propensity to be an entrepreneur.

The problem of omitted variables could obviously be reduced by introducing into the model a wide range of control variables for relevant personal characteristics. However, due to data restrictions and since there is no clear theory on the determinants of social contact formation, the problem cannot be completely resolved in this way in practice (Durlauf 2002). Heckman *et al.* (1998) present a feasible way to overcome selection biases by using difference-in-differences estimators. Inspired by this approach, we draw on the differing impact of social contacts on being an entrepreneur *across* regions—rural and agglomerated areas. Thereby, the region-specific impact of social contacts is estimated by exploiting the variation between individuals *within* regions. Our identification strategy is explained in more detail below.

As discussed in Section 2, we expect differences in the coefficients of social contacts across regions on a person's propensity to be an entrepreneur. These differences are believed to occur due to the following conjectures. If one assumes that being a member of many associations in a rural area results in more interconnected social contacts, but that the same pattern of membership behavior in cities creates a wide network of sparsely interconnected social contacts, we can identify the importance of network closure in being an entrepreneur. Thus, in rural areas, individuals have better access to resources as the dense social networks in

those areas make the entrepreneur's future outcome more predictable due to knowledge of his or her history and because of a greater possibility of punishing or preventing noncooperative action. In short, trust facilitates the process of resource mobilization in less densely populated rural areas. By contrast, in congested areas where social networks are composed of less interconnected social contacts formal institutions adopt this role. Along this line, we assume that greater network closure has an impact on occupational choice and that this impact should vary with a region's network structure.

Our goal is to estimate the impact of social contacts based on differing underlying network structures on the propensity to be an entrepreneur. The dependent variable, *occupational status*, has two outcomes: being an employee or being an entrepreneur. Our main independent variables of interest are interaction terms of social contact variables with regional type dummies. Comparing the coefficients of these interaction terms provides insight into the way social contact effects differ solely due to regional heterogeneity. We argue that this regional heterogeneity can be explained by community size and, eventually, the interconnectedness of social contacts. In this quasi-difference-in-differences approach, unobserved individual heterogeneity cancels out if one assumes that the self-selection process is the same across regional types. To put it in different terms, we assume that unobserved characteristics work in the same way in rural and agglomerated areas. If this assumption holds, statistically significant differences in the coefficients of our interaction terms signify differences in regional-specific social contacts alone; unobserved individual heterogeneity cannot be responsible for any differences across regional types. The Global Entrepreneurship Monitor's (GEM) findings for Germany provide some initial support for the validity of our assumption. Individuals who state that they would prefer to be entrepreneurs, or "latent entrepreneurs" according to GEM, are assumed to be a good proxy for entrepreneurial spirit and they are relatively evenly distributed across Germany (Brixey *et al.* 2007). This situation leads to a sufficiently high rate of business startup rates across all types of German regions. This observation may be the result of a more evenly distributed settlement system across German regions in comparison to, for example, the United States, where we observe considerable segregation, e.g., rich homogenous suburbs and poor city centers (Schelling 1971). However, we are not concerned here with whether either this or that system of settlement is better; we simply note that the observed settlement structure across German regions supports the validity of our identification strategy.

Our modified difference-in-differences approach can be described by the following equations:

$$OCC_{rur} = \alpha_{rur} * SC + \beta * Z + \chi * UV \quad (1)$$

$$OCC_{agg} = \alpha_{agg} * SC + \beta * Z + \chi * UV \quad (2)$$

$$OCC_{rur} - OCC_{agg} = (\alpha_{rur} - \alpha_{agg}) * SC, \quad (3)$$

where OCC_{rur} stands for occupational status of people from rural areas and captures the average probability of being an entrepreneur in rural areas. OCC_{agg} is the respective variable for agglomerations. SC stands for the average number of social contacts. Z is a vector of the means of observable control variables in rural and agglomerated areas respectively. For simplicity, we assume that SC and Z are equal in rural and agglomerated areas, an assumption that is relaxed in the empirical assessment. Indeed, Brueckner and Largey (2008) show that low-density living does not reduce the number of social contacts. UV is a vector including the means of all unobserved variables that have an impact on a person's occupational choice. As can be seen from Equations (1) and (2), we assume that the betas and gammas are the same in rural and agglomerated areas. Assuming the gammas to be the same across regions means that the self-selection process does not differ systematically across regions. The advantage of this identification strategy is that unobserved heterogeneity (UV), disappears in Equation (3). Following these formulas, one can see that we are able to explain the difference in the average propensity to be an entrepreneur across regions solely by differences in the coefficients of the observed social contacts variables. The differences in the coefficients of social contacts can only be explained by regional differences in the interconnectedness of social contacts. Thus, this difference can be interpreted as the effect of intensifying social contacts, holding the number of social contacts (as well as all other covariates) constant.

We thus have overcome potential endogeneity problems. Equation (4) sums up our estimation strategy. A person's occupational choice OCC , i.e., his or her decision to become and stay an entrepreneur, depends on social contacts SC . However, we believe that the interconnectedness of *social contacts* SC is different in rural and agglomerated areas. Given the difference in the interconnectedness of social contacts across regions, we assume that the same memberships have greater value in rural areas than in agglomerated areas. To disentangle these effects, we implement interaction terms of social contacts SC with 0-1 dummies for rural areas R_{rur} and agglomerated areas R_{agg} . To capture differences in personal, regional, and industry characteristics that might influence the propensity to be an entrepreneur, we introduce several control variables at the individual, regional, and industry levels. Particularly, we implement region type, federal state, industry, and year fixed effects. With regard to the theoretical

discussion in Section 2, the region type fixed effects should be particularly well suited to capture differences in regional (formal) institutions. e is the error term.

$$\text{OCC:I} \left(1 \mid \text{SC} * \text{R}_{\text{rur}}, \text{SC} * \text{R}_{\text{agg}}, \text{R}_{\text{rur}}, \text{R}_{\text{agg}}, \text{individual characteristics, federal state, industry, year, } e \right) \quad (4)$$

4. Data

The ALLBUS survey (Allgemeine Bevoelkerungsumfrage der Sozialwissenschaften) is a valuable data source for our research question. It can be viewed as the German equivalent to the U.S. General Social Survey (GSS) and currently covers the period from 1980 to 2006. The ALLBUS program was financially supported by the German Research Foundation (DFG) from 1980 to 1986 and in 1991. Further surveys were financed on a national and federal state (Laender) level via the GESIS network (Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen). The dataset is based on regularly repeated, representative surveys of the German population conducted through personal interviews. ALLBUS covers a wide range of topics pivotal to empirical research in social sciences. A core set of questions is asked in every wave of the survey, with various sets of additional questions added in different years. Terwey et al. (2007) provide detailed information on the ALLBUS surveys in general and present all variables available in the cumulated dataset from 1980 till 2006.

We focus on the period from 1980 to 1992 because of the availability of information on club activities and industries. For these years, we have information on 24,754 individuals. From 1991 onward, ALLBUS includes individuals from Eastern Germany. However, due to a lack of comparability regarding private club memberships, we drop all 2,692 observations for Eastern Germany for 1991 and 1992. Furthermore, we retain only employed or entrepreneurs, leaving us with 10,010 observations. As to the concept of “entrepreneurship”, ALLBUS covers two different kinds of activities. First, it includes self-employed individuals who work for themselves and gain income by operating all tasks personally. And second, it also contains entrepreneurial individuals who are business owners, i.e. who run their own businesses *and* employ other persons. Due to data limitations, it is not possible to distinguish between these two kinds of individuals; therefore, both are regarded as being entrepreneurs in our context.

Additionally, we drop 3,034 observations from entities inappropriate to our purpose. These include industries such as agriculture, nonprofit organizations, private households, local authorities, and the social security system, but also individuals from the energy and water supply industry as well as from the mining industry where no entrepreneurs are found in our

sample. Individuals working for the German postal service (Bundespost) and for the German railway (Bundesbahn) were dropped as well, since these were state-owned enterprises during the period our research covers. These zero event industries differ from the other industries in our sample in various ways, e.g. barriers to entry or the employment status of its employees. For instance, telecommunication (formerly part of the German postal service) was a state monopoly until 1998 and, even today, more than 40 % of Deutsche Telekom's employees have the status of civil servants (cf. Czernich *et al.* 2008).

Information on club membership is missing for 898 individuals. After checking for the randomness of the missing values, we drop these observations. Our final sample contains 6,078 individuals, a sound database for our empirical analysis.

As a dependent variable we use the binary variable *occupational status*, which has the value 0 if the individual is employed and 1 if he or she is an entrepreneur. It might be argued that our identification strategy potentially suffers from reverse causality. However, occupational choice theory models the decision to become an entrepreneur not as a once and for all decision but rather as a decision that is made repeatedly every period (Kihlstrom and Laffont 1979). This means that social contacts that are developed during times of running a business can have an impact on the propensity of staying an entrepreneur, i.e., the decision to be an entrepreneur in the next period. Therefore, we argue that it does not matter at which point in time social contacts are developed. They always influence an individual's occupational choice—either the decision to become an entrepreneur for the very first time or the repeated decisions to stay an entrepreneur. However, in the robustness checks, we will also run separate regressions for young entrepreneurs who have started their own business during the last twelve months preceding the survey. If some of these individuals are members of clubs and, making the plausible assumption that the point of time they joined these clubs lies before their recently made occupational choice, this should additionally counteract reverse causality.

To capture the impact of social contacts, we use a variable that counts the number of an individual's memberships in private associations and clubs. ALLBUS contains information on an individual's membership in a political party, a religious association, a choir, a sports club, any type of hobby club, a citizens' group or initiative, any type of social club, a charity, a displaced persons' group, or a youth club. It is important to stress that we deal only with those associations and clubs that can be described as involving private leisure activities.

Memberships in employer, entrepreneur, or trade associations, as well as union memberships, are not included in our associational activity variable in order to avoid severe problems of reverse causality. Clearly, it is safe to assume that people are more likely to join an employers' association, entrepreneurs' association, or trade association *after* they have become an entrepreneur themselves. Moreover, if rate of unionization is smaller than the rate of entrepreneurs being members in employers' association, entrepreneurs' association, or trade associations, we might find a positive correlation between being an entrepreneur and the number of associational activities. However, this correlation coefficient can not properly be interpreted as evidence for positive impacts of associational activities on the propensity to become an entrepreneur. Thus, excluding occupation-related club memberships seems essential.

To identify the different effects of social contacts across region types, we interact the club membership variable with a binary variable taking the value of 1 for rural areas and a binary variable taking the value of 1 for agglomerated areas, respectively. We are well aware that measuring community size simply according to administrative borders can be misleading. A small community adjacent to a big city differs in many respects from a community with the same number of inhabitants that is located somewhere in the periphery. To overcome these problems, Boustedt (1975) developed community type variables for Germany that describe regions in socioeconomic terms, largely independent of their administrative regional borders. The original Boustedt variable groups regions into seven categories according to the number of inhabitants of the specified socioeconomic region. We aggregate the available data by defining a regional variable that comprises only two categories: one category is comprised of individuals living in rural areas with less than 5,000 inhabitants, the other is comprised of individuals living in more densely populated agglomerations with 5,000 or more inhabitants. To avoid being completely arbitrary in this matter, we tried several alternative classifications, which are presented in the robustness checks. However, after looking at these alternatives, it becomes obvious that our first categorization method is the clearest and most consistent given the goal of our research.

We use various control variables to take into account differences in individual, industrial, and regional characteristics, as well as time that might influence a person's occupational choice. Since research has shown pronounced differences between men and women as to the likelihood of becoming an entrepreneur (e.g., Blanchflower 2000), we implement a gender dummy: 0 for men, 1 for women. To control for regional embeddedness, we introduce a

dummy variable indicating whether a person is a tenant or owns his or her own house or flat. According to DiPasquale and Glaeser (1999), home ownership decreases mobility, which, in turn, provides an incentive to interact with the local community and thus create more contacts with neighbors. Another dummy variable is included to capture the effects of marital status on an individual's propensity to be an entrepreneur. It takes the value 1 if the interviewed person is married and lives with his or her spouse, 0 otherwise. We suspect that singles are more likely to be an entrepreneur, since any risk involved in such an endeavor would be their alone, that is, they are not responsible for the safety, financial or otherwise, of a spouse or children. Or, in other words, one could say that being married shows a time allocation preference for family. Apart from marital or family status, we control for a person's nationality by including a binary variable with the value 1 for non-German, 0 otherwise. To capture human capital effects, we include a person's education and working experience. Information on the respondents' secondary (and higher) education is available for the entire time period analyzed. We generate a categorical variable signifying whether an individual has no secondary school education, lower ("Hauptschulabschluss"), medium ("Mittlere Reife"), upper ("Hoch-/Fachhochschulreife") secondary education or whether she holds a degree of a university or a university of applied sciences ("Fachhochschule"). As a proxy for working experience, we compute a variable where we subtract the time before school enrollment and the number of years spent in school from an individual's age. In Germany, children are enrolled at the age of six. There are nine years of compulsory schooling; medium secondary education is meant to take ten years, while upper secondary education comprises thirteen years of schooling altogether. For individuals that hold a university degree we added six more years of higher education. Additionally, we introduce a variable that groups people into one of three categories according to their previous time of unemployment during the last 10 years; these categories are "no unemployment at all", "less than a year", or "more than a year".

As to regional characteristics, we include dummies for the federal states (Laender) as well as the rural and agglomerated region dummies discussed previously. In addition to these regional characteristics, we have information on the industry or field in which a person is working. This information enters our model in the form of industry dummies. Finally, we use year dummies to capture time effects, i.e., macroeconomic effects and legal changes.

5. Results

5.1 Descriptive Statistics

Before running the multivariate regressions, we provide descriptive statistics of our dataset. Table 1 shows the absolute numbers and the ratios of entrepreneurs and employed individuals, respectively. Out of the 6,078 observations in our sample, 5,246 individuals are classified as employees (86.3 percent), whereas 832 are entrepreneurs (13.7 percent). Furthermore, it can be seen that the share of entrepreneurs as compared to the entire sample varies only slightly between rural and agglomerated areas. Analyzing this in more depth by further disaggregating agglomerated areas into three categories, no perfectly clear structure can be found. However, on average, it seems that rural areas might be slightly more beneficial for entrepreneurship.

Table 1: Occupational status across region types

Regional type	Occupational status		
	Employee	Entrepreneur	Total
Rural area	684 84.76	123 15.24	807 100.00
Agglomerated area	4,562 86.55	709 13.45	5,271 100.00
5,000 - 49,999 inhabitants	1,157 87.45	166 12.55	1,323 100.00
50,000 - 499,999 inhabitants	938 85.66	157 14.34	1,095 100.00
500,000 and more inhabitants	2,467 86.47	386 13.53	2,853 100.00
Total	5,246 86.31	832 13.69	6,078 100.00

The Appendix contains tables and graphs describing the dataset in more detail. Information is given on the distribution of the individuals across the German federal states as well as across industries. Moreover, we present summary statistics for all the remaining independent variables used later in the multivariate regressions, namely, gender, age, house ownership, marital status, nationality, secondary education, and previous unemployment.

Looking at associational activities, we find that, in total, the average number of club memberships is higher for entrepreneurs (1.0) than for employees (0.8). Furthermore, Table 2 illustrates that the average number of club memberships steadily decreases for both employees and entrepreneurs as the area becomes more agglomerated. It is remarkable,

though, that in every type of region, the mean is higher for the entrepreneurs than it is for the employees. While in rural areas with less than 5,000 inhabitants, the entrepreneurs (employees) join 1.7 (1.1) associations/clubs on average, this figure drops to 0.8 (0.7) in big agglomerations with 500,000 or more inhabitants.

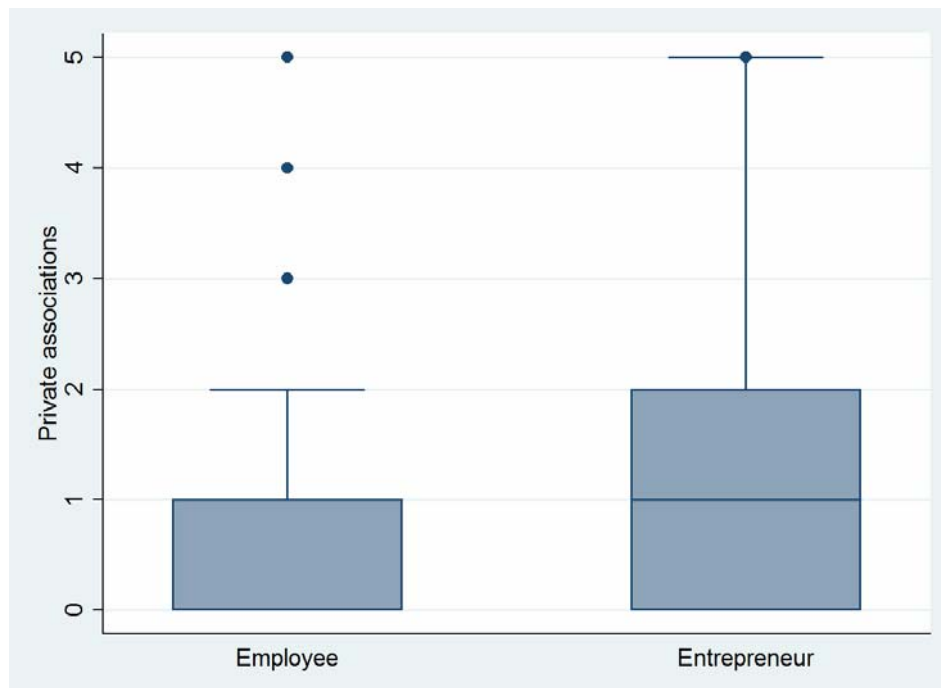
Given that rural and agglomerated areas differ in their average levels of club memberships, our difference-in-differences estimator could also depend to some extent on the functional form assumptions about the effects of club memberships on entrepreneurship. In order to analyze how severe this problem is, we re-run the estimation on the basis of a matched sample with common support between rural and agglomerated areas in the robustness checks.

Table 2: Number of club memberships across region types and occupational status

Regional type	Average number of club memberships		
	Employee	Entrepreneur	All
Rural area	1.09 (1.04)	1.65 (1.28)	1.18 (1.10)
Agglomerated area	.75 (.87)	.91 (1.06)	.78 (.90)
5,000 - 49,999 inhabitants	.91 (.98)	1.16 (1.23)	.94 (1.01)
50,000 - 499,999 inhabitants	.80 (.90)	.99 (1.16)	.83 (.94)
500,000 and more inhabitants	.66 (.79)	.78 (.91)	.68 (.81)
Total	.80 (.90)	1.02 (1.13)	.83 (.94)

Note: standard deviations in parentheses.

Graph 1 depicts box plots of associational activity over occupational status. Again, it can be seen that entrepreneurs join more associations/clubs on average than do employees. A *t*-test rejects the null hypothesis that the means are not statistically different from each other at a 1 percent level of significance. To ensure that our estimation results are not driven by outliers, we exclude individuals with six or more club memberships later when checking the robustness of our multivariate regressions.

Graph 1: Occupational status and associational activity

In sum, the descriptive statistics suggest that entrepreneurs differ from employed people in terms of associational activity. It might well be that people are accumulating social contacts through membership in several associations or clubs, which, in turn, creates a solid foundation to build on when being an entrepreneur. We test if these effects still hold in a multivariate regression framework and at the same time we endeavor to avoid potential omitted variable bias.

5.2 Multivariate Analysis

Applying standard multivariate regression techniques might lead to misleading results regarding our research question. We have concerns about omitted variables, which might account for people being members in associations/clubs but at the same time make them more likely to be an entrepreneur. To tackle this potential self-selection bias, we use modified difference-in-differences methods in our multivariate regressions.

In all the following estimations, occupational status is implemented as a binary dependent variable indicating whether a person is employed or an entrepreneur. Therefore, probit models are used to estimate the effects of social capital on an individual's occupational choice in rural and agglomerated areas. For all estimations, we use clustering robust standard errors, as suggested by Huber (1967) and White (1980), where year and federal state together form one

cluster.⁶ Table 3 sets out the results of our multivariate regressions. In the first probit model, only the control variables are used to predict a person's propensity to be an entrepreneur. The coefficients of the independent variables show the expected signs. Women are less likely than men to be an entrepreneur, which is in line with the findings of e.g. Blanchflower (2000). Being a house or a flat owner, i.e., being a person with assumed low mobility, leads to a higher probability of being an entrepreneur. Married men and women have different time allocation preferences than singles, which makes them less likely to start their own enterprises. Moreover, experience seems to pay off: the age variable as an indicator of experience shows a highly significant, positive coefficient on occupational status. People with a better secondary education more often become an entrepreneur than their less qualified counterparts. Nationality and previous unemployment seem to have no significant effect on a person's occupational choice.

In a second step, we introduce social contacts, measured by associational activity, into our model. Membership in associations and clubs does appear to help explain a person's propensity to be an entrepreneur. The coefficient of the social contacts variable is highly significant and positive, while the coefficients of the control variables still show the same signs and levels of significance; in fact, they hardly change at all. The predictive power of our model increases by introducing social contacts as an additional explanatory variable. However, introducing associational activities as an independent variable and interpreting the positive sign of the coefficient is misleading. This simple strategy cannot show real causal effects of social contacts on individual occupational choice because omitted variables might be responsible for the sign and size of the coefficient. Thus, it might not be social contacts but, instead, underlying unobserved factors that account for the positive impact on a person's propensity to be an entrepreneur.

To overcome these potential problems of endogeneity, we substitute the associational activity variable with two interaction terms and focus on the difference of their coefficients. In these pivotal variables, associational activity is interacted with an agglomerated and a rural dummy, respectively. Table 3 shows the rural area interaction term to be positive and highly significant, whereas the interaction term for agglomerations is not significantly different from zero. A *chi*²-test rejects the null hypothesis that the coefficients of the interactions terms are the same at a 5 percent level of significance. Only this statistically significant difference in the

⁶ Angrist (2001) suggests using simple linear identification strategies when estimating causal effects even though the dependent variable is a 0-1-dummy variable. Therefore, we additionally use a linear probability regression

interaction terms allows us now to discuss a causal impact of social contacts on a person's propensity to be an entrepreneur in rural areas. Assuming that the selection bias is the same in both types of regions and that rural areas are characterized by more interconnected social contacts while agglomerations are characterized by less interconnected social contacts, this difference signifies that more interconnected social contacts are more supportive of being an entrepreneur than are less interconnected social contacts, holding the underlying number of social contacts constant.

The last column of Table 3 lists the coefficients of a linear probability model, as suggested by Angrist (2001). In this least square model with cluster robust standard errors, we qualitatively obtain the same results as in Column 3 of the probit model, which is cautiously interpreted as a first indication that our findings are robust. An F -test rejects the null hypothesis that the coefficients of the interaction terms are the same on a 5 percent level of significance.

Table 3: Determinants of occupational choice

Independent variable	Probit coefficient on OCCUPATIONAL STATUS			Linear coefficient on OCCUPATIONAL STATUS
Associational activity	.0897553 *** (.0267667)			
Associational activity * Rural area	.1854527 *** (.0460897)			.0367856 *** (.0105222)
Associational activity * Agglomerated area	.066073 ** (.0322465)			.0118737 * (.006526)
Gender	-.4969456 *** (.0501642)	-.4801513 *** (.0505616)	-.4784654 *** (.0512638)	-.0896943 *** (.0093752)
Houseowner	.3388627 *** (.0580854)	.3125161 *** (.0593465)	.3169403 *** (.060264)	.0545353 *** (.0099225)
Marital status	-.0956469 * (.0543398)	-.109762 ** (.0554873)	-.1088271 * (.0558204)	-.020894 ** (.0096572)
Foreign	.17684 (.1997472)	.1953591 (.1935641)	.1875131 (.1935347)	.0398791 (.0300495)
Experience	.0280216 *** (.0098722)	.0276995 *** (.0093972)	.028103 *** (.0094792)	.0030666 (.0019799)
Experience ²	-.0000507 (.0001782)	-.0000469 (.0001787)	-.0000552 (.0001794)	.0000329 (.000042)
Education dummies	Yes	Yes	Yes	Yes
Previous unemployment dummies	Yes	Yes	Yes	Yes
State dummies	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Regional type dummies	Yes	Yes	Yes	Yes
Number of observations	5,519	5,519	5,519	5,519
Wald chi2	7,350.58	10,158.78	11,305.21	
Prob > chi2	0.0000	0.0000	0.0000	
F				210.64
Prob > F				0.0000
R-squared				.30738

Note: *** 1% level of significance, ** 5% level of significance, * 1% level of significance; clustering robust standard errors in parentheses.

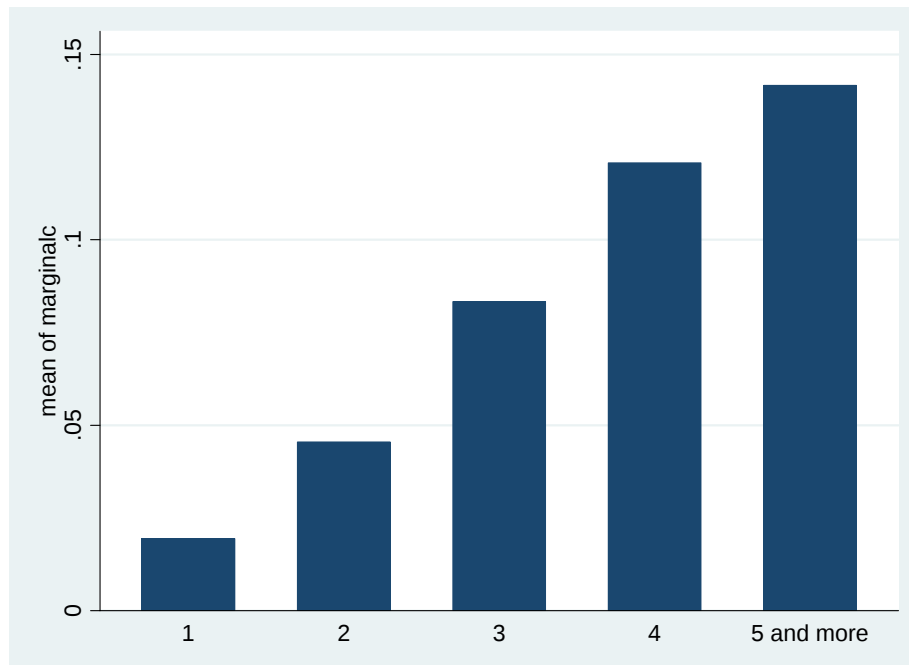
Additionally to the probit coefficients presented in Table 3, marginal effects of social contacts are computed. We want to investigate the effect of intensifying social contacts when all other control variables are held constant. The marginal effects are calculated with the following equation:

$$effect_{mar} = \Phi(\alpha[agg]*agg + \beta_1[sc*rur]*sc*agg + \chi[x]*x) - \Phi(\alpha[agg]*agg + \beta_2[sc*agg]*sc*agg + \chi[x]*x) \quad (4)$$

where Φ is the cumulative standard normal distribution, $\alpha[agg]$ is the probit coefficient of the agglomeration dummy, $\beta_1[sc*rur]$ is the probit coefficient of the interaction term of social contacts with the rural area dummy, $\beta_2[sc*agg]$ is the probit coefficient of the interaction term of social contacts with the agglomerated area dummy, $\chi[x]$ is a vector of probit coefficients of the control variables, and x is a matrix of control variables. Graph 2 shows the average marginal effects over individuals in agglomerations at different numbers of club memberships.

Intensifying social contacts induced by one (two, three, four, five and more) club membership(s) increases a person's probability to be an entrepreneur by 1.9 (3.2, 4.6, 8.3, 12.1, 14.2) per cent.

Graph 2: Marginal effects of intensifying social contacts



5.3. Robustness Checks

We estimate regressions varying from the original ones in several respects in order to test the robustness of our results. First of all, we deal with outliers that could have distorted our results. Graph 1 suggests that employed individuals with three or more club memberships might be treated as outliers. Therefore, we drop these 254 individuals and re-run our regressions on the adjusted sample. Now, the estimation results are even more clear and convincing. The effect of intensifying social contacts becomes even larger and more than doubles for particular groups. This is true for our main identification strategy as well as for all the following robustness checks. However, we feel that the number of observations that can be classified as outliers is rather high. This is why we decide to stick to our original sample. If our results are robust even when including outliers, this makes them even more convincing.

To check that the regional type boundaries are not arbitrary, we introduce a regional variable that groups people into one of four different Boustedt regional types, instead of just the two types (rural and agglomerated areas) used so far. Our rural areas with less than 5,000 inhabitants are classified as *Boustedt type 1* in Table 4. *Boustedt type 2* consists of regions with 5,000 to 49,999 inhabitants, *Boustedt type 3* comprises regions with 50,000 to 499,999

inhabitants, and regions with 500,000 or more inhabitants fall into the category *Boustedt type 4*.

Table 4: Determinants of occupational choice using four region types

Independent variable	Probit coefficient on OCCUPATIONAL STATUS	Linear coefficient on OCCUPATIONAL STATUS
Associational activity * Boustedt type 1	.1857276 *** (.046365)	.0368227 *** (.0105457)
Associational activity * Boustedt type 2	.0801105 (.0508838)	.0172015 * (.010278)
Associational activity * Boustedt type 3	.037919 * (.0208832)	.0056286 (.0044915)
Associational activity * Boustedt type 4	.0262672 (.0429043)	.0048357 (.0081459)
Gender	-.4780266 *** (.0510391)	-.0892134 *** (.0093266)
Houseowner	.3183257 *** (.0611814)	.0541893 *** (.0099975)
Marital status	-.1082081 * (.0557378)	-.0212158 ** (.0096297)
Foreign	.2017466 (.1890388)	.0415406 (.0298528)
Experience	.0283002 *** (.0094091)	.0031095 (.0019677)
Experience ²	-.0000605 (.0001788)	.000032 (.0000419)
Education dummies	Yes	Yes
Previous unemployment dummies	Yes	Yes
State dummies	Yes	Yes
Industry dummies	Yes	Yes
Year dummies	Yes	Yes
Boustedt community type dummies	Yes	Yes
Number of observations	5,519	5,519
Wald chi2	16,600.35	
Prob > chi2	0.0000	
F		299.52
Prob > F		0.0000
R-squared		0.30736

Note: *** 1% level of significance, ** 5% level of significance, * 10% level of significance; clustering robust standard errors in parentheses.

The results of the probit and linear probability regressions, which are presented in Table 4, are in agreement with our original results: associational activities in rural areas have a considerably higher positive effect on the propensity to be an entrepreneur than associational activities in agglomerated areas. In fact, the *chi*²-tests on the coefficients of the interaction terms of the probit model prove that the regions with 5,000 or more inhabitants do not statistically differ from each other with respect to the influence of associational activities on a person's occupational choice. The same can be shown by an *F*-test for the linear probability

model, although the results are not presented here. These findings support our strategy to use only two regional categories, namely, rural and agglomerated areas.

Table 5: Chi²-tests on the coefficients of the interaction terms

Chi²-tests on the probit coefficients	
(Associational activity * Boustedt type 1) - (Associational activity * Boustedt type 2) = 0	
chi2 =	2.30
Prob > chi2 =	0.1295
(Associational activity * Boustedt type 1) - (Associational activity * Boustedt type 3) = 0	
chi2 =	7.93
Prob > chi2 =	0.0049
(Associational activity * Boustedt type 1) - (Associational activity * Boustedt type 4) = 0	
chi2 =	5.62
Prob > chi2 =	0.0178
(Associational activity * Boustedt type 2) - (Associational activity * Boustedt type 3) = 0	
chi2 =	0.66
Prob > chi2 =	0.4179
(Associational activity * Boustedt type 2) - (Associational activity * Boustedt type 4) = 0	
chi2 =	0.62
Prob > chi2 =	0.4321
(Associational activity * Boustedt type 3) - (Associational activity * Boustedt type 4) = 0	
chi2 =	0.07
Prob > chi2 =	0.7898

We additionally use matching techniques to check the reliability of our findings. Angrist (1998) demonstrates that regression and matching estimates can differ substantially from each other even though they control for exactly the same variables. In our matching approach, we keep only those individuals who live in different regions, i.e. agglomerated areas on the one hand and rural areas on the other hand, but apart from this have similar observable characteristics, among them an individual's gender, age, education, or the industry he or she is working in. By doing this, we tackle a potential selection bias through migration and thus enhance the comparability of these two population groups in our sample. However, due to the wide range of control variables, exact matching is not practicable. Rosenbaum and Rubin (1983, 1984, 1985) show that using propensity score matching, i.e., matching on the conditional probability of treatment, is a feasible way to overcome this problem. We thus create five subsamples according to the number of club memberships in a way that, in the end, individuals with no club memberships are in subsample one, individuals with one club membership are in subsample two, and so on. The fifth and last subsample comprises individuals with four or more club memberships. Thus, we guarantee that our results are not only driven by functional form assumptions as average levels of club memberships differ in

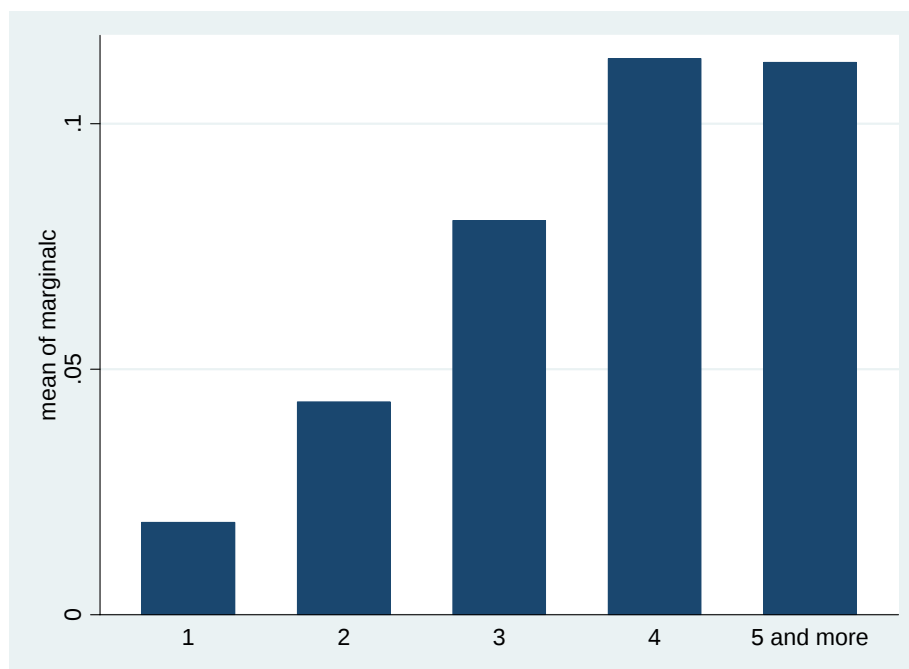
rural and agglomerated areas (cf. Table 2). In a second step, we compute propensity scores for the individuals from the respective subsamples with *rural area* as a binary dependent variable and our remaining established control variables as independent variables. Then, we apply nearest-neighbor matching separately for each of the five subsamples and drop all individuals who could not be matched, i.e., who are neither in the treatment nor in the control group. Thus, we arrive at a sample including only those individuals from agglomerations who have counterparts that strongly resemble them in terms of their probability of living in an agglomerated area but live in rural areas—and the other way round. We then estimate our probit and linear probability models on this matched sample. The results reported in Table 6 confirm our previous findings: associational activities are more supportive of entrepreneurship in rural areas, where they are assumed to represent more interconnected social contacts and thereby trust. An *F*-test can reject the null hypothesis that the coefficients of the interaction terms are the same in the linear probability model at a 10 percent level of significance. For the coefficients of the interaction terms in the probit regression, a *chi*²-test rejects the null hypothesis of the coefficients being the same at a 12 percent level of significance.

Table 6: Determinants of occupational choice after matching

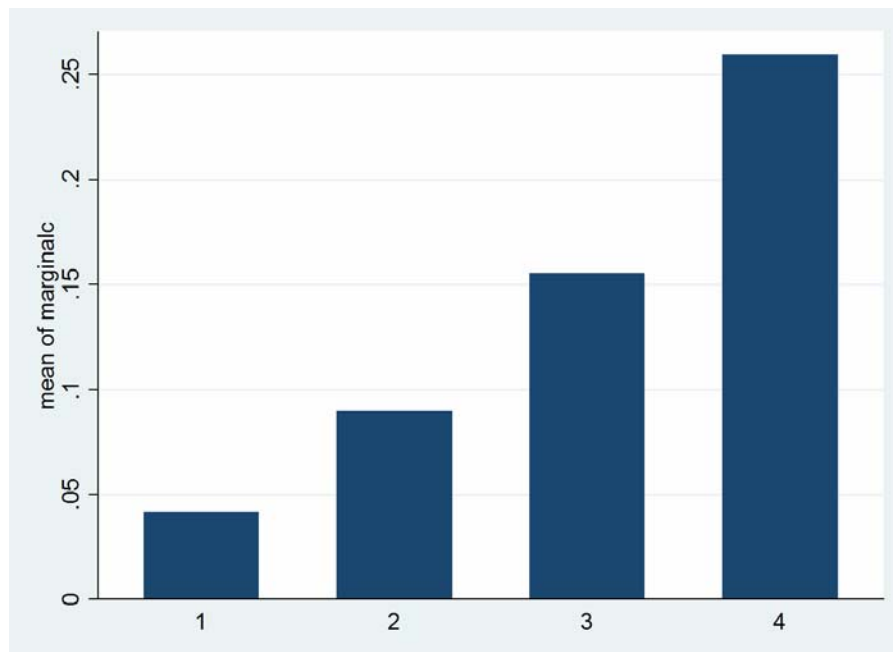
Independent variable	Probit coefficient on OCCUPATIONAL STATUS	Linear coefficient on OCCUPATIONAL STATUS
Associational activity * Rural area	.2068062 *** (.0487698)	.0403045 *** (.0108841)
Associational activity * Agglomerated area	.0932725 ** (.0453631)	.0147578 * (.0088052)
Gender	-.5077738 *** (.056824)	-.0965429 *** (.0105714)
Houseowner	.2588603 *** (.0597815)	.0451398 *** (.0103613)
Marital status	-.1564806 ** (.0665254)	-.0277183 ** (.0116227)
Foreign	.1294084 (.442935)	.0460395 (.156085)
Experience	.0271819 *** (.0091563)	.0029011 (.0019363)
Experience ²	-.0000339 (.0001722)	.0000381 (.0000418)
Education dummies	Yes	Yes
Previous unemployment dummies	Yes	Yes
State dummies	Yes	Yes
Industry dummies	Yes	Yes
Year dummies	Yes	Yes
Regional type dummies	Yes	Yes
Number of observations	4,262	4,262
Wald chi2	10,939.85	
Prob > chi2	0.0000	
F		443.59
Prob > F		0.0000
R-squared		0.3352

Note: *** 1% level of significance, ** 5% level of significance, * 10% level of significance; clustering robust standard errors in parentheses.

Again, we compute marginal effects of intensifying social contacts for the different numbers of club membership following Equation (4). The results are comparable to the ones found above, as can be seen in Graph (3). Intensifying social contacts induced by one (two, three, four, five and more) club membership(s) increases a person's probability to be an entrepreneur by 1.9 (4.3, 4.6, 11.3, 11.2) per cent.

Graph 3: Marginal effects of intensifying social contacts after matching

As already explained earlier in this paper, we do not consider reverse causality to be a major issue in this analysis because occupational choice theory models the decision to be an entrepreneur as a decision that is made repeatedly every period (Kihlstrom and Laffont 1979). Nevertheless, to further check the robustness of our results, we adjust our sample in a way that it only includes individuals who, no longer than one year ago, made a particular kind of occupational choice which either made them entrepreneurs (young entrepreneurs) or choose a new job as employee. If some of these individuals are members of clubs, the point of time they joined these clubs is very likely to lie before their recently made occupational choice. Running the regressions, we find that low observation figures (235 employees and 56 entrepreneurs) render the estimated coefficients statistically insignificant. However, the general effects are still in line with our findings and can be considered to be economically relevant. This is shown in Graph 4, where the marginal effects of intensifying social contacts for the described subsample are depicted.

Graph 4: Marginal effects of intensifying social contacts (young entrepreneurs)

In a final robustness check, we drop all observations from the service sector and estimate our models for the subsample of manufacturing industries only. Again, the overall picture remains the same. The effect of intensifying social contacts for this subsample becomes even larger than in the original regressions.

6. Conclusions

The paper's goal is to show that an individual's social contacts influence his or her decision to become or stay an entrepreneur. Overlapping social contacts can create trust that, in turn, helps overcome potential resource constraints. The underlying mechanism that supports cooperation is determined by network closure. In rural areas, we expect more overlappings between individuals' neighborhoods, leading to more interconnected social contacts. We hypothesize that trust should facilitate the process of resource mobilization in rural areas of smaller community size, which might influence occupational choice in these places as compared to agglomerated areas. Given that the ratio of entrepreneurs does not change considerably with the number of inhabitants within a region, more interconnected social contacts in rural areas take the role of formal institutions in agglomerated areas.

Our empirical identification strategy draws on the differing impact of social contacts on becoming an entrepreneur across regions in order to test the importance of more interconnected social contacts in being an entrepreneur. The advantage of our identification

strategy is that unobserved heterogeneity between individuals that influences both the likelihood of being a social network member as well as the likelihood of being an entrepreneur cancels out and does not interfere with the results. We find that intensifying social contacts increases the likelihood of being an entrepreneur by 1.9 to 14.2 percentage points, depending on the number of underlying social contacts. These results remain robust when applying various approaches, including matching techniques.

If it is shown that social contacts have a significant effect on different economic outcomes such as entrepreneurship (as illustrated in this paper), firm location (Michelacci and Silva 2007), job availability (Bayer *et al.* 2008), or growth (e.g., Knack and Keefer 1997), it would especially be interesting to ask to what extent social contacts or social capital can be affected by policies (cf. Tabellini 2008). Along this line, Aghion *et al.* (2008) present the example of a minimum wage policy having a deleterious effect on the willingness of labor market participants to cooperate. The authors find that in the case of strong state regulations regarding the minimum wage, the labor market becomes characterized by distrustful labor relations and low union density. These findings demonstrate the value of future research on the determinants of social capital or, social contacts.

References

- Aghion, Philippe, Yann Algan, and Pierre 2008.** “Can Policy Interact with Culture? Minimum Wage and the Quality of Labor Relations.”, NBER Working Paper 14327.
- Ajayi-obe, Olufunmilola, and Simon C. Parker.** 2003. “The changing nature of work among the selfemployed in the 1990s: Evidence from Britain.” *Journal of Labor Research*, 26(3): 501–517.
- Akerlof George A., and Rachel E. Kranton.** 2002. “Identity and Schooling: Some Lessons for the Economics of Education.” *Journal of Economic Literature*, 40(4): 1167-1201.
- Allcott, Hunt, Dean Karlan, Markus M. Möbius, Tanya S. Rosenblat, and Adam Szeidl.** 2007. “Community Size and Network Closure.” *American Economic Review Papers and Proceedings*, 97(2): 80-85.
- Amit, Raphael , Lawrence Glosten, and Eitan Muller.** 1990. “Entrepreneurial Ability, Venture Investments, and Risk Sharing.” *Management Science*, 36(10): 1232–1245.
- Angrist, Joshua D.** 1998. “Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants.” *Econometrica*, 66(2): 249–288.
- Angrist, Joshua D.** 2001. “Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice.” *Journal of Business and Economic Statistics*, 19(1): 2–16.
- Audretsch, David B., and Maryann P. Feldman.** 1996. “R&D Spillovers and the Geography of Innovation and Production.” *American Economic Review*, 86(3): 630–640.
- Bayer, Patrick, Stephen Ross, and Giorgio Topa.** 2008. “Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes.” *Journal of Political Economy* (in press).
- Bentolila, Samuel., Claudio Michelacci, and Javier Suarez.** 2008. “Social Contacts and Occupational Choice.” *Economica* (in press)
- Blanchflower, David G.** 2000. “Self-Employment in OECD Countries.” *Labor Economics*, 7(5): 471–505.
- Blanchflower, David G., and Bruce Meyer.** 1994. “A Longitudinal Analysis of Young Entrepreneurs in Australia and the United States.” *Small Business Economics*, 6(1): 1–20.

- Blau, David M.** 1987. "A Time-Series Analysis of Self-Employment in the United States," *Journal of Political Economy*, 95(3): 445–467.
- Bourdieu, Pierre.** 1986. "The Forms of Capital." In *Handbook of Theory: Research for the Sociology of Education*, ed. J. G. Richardson, 241–258. Westport, CT: Greenwood Press.
- Boustedt, Olaf.** 1975. "Grundriss der empirischen Regionalforschung." Vol. 3: Siedlungsstrukturen, Akademie fuer Raumforschung und Landesplanung (ARL), Hannover: Schroedel.
- Bowles, Samuel, and Herbert Gintis.** 2002. "Social Capital and Community Governance." *Economic Journal*, 112(483), F419–F436.
- Brixy, Udo, Rolf Sternberg, and Heiko Stüber.** 2007. "The Relationship Between the Potential of Entrepreneurs and Actual Entrepreneurship at the Regional Level." Paper presented at the North American Regional Science Council Conference, Savannah, GA.
- Brueckner, Jan K., and Ann G. Largey.** 2008. "Social Interaction and Urban Sprawl," *Journal of Urban Economics*, 64, 18-34.
- Burt, Ronald.** 1992. *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard University Press.
- Camerer, Colin, and Dan Lovallo.** 1999. "Overconfidence and Excess Entry: An Empirical Approach." *American Economic Review*, 89(1): 306–318.
- Chwe, Michael.** 1999. "Structure and Strategy in Collective Action." *American Journal of Sociology*, 105(1): 128–156.
- Coleman, James S.** 1988. "Social Capital in the Creation of Human Capital." *American Journal of Sociology*, 94: 95–121.
- Czernich, Nina, Oliver Falck, Thomas Kiessl, and Tobias Kretschmer.** 2008. *Regulierung in Telekommunikationsmärkten: Technologische Dynamik und Wettbewerbspotenziale*. Munich: Ifo Contributions to Economic Research 32.
- Darity, William A., and Arthur H. Goldsmith.** 1996. "Social Psychology, Unemployment and Macroeconomics." *Journal of Economic Perspectives*, 10(1): 121–140.
- Dasgupta, Partha, and Ismail Serageldin, ed.** 1999. *Social Capital: A Multifaceted Perspective*. Washington, DC: World Bank.

- DiPasquale, Denise, and Edward L. Glaeser.** 1999. "Incentives and Social Capital: Are Homeowners Better Citizens?" *Journal of Urban Economics*, 45(2): 354–385.
- Dixit, Avinash.** 2004. *Lawlessness and Economics*. Princeton, NJ: Princeton University Press.
- Durlauf, Steven.** 2002. "On the Empirics of Social Capital." *Economic Journal*, 112(483): F459–F479.
- Ellickson, Robert C.** 1991. *Order Without Law*. Cambridge, MA: Harvard University Press.
- Evans, David, and Boyan Jovanovic.** 1989. "An Estimated Model of Entrepreneurial Choice Under Liquidity Constraints." *Journal of Political Economy*, 97(4): 808–827.
- Evans, David S., and Linda S. Leighton.** 1989a. "The Determinants of Changes in U.S. Self-Employment, 1968–1987." *Small Business Economics*, 1(2): 111–119.
- Evans, David S., and Linda S. Leighton.** 1989b. "Some Empirical Aspects of Entrepreneurship." *American Economic Review*, 79(3): 519–535.
- Fehr, Ernst, and Simon Gächter.** 2000. "Cooperation and Punishment in Public Good Experiments—An Experimental Analysis of Norm Formation and Norm Enforcement." *American Economic Review*, 90(4): 980–994.
- Feldman, Maryann P.** 2001. "The Entrepreneurial Event Revisited: Firm Formation in a Regional Context." *Industrial and Corporate Change* 10(4): 861–891.
- Glaeser, Edward L., Jed Kolko, and Albert Saiz.** 2001. "Consumer City." *Journal of Economic Geography*, 1(1): 27–50.
- Glaeser, Edward L., David Laibson, and Bruce Sacerdote.** 2002. "An Economic Approach to Social Capital." *Economic Journal*, 112(483): 437–458.
- Granovetter, Marc.** 1985. "Economic Action and Social Structures: The Problem of Embeddedness." *American Journal of Sociology*, 91(3): 481–510.
- Granovetter, Marc.** 2005. "The Impact of Social Structure on Economic Outcomes." *Journal of Economic Perspectives*, 19(1): 33–50.
- Hayek, Friedrich A. von.** 1937. "Economics and Knowledge." *Economica, New Series*, 4(13): 33–54.

- Heckman, James J., Hidehiko Ichimura, Jeffrey Smith, and Petra E. Todd.** 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica*, 66(5): 1017–1098.
- Huber, Peter J.** 1967. *The Behavior of Maximum Likelihood Estimates Under Non-Standard Conditions*. Berkeley, CA: University of California Press.
- Jackson, Matthew O., and Brian W. Rogers.** 2007. "Meeting Strangers and Friends of Friends: How Random Are Socially Generated Networks?" *American Economic Review*, 97(3): 890-915.
- Jacobs, Jane.** 1961. *The Death and Life of great American Cities*. New York: Vintage.
- Kihlstrom, Richard E., and Jean-Jaques Laffont.** 1979. "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion." *Journal of Political Economy*, 87(4): 719–748.
- Knack, Stephen, and Philip Keefer.** 1997. "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *Quarterly Journal of Economics*, 112(4): 1251–1288.
- Lazear, Edward P.** 2005. "Entrepreneurship." *Journal of Labor Economics*, 23(4): 649–680.
- Michelacci, Claudio, and Olmo Silva.** 2007. "Why So Many Local Entrepreneurs?" *Review of Economics and Statistics*, 89(4): 615–633.
- Möbius, Markus M., and Adam Szeidl.** 2007. "Trust and Social Collateral." NBER Working Paper No. 13126.
- North, Douglas C.** 1990. *Institutions, institutional change, and economic performance*. Cambridge, MA: Harvard University Press.
- Parker, Simon C.** 2004. *The Economics of Self-Employment and Entrepreneurship*. Cambridge: Cambridge University Press.
- Putnam, Robert D.** 2000. *Bowling Alone*. New York: Simon & Schuster.
- Putnam, Robert D., Robert Leonardi, and Raffaella Y. Nanetti.** 1993. *Making Democracy Work*. Princeton: Princeton University Press.
- Rosenbaum, Paul R., and Donald B. Rubin.** 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, 70(1): 41–55.

- Rosenbaum, Paul R., and Donald B. Rubin.** 1984. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of the American Statistical Association*, 79(387): 516–524.
- Rosenbaum, Paul R., and Donald B. Rubin.** 1985. "Constructing a Control Group Using Multi-Variate Matching Methods that Include the Propensity Score." *American Statistician*, 39(1): 33–38.
- Sanders, Jimmy M., and Victor Nee.** 1996. "Immigrant Self-Employment: The Family as Social Capital and the Value of Human Capital." *American Sociological Review*, 61(2): 231–249.
- Schelling, Thomas C.** 1971. "Dynamic Models of Segregation." *Journal of Mathematical Sociology*, 1: 143–186.
- Schumpeter, Joseph A.** 1912. *The Theory of Economic Development*. New York: Oxford University Press.
- Sobel, Joel.** 1985. A Theory of Credibility, *Review of Economic Studies*, 52(4): 557–573.
- Sobel, Joel.** 2002. Can We Trust Social Capital? *Journal of Economic Literature*, 40(1): 139–154.
- Tabellini, Guido.** 2008. "The Scope of Cooperation: Values and Incentives." *Quarterly Journal of Economics*, 123(3): 905-950.
- Terwey, Michael, Arno Bens, and Stefan Baltzer.** 2007. *Datenhandbuch ALLBUS 1980-2006*, ZA-No. 4241, Cologne, Mannheim: GESIS.
- Watson, Joel.** 1999. "Starting Small and Renegotiation", *Journal of Economic Theory*, 85(1): 52–90.
- White, Halbert.** 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica*, 48(4): 817–838.
- Williamson, Oliver E.** 1979. "Transaction Cost Economics: The Governance of Contractual Relations." *Journal of Law and Economics*, 22(2): 233–261.
- Wu, Brian, and Anne Marie Knott.** 2006. "Entrepreneurial Risk and Market Entry." *Management Science*, 52(9): 1315–1330.

Zak, Paul J., and Stephen Knack. 2001. "Trust and Growth." *Economic Journal*, 111(470): 295–321.

Appendix

Appendix 1: Distribution of individuals across the Bundeslaender

Bundesland	Frequency	Percent
Schleswig-Holstein	258	4.24
Hamburg	230	3.78
Niedersachsen	638	10.50
Bremen	77	1.27
Nordrhein-Westfalen	1,583	26.04
Hessen	557	9.16
Rheinland-Pfalz	365	6.01
Baden-Wuerttemberg	909	14.96
Bayern	1,131	18.61
Saarland	102	1.68
Berlin (West)	228	3.75
Total	6,078	100.00

Appendix 2: Distribution of individuals across industries

Industry	Frequency	Percent
Chemical industry, petroleum processing	212	3.76
Plastics, rubber industry	79	1.40
Earths, stones, fine ceramics	77	1.36
Metal industry	337	5.97
Steel, machines, car manufacturing	663	11.75
Electrical engineering, optics	469	8.31
Wood, paper, printing industry	211	3.74
Leather, textile industry	193	3.42
Food, beverages and tobacco industry	207	3.67
Primary building industry	346	6.13
Secondary building industry	225	3.99
Wholesale industry	358	6.34
Trade negotiations	41	0.73
Retail industry	765	13.56
Transport, information transmission	182	3.23
Credit institutions, banks	200	3.54
Insurances	148	2.62
Diverse services	930	16.48
Total	5,643	100.00

Appendix 3: Industries of the entrepreneurs across regional types

Industry	Regional type		Total
	Rural area	Agglomerated area	
Chemical industry, petroleum processing	0 <i>0.00</i>	1 <i>0.15</i>	1 <i>0.12</i>
Plastics, rubber industry	0 <i>0.00</i>	4 <i>0.58</i>	4 <i>0.50</i>
Earths, stones, fine ceramics	0 <i>0.00</i>	4 <i>0.58</i>	4 <i>0.50</i>
Metal industry	3 <i>2.54</i>	8 <i>1.17</i>	11 <i>1.37</i>
Steel, machines, car manufacturing	2 <i>1.69</i>	7 <i>1.02</i>	9 <i>1.12</i>
Electrical engineering, optics	1 <i>0.85</i>	14 <i>2.04</i>	15 <i>1.87</i>
Wood, paper, printing industry	5 <i>4.24</i>	13 <i>1.90</i>	18 <i>2.24</i>
Leather, textile industry	2 <i>1.69</i>	21 <i>3.06</i>	23 <i>2.86</i>
Food, beverages and tobacco industry	7 <i>5.93</i>	10 <i>1.46</i>	17 <i>2.11</i>
Primary building industry	4 <i>3.39</i>	14 <i>2.04</i>	18 <i>2.24</i>
Secondary building industry	10 <i>8.47</i>	42 <i>6.12</i>	52 <i>6.47</i>
Wholesale industry	3 <i>2.54</i>	21 <i>3.06</i>	24 <i>2.99</i>
Trade negotiations	2 <i>1.69</i>	31 <i>4.52</i>	33 <i>4.10</i>
Retail industry	23 <i>19.49</i>	152 <i>22.16</i>	175 <i>21.77</i>
Transport, information transmission	3 <i>2.54</i>	29 <i>4.23</i>	32 <i>3.98</i>
Credit institutions, banks	1 <i>0.85</i>	3 <i>0.44</i>	4 <i>0.50</i>
Insurances	7 <i>5.93</i>	27 <i>3.94</i>	34 <i>41.04</i>
Diverse services	45 <i>38.14</i>	285 <i>41.55</i>	330 <i>16.48</i>
Total	118 <i>100.00</i>	686 <i>100.00</i>	804 <i>100.00</i>

Note: percentages in italics

Appendix 4: Further distribution figures for independent variables

Independent variables		Frequency	Percent
Gender	Male	3,952	65.02
	Female	2,126	34.98
	Total	6,078	100.00
Houseowner	Tenant	3,323	55.02
	House/flat owner	2,717	44.98
	Total	6,040	100.00
Marital status	Single	2,155	35.46
	Married	3,922	64.54
	Total	6,077	100.00
Nationality	German	6,022	99.13
	Not German	53	0.87
	Total	6,075	100.00
Education	No secondary	66	1.09
	Lower secondary	3,360	55.55
	Medium secondary	1,627	26.90
	Upper secondary	406	6.71
	University degree	590	9.75
	Total	6,049	100.00
Previous unemployment	No	5,008	83.24
	Less than a year	715	11.99
	More than a year	293	4.87
	Total	6,016	100.00

Appendix 5: Experience distribution of individuals

