The goldilocks effect of diversity?

The effect of team diversity on new venture performance

Copernicus Institute of Sustainable Development

Department Innovation, Environment and Energy Sciences

Matthee le Cointre - 3510271
W.M.leCointre@students.uu.nl

Supervisors

Koen Frenken - Professor
K.Frenken@uu.nl

Frank van Rijnsoever - Assistant Professor
F.J.vanRijnsoever@uu.nl

Martin Andersson - Professor
Martin.Andersson@circle.lu.se

This page intentionally left blank

Abstract

Conflicting arguments and inconclusive findings on the effect new venture team diversity on the performance of new ventures dominate existing literature. In this research, I investigate the relationship between educational, industrial, organizational and occupational diversity and the performance of new ventures. I address several research gaps by including multiple dimensions of diversity, allowing for non-linear effects and including longer career histories of individuals. The effect of diversity is quantitatively tested on the survival and sales of new ventures using data from Statistics Sweden. A total of 891 software programming ventures founded between 2002 and 2009 doing are included. The results show that each diversity dimension has its own distinctive influence on the performance of new ventures, and this influence varies for different performance measures. First, educational diversity has a negative effect on sales and is unrelated to survival. Second, industrial diversity does not appear to have any relation at all to the performance of new ventures. Third, organizational diversity positively effects sales and shows an inverted U-shaped relationship to survival. Last, occupational diversity shows an inverted U-shaped effect to sales but has no effect on survival. The results imply that scholars should be careful with applying diversity arguments uniformly. Moreover, the prominent role of industrial experiences and similar organizational experiences in literature might need to be reconsidered.

Acknowledgements

I am grateful to everyone at CIRCLE for giving me the opportunity to work in its intellectual and open environment and providing me with the necessary resources to perform this research. Special thanks are reserved for my supervisors – Koen Frenken, Frank van Rijnsoever & Martin Andersson - for taking the time to provide valuable feedback and inspiration during the process. Last, I would like to thank my fellow students who helped me put my thoughts into words and offered criticism when needed.

Table of contents

1. Introduction

New ventures are regarded as being of significant importance for economic development and technological change (Klepper & Sleeper, 2005). They create new job opportunities and drive competition within industries (Dahl & Reichstein, 2007; Klepper, 2009). New ventures have an initial resource base and starting capabilities that are embedded in the individuals involved (Helfat & Lieberman, 2002; Wright et al., 2007). Most new ventures are started by multiple individuals that form a new venture team (NVT) rather than by lone entrepreneurs (Kamm et al., 1990; Klotz et al., 2013; Reich, 1987). NVT members can reproduce routines and inherit knowledge from their previous employers, gain knowledge through education and accumulate human capital (Agarwal & Echambadi, 2004; Burton et al., 2002; Colombo & Grilli, 2005). These experiences shape the strategic positioning, routines and resource mobilization of the new venture and have a salient and long-lasting impact on the performance of new ventures (e.g. survival and sales) (Agarwal & Echambadi, 2004; Klepper, 2001). Because the backgrounds of individuals vary, NVT members may bring different experiences to the table (Melillo et al., 2012). But what are the consequences of differences between experiences of NVT members for the new venture?

Some scholars argue that NVTs with different experiences have a positive effect on the performance of new ventures (van Knippenberg & Schippers, 2007). Diverse NVTs have a broader knowledge base, a variety of resources available and possess more unique ideas and perspectives (Beckman, 2006; Frederiksen & Wennberg, 2011; Helfat & Lieberman, 2002). On the other hand, overlap between NVT members' experiences – i.e. low diversity - also seems to be important for the performance of new ventures (Klepper, 2001; van Knippenberg & Schippers, 2007). Similar individuals communicate better due to shared understanding, enabling them to act faster, increase coordination and build trust (Beckman, 2006; Eisenhardt & Schoonhoven, 1990; Eriksson & Kuhn, 2006). Thus, two conflicting explanations exist for the effect of NVT diversity on performance of new ventures. However, neither the positive nor the negative effect can be empirically supported (Kakarika & Biniari, 2011; Klotz et al., 2013). The integration of the conflicting explanations leads to the hypothesis of an inverted U-shaped effect of NVT diversity on performance (van Knippenberg & Schippers, 2007; Webber & Donahue, 2001; Williams & O'Reilly, 1998). In other words, there may be an optimal level of diversity within NVTs of new ventures that enables them to reap the benefits from different experiences, while maintaining a level of similarity necessary for effective cooperation. This resembles a goldilocks effect between NVT diversity and new venture performance; not too much, not too little but just right.

The literature on team diversity has mainly been concerned with demographic characteristics and professional experiences (Jackson et al., 2003; Klotz et al., 2013). The former consists of gender, age and ethnic background of team members (Østergaard et al., 2011). However, NVTs are often homogeneous with respect to the demographic diversity dimensions (Hellerstedt, 2009; Ruef et al., 2003; Steffens et al., 2012) which are therefore not the focus of this study. The latter - professional experiences - can be obtained by education or in a job (Horwitz & Horwitz, 2007). This is reflected by research on for example diversity in types of education or diversity of managerial, technical and commercial experience (Hellerstedt, 2009; Ucbasaran et al., 2003). It is these professional experiences that NVTs mainly rely on for resource mobilization and decision making (Dencker et al., 2009).

The existing research on diversity of NVTs shows several research gaps. First, existing studies on diversity only include one or a few dimensions of diversity (Hellerstedt, 2009). However, diversity should be seen as a multidimensional concept because differences between individuals' experiences can stem from different sources (Williams & O'Reilly, 1998). I intend to further delve into the sources of diversity by including multiple dimensions and investigate their relative importance. Therefore, I combine educational, organizational, industrial and occupational diversity in this research. These dimensions are inspired by two bodies of literature dealing with diversity in teams: the psychology and management literature on the one hand (Jackson et al., 2003; Milliken & Martins, 1996) and entrepreneurship literature on the other hand (Dahl & Reichstein, 2007; Klepper, 2009).

Second, existing research has mainly focused on assigning positive or negative linear effects to diversity dimensions (van Knippenberg & Schippers, 2007). However, van Knippenberg et al. (2004) argue that each dimension may have positive as well as negative effects on performance, possibly leading to curvilinear relationships (van Knippenberg & Schippers, 2007). Thus far, non-linear effects of diversity dimensions within teams has received little attention (Ali et al., 2013; Chi et al., 2009; Vegt & Bunderson, 2005). To investigate the goldilocks effect of diversity, I look at inverted U-shaped relationships between diversity dimensions and performance. Last, the inclusion of industrial, organizational and occupational experiences of individuals in existing studies is limited to the previous employer of the NVT members rather than their whole career history (Klepper, 2001; Klotz et al., 2013; van Knippenberg & Schippers, 2007). As a result, earlier job experiences that could influence diversity within the NVT are excluded. This research includes a longer organizational and industrial career history in determining the diversity of NVTs[1], and sets out with the following research question:

> *What is the effect of educational, industrial, organizational and occupational diversity of new venture teams on the performance of new ventures?*

The effect of diversity on performance is quantitatively tested on new ventures in the Information Communication Technology (ICT) industry involved in software programming by using data from Swedish matched employer–employee databases (Andersson & Klepper, 2013). This research contributes to existing literature by showing that each diversity dimension has a distinctive effect on the performance of new ventures. Both positive and negative linear effects as well as inverted U-shaped effects on performance are found. Moreover, industrial diversity does not seem to be related to the performance of new ventures. From a managerial perspective, the role of the NVT is considered paramount to the performance of new ventures. For example, investors are more prone to bet on the NVT than the business idea (Delmar & Shane, 2006). This research can provide helpful guidelines on how to compose teams to increase the performance of new ventures, and hence stimulate economic development and technological change. The remainder of this thesis consists of a theoretical section, methodology, results and a discussion.

[1] Due to data limitations, a longer history of occupational experiences can unfortunately not be included.

2. Theory

In this section, first an introduction is given to NVTs, the performance of new ventures and the role of past experiences of individuals. Thereafter, I delve into the concept of diversity and discuss mechanisms that relate diversity of NVTs to performance. Last, I formulate hypotheses for the diversity dimensions.

2.1 NVTs, performance & experiences

New ventures are usually considered small and young firms that are in the process of developing their business (Klotz et al., 2013). In line with, Klotz et al. (2013) I define new venture broadly as: "a firm that is in its early stages of development and growth". This research is concerned with the group of individuals that are involved in the new ventures, called the NVT[2]. All individuals initially involved in the new venture are assumed part of the NVT, because they are considered to be involved in running the venture and contribute to decision-making (Hellerstedt, 2009). The majority of the teams are formed by people who trust each other, due to friendship or shared personality (Kakarika, 2013). However, individuals also try to find NVT members with complementary skills and resources (Kamm et al., 1990; Ruef, 2002). The process of team formation is associated with searching costs for appropriate individuals, but also costs for dividing ownership and tasks (Kamm et al., 1990). Over time, NVTs are subject to continuous development and occasional change in membership (Vanaelst et al., 2006). Team composition might change as a result of external shocks or internal conflicts and needs (Clarysse & Moray, 2004). New members generally share the perspectives of the NVT on the venture, but bring in new resources and skills (Vanaelst et al., 2006). Moreover, NVT members might acquire new skills and knowledge during the initial phases of the new venture (Clarysse & Moray, 2004). Defining a NVT at a particular moment in time poses difficulties for incorporating these dynamics. In this research, I use the moment of formal establishment of the new venture to define NVTs, because at that point all individuals show entrepreneurial commitment to the venture (Vanaelst et al., 2006).

The NVT members are considered to make essential strategic decisions that have long-lasting impacts on the firm (Beckman, 2006). The main challenge of new ventures is to overcome the liability of newness (Shane & Khurana, 2003; Stinchcombe, 1965). The first step in overcoming it, is the identification of a business opportunity, which is the potential to achieve a profit (Shane, 2000). Second, new ventures need resources to capitalize on the business opportunity and must overcome skepticism of resource providers by creating legitimacy (Burton et al., 2002). Third, new ventures lack the internal structures and processes of established organizations, which can be costly and time-consuming to build (Stinchcombe, 1965). New ventures can be considered successful when they (1) survive by overcoming the liability of newness, and (2) grow as a result of adequate business opportunities, resources and processes (Cooper et al., 1994).

The previous experiences of NVT members help them to overcome the liability of newness (Shane & Khurana, 2003). For example, the identification of a business opportunity is based on information that individuals already possess and is formed in interaction with other organizations and individuals (Furlan & Grandinetti, 2014; Shane, 2000). Moreover, the availability of resources of the new venture is affected by networks and knowledge of NVT members (Eisenhardt & Schoonhoven, 1990) and members base their own roles and that of

[2] 'Founding team' and 'entrepreneurial team' are used as synonyms in different literature streams (Hellerstedt, 2009).

others within the new venture on past experiences (Shane & Khurana, 2003). A primary source of experience of individuals is previous employers, from which individuals inherit knowledge and know-how (Agarwal & Echambadi, 2004). More specifically, employees attempt to imitate or recombine their previous employers' resources to help their new venture (Franco & Filson, 2006; Klepper & Sleeper, 2005). However, the ability of NVT members to inherit resources is limited to non-technical knowledge that is embedded in the individual (Andersson et al., 2010; Chatterji, 2009; Helfat & Lieberman, 2002). The work experience of individuals can also result in specific skills gained through occupational background or knowledge and networks resulting from experience in specific industries (Klotz et al., 2013). Another source from which individuals inherit resources, is their education (Brüderl et al., 1992). Educational experience shapes individuals' working methods in how to identify and solve problems and gives them an initial knowledge base (Østergaard et al., 2011). Besides direct inheritance of know-how, NVT members may mobilize resources that are not embedded in the individual by using formal and informal interpersonal networks (Furlan & Grandinetti, 2014).

2.2 Diversity & dimensions

Diversity may be seen as a characteristic of a social group, in this case the NVT, and reflects the degree to which there are differences between group members (van Knippenberg & Schippers, 2007)[3]. At the low end of the diversity scale, there are individuals that have similar experiences and at the high end of the scale, individuals differ in their experiences (van Knippenberg et al., 2004). Stirling (2007) poses three properties of diversity: variety, balance and disparity. Variety is the number of categories to which individuals are apportioned, where more categories indicate higher diversity. A category refers to a specific sort of experience, for example a certain job position or type of education. Balance refers to the pattern of occurrence of individuals in each category, where a higher occurrence of individuals in a category indicates low diversity. Disparity is the degree to which categories are different, but in this study only variety and balance are taken into account[4]. With respect to balance, it is important to note that individuals may have gathered different experiences throughout their entire career and as a result may not only occur in one category but in multiple categories. For example, individuals may have worked in multiple industries. In sum, the concentration of experiences in one category leads to lower diversity, where the addition of more categories leads to higher diversity.

To organize the theoretically limitless dimensions of diversity, several typologies of diversity have been proposed[5]. The first typology contains ascribed attributes which are given by birth (Hellerstedt, 2009). They are also called readily observable or demographic attributes, and exist of gender, age and ethnic dimensions (Horwitz & Horwitz, 2007; Jackson et al., 2003). The second typology is achieved or professional attributes, which are the result of individuals' choices and achievements (Hellerstedt, 2009). Alternatively, achieved attributes are also called underlying attributes because they are not directly observable (Milliken & Martins, 1996). In the majority of the NVT diversity literature, these achieved attributes are conceptualized as functional diversity (Beckman et al., 2007; Cantner et al., 2010; Chandler & Jansen, 1992; Ensley

[3] Within different literature streams, the terms 'diversity' and 'heterogeneity' are used to refer to the same concept (Hellerstedt, 2009).

[4] I do not take disparity into account because no proven measure can deal with all three properties of diversity (Stirling, 2007). Moreover, this research is concerned with overlap and differences between experiences of NVT members, which reflect the balance and variety properties. Consequently, disparity is the least relevant property of diversity in this research and is therefore not included.

[5] An alternative set of dimensions - such as personality, values and beliefs - are also recognized in literature, but often ignored for methodological reasons (Jackson et al., 2003; Milliken & Martins, 1996).

et al., 1999; Jones-Evans, 1996; Leary & DeVaughn, 2009). However, the concept of functional diversity is poorly defined and its operationalization is inconsistent (Cantner et al., 2010; Kakarika & Biniari, 2011). Moreover, all professional experience is consolidated in a single diversity measure when using functional diversity and thus the sources of this experience remains unclear (Cantner et al., 2010; Klotz et al., 2013). Therefore, I use diversity dimensions that relate to the sources of professional experience and include educational, industrial, organizational and occupational diversity (Hellerstedt, 2009; Horwitz & Horwitz, 2007; van Knippenberg et al., 2004).

2.3 Mechanisms of diversity

Two conflicting arguments explain the effect of diversity on performance of new ventures: the high diversity argument and the low diversity argument (Williams & O'Reilly, 1998). The former argues that diverse NVTs have a broader knowledge base, different resources available and possess more unique ideas and perspectives (Beckman, 2006; Frederiksen & Wennberg, 2011; Helfat & Lieberman, 2002; van Knippenberg et al., 2004). Moreover, NVT members have increased access to informational networks outside the group (Williams & O'Reilly, 1998). Access to different networks enable them to mobilize a broader variety of resources (Frederiksen & Wennberg, 2011; Furlan & Grandinetti, 2014). To acquire scarce and unique resources from their networks, individuals need to be firmly embedded in the network (Martinez & Aldrich, 2011). NVT members can use weak ties as a source of information on alternatives and market opportunities, which open opportunity for new innovative combinations (Burt, 2004; Stam & Elfring, 2008). The variety of insights available in diverse NVTs increases conflict in decision making, which leads to higher quality problem solving skills and promotes creativity and innovation (Eisenhardt & Schoonhoven, 1990; Horwitz & Horwitz, 2007). Diverse teams are especially helpful in solving non-routine problems (van Knippenberg & Schippers, 2007). Problems faced by NVTs, such as innovations and product design, benefit most from multiple perspectives (Williams & O'Reilly, 1998). Thus, high diversity of NVTs' experiences enable better opportunity recognition, broaden the set of available resources and advance problem solving skills, which positively affects the performance of new ventures.

On the other hand, the low diversity argument rests on effective cooperation and communication which is needed to streamline group processes and increase performance (van Knippenberg & Schippers, 2007). The underlying mechanism works as follows: individuals perform social comparisons between different group members and classify others based on differences they observe. As a consequence, NVT members are fictionally divided into various subgroups (van Knippenberg & Schippers, 2007; Williams & O'Reilly, 1998). Individuals are more prone to cooperate with people in the same subgroup (van Knippenberg & Schippers, 2007). As a consequence, in case of highly diverse NVTs, communication and cohesion between all members of the group is hindered and may even lead to personal conflicts (Williams & O'Reilly, 1998). Next to communication at the personal level, low diversity also promotes task related communication. Due to shared understandings, NVTs can obtain and use external knowledge and act more quickly (Beckman, 2006; Cohen & Levinthal, 1990). Moreover, individuals that have worked together learn to communicate together, resulting in better coordination and trust (Eisenhardt & Schoonhoven, 1990). Evidence suggests that trust among NVT members is influential for the performance of new ventures and positively influences the speed of decision making (Eriksson & Kuhn, 2006; Talaulicar et al., 2005). Cohesion between members of the NVT increases task-related conflict but reduces personal conflicts (Ensley et al.,

2002). In sum, both high and low diversity of NVTs are argued to be essential for the performance of new ventures. If the diversity is too high, NVTs lack a common basis for effective cooperation but if diversity is too low, a lack of diverse resources and skills hinder the performance. Hence, I suspect a goldilocks effect of diversity on the performance of new ventures, where moderate diversity allows NVTs to cooperate and profit from diverse insights at the same time.

2.4 Hypotheses

The conflicting mechanisms can possibly be applied to all dimensions of diversity (van Knippenberg et al., 2004). However, because each dimension may have a unique effect on performance, the diversity dimensions are not aggregated into a single diversity measure but analyzed separately (Harrison & Klein, 2007; van Knippenberg & Schippers, 2007). Due to conflicting findings in literature, linear positive or negative effects of diversity may also exist and are included in the empirical research while no explicit hypotheses are formulated (van Knippenberg & Schippers, 2007). Moreover, diversity dimensions may have different effects on other sorts of performance (Murphy et al., 1996; Timmermans & Coad, 2013). However, due to the complex nature of the relationship (see 2.3 Mechanisms of diversity) no theoretical reasons ex ante predict divergent effects of diversity on a specific sort of performance. Similarly, the relative importance of the diversity dimensions is unclear beforehand, due to the inconsistent findings in literature and novel combination of diversity dimensions.

Educational diversity

Education relates to the type of education of the NVT members, where large differences indicate high diversity and similar education indicates low diversity (Milliken & Martins, 1996). Differences in education type are a source of informational variety leading to higher absorptive capacity of firms (Cohen & Levinthal, 1990; Hellerstedt, 2009; Østergaard et al., 2011). Educational diversity is found to have a positive influence on NVT and new venture performance (Milliken & Martins, 1996; Østergaard et al., 2011; Timmermans & Coad, 2013). Schlosser (2013) finds that key individuals in new ventures often differ in educational background. At the other end of the spectrum, educationally similar NVTs are likely to have higher attraction and cohesion (Hellerstedt, 2009). Moreover, low educational diverse teams are found to be better in diffusing information within the NVT (Østergaard et al., 2011). In new ventures with a high degree of novelty or ventures experiencing high growth, educational diversity has a negative effect on performance (Amason et al., 2006; Ensley et al., 1999). The middle ground could allow NVTs to have best of both extremes, and therefore I hypothesize:

> *H1: Educational diversity of NVTs has an inverted U-shaped relationship to new venture performance*

Industrial diversity

The industry dimension of diversity consists of the different industries in which the NVT members have worked. High diversity in industry experience of the NVT is found to be positively related to new venture performance (Frederiksen & Wennberg, 2011; Hellerstedt, 2009). Fern et al. (2012) finds that industrial diversity lessens the constraints on strategic choice and that unique rather than shared experiences mainly influence this decision. On the other hand, industrial diverse groups communicate poorly and were less integrated, limiting their ability to make successful strategic decisions (Milliken & Martins, 1996). Other studies find that industry specific experience rather than many different experiences is essential for performance (Dahl &

Reichstein, 2007; Klepper, 2001; Roure & Maidique, 1986). In a recent study, Timmermans & Coad (2013) confirm the positive effect of specific industry experience of NVTs on survival but find a negative effect on employee growth, which implies a more complex interaction between industry experience and performance. The effect of industrial experience is found to be non-linear to performance of new ventures, where NVTs with between 11 and 15 years of experience perform better (Delmar & Shane, 2006). The overall findings suggest that NVTs need specific industry experience but need to complement it with experiences from other industries as well. Thus, I hypothesize:

> *H2: Industrial diversity of NVTs has an inverted U-shaped relationship to new venture performance*

Organizational diversity

Organizational diversity encompasses the previous employers for which the NVT members have worked. Each organization has a specific culture and set of norms and routines, which individuals transfer to the new venture (Agarwal & Echambadi, 2004; Burton et al., 2002). When individuals have been employed at multiple different organizations, they bring a variety of perspectives that enhance creative solutions to problems and performance (Milliken & Martins, 1996). Startups with NVT members that have worked for many different employers are found to have better access to venture capital and have higher chances of successful public offerings (Beckman et al., 2007). However, Klepper, (2001), Timmermans (2010) and Roure & Maidique (1986) find that NVT members that have worked together at the same organization perform better. Teams that have a similar organizational background learn to communicate together resulting in better coordination and trust which improves the performance of the new venture (Eisenhardt & Schoonhoven, 1990). Moreover, prior joint work experience helps to overcome the lack of organizational culture and routines (Timmermans, 2010). New ventures with NVTs that have both overlap and variety of experiences at organizations have additional advantages over teams that lack one of them (Beckman, 2006). Hence, I hypothesize:

> *H3: Organizational diversity of NVTs has an inverted U-shaped relationship to new venture performance*

Occupational diversity

Occupational diversity of NVTs is reflected by the different job positions that NVT members have fulfilled in the past. High occupational diversity of NVT members is found to have a positive effect on performance (Klepper, 2001; Williams & O'Reilly, 1998). Employees with technical and management experiences differ in their occupational background (Jones-Evans, 1996). Roure & Maidique (1986) find that successful new ventures fulfill more critical roles in the NVT and tend to have previous experience in that particular function. However, Milliken & Martins (1996) report a negative relation to team innovation and performance. The lack of effective communication, which stems from limited mutual understanding, leads to less effective cooperation (Milliken & Martins, 1996). In a study on bank start-ups, Leary & DeVaughn (2009) do not find any significant effect if occupational diversity on a successful launch. The conflicting effects of occupational diversity on performance of new ventures lead to the following hypothesis:

> *H4: Occupational diversity of NVTs has an inverted U-shaped relationship to new venture performance*

3. Methodology

3.1 Data

The hypotheses formulated in the theory section are tested using a quantitative longitudinal approach. The unit of analysis is the new venture and the data stems from multiple Swedish matched employer and employee databases. These databases contain information on all firms and employed individuals in the country and are updated yearly by Statistics Sweden (SCB, 2015). The data gathered and processed by the SCB is widely used and considered to be highly consistent and reliable (Andersson & Klepper, 2013; Frederiksen & Wennberg, 2011; Hellerstedt, 2009). The first database is RAMS[6], which uses labor market statistics to provide basic individual characteristics as well as matching individuals to employers. This data is limited from 1985 to 1989, after which its successor LISA was introduced. This database includes all legal residents of Sweden over the age sixteen and offers more enriched individual information gathered from individual tax statements, financial records, birthplace registries, and school records (Frederiksen & Wennberg, 2011). A third source is the FAD database, which was constructed based on the data of previously mentioned matched employee-employer databases. FAD offers more accurate data for identifying and tracking organizations, by making use of labor mobility patterns rather than legal identification methods (Andersson & Klepper, 2013). In addition to the data matching employees to employers, FRIDA is used to match stakeholders to new ventures as well. The data for FRIDA stems from Sweden's firm registers. Finally, FEK - a database on business statistics constructed based on tax reports of the Swedish fiscal authorities – is used to obtain financial information of new ventures (Frederiksen & Wennberg, 2011).

To select a set of comparable new ventures for this research, a list of strict conditions is used. First, the new ventures have to be founded in the period of 2002 till 2009. The upper limit of the timeframe is restricted by the availability of financial data three years after founding. The lower limit is set at 2002, because information on the occupational status of individuals is only available from 2001 onwards. Second, I only want to include ventures that are genuinely new and not continuations of existing firms (see Dahl & Reichstein, 2007; Timmermans & Coad, 2013). The FAD database entails a coding scheme that uses the flow of employees rather than ownership data, location or sector code to identify new firms. Based on the flow of employees, it determines whether the firm is new or a continuation of an existing firm (SCB, 2015). Besides the firm level, the employees are also followed at the establishment level (i.e. company locations or plants). I require new ventures to have a new configuration of employees at both the firm and establishment level. Hence, I exclude continuations of existing firms or parts of them. To exclude mergers and takeovers as well, the number of establishments in the founding year of the new venture is a maximum of 1. Third, to ensure that the ventures are new it requires a new legal entity. Thus, firms with the same legal number as the year before are excluded, as are firms that are identified as being new firms in multiple years but share the same legal number. Using this method eliminates the self-selection and survival bias widespread in research on new ventures, because all new ventures in the population are included in the research (Fern et al., 2012).

I further delimit to the Information Communication Technology (ICT) industry because ICT is a strongly growing sector in advanced economies and represents a large fraction of new ventures (Davis & Schaefer, 2003). Industry information is based on the 5-digit Statistical Classification of

[6] Database abbreviations: RAMS - Register-based Labour Market Statistics; LISA – Longitudinal Individual Database; FAD –Firm and Establishment Dynamics; FRIDA – Firm Register and Individual Database; FEK - Företagens Ekonomi

Economic Activities in the European Community (NACE) present in the RAMS and LISA database (SCB, 2015). Specifically, I select ventures that are involved in the publishing of software and hence exclude publishing, TV, radio and telecommunication organizations. Software programming is highly skill intensive and resources are tacit and not embedded in physical infrastructure. Therefore, the inheritance process can more easily occur in the software industry, strengthening the effect of individuals' experiences on their new venture (Andersson et al., 2010). Sampling based on this specific industry removes contextual differences between industries that may affect the findings, but limits the generalizability of the results to other industries.

Next, the individuals forming the NVT of the new ventures need to be identified. Making use of the FAD database, the initial employees are matched to the selected ventures. However, the owners of the company are also part of the NVT and should therefore be included. The owners of the new ventures are identified in the FRIDA database. Double occurrences of individuals that are both employee and stakeholder of the new ventures are removed, to ensure that each individual is only counted once. Based on the total count of individuals involved in the NVT a new variable indicating the NVT size is constructed. Because this research focuses on mechanisms within teams, only ventures with a minimum NVT size of two are selected. To function as a team, the individuals involved should have more or less a similar position and influence. In larger teams, a subset of individuals might have a more dominant position and function as management whereas others have a more limited influence on routine building and decision making (Schlosser, 2013). Therefore, in line with Steffens et al. (2012) only new ventures with a NVT size maximum of five are selected.

3.2 Operationalization

Performance

For measuring performance I take a delay between founding year and the measurement of performance of three years[7]. This period is taken because it is seen as a critical threshold for ICT startups, after which successful and unsuccessful ventures can be identified (Lasch et al., 2007). Performance of new ventures is a multidimensional concept and may reflect different qualities of the venture, such as its efficiency or growth (Murphy et al., 1996; Timmermans & Coad, 2013). Performance measures might be uncorrelated or even negatively correlated to each other. As a result independent variables might be positively correlated to one performance measure and negatively to another. Therefore, scholars should include multiple performance measures and make specific conclusions rather than generalize to overall performance (Murphy et al., 1996). Because of the differing characteristics of performance measures, the performance of new ventures is measured in two ways: survival and sales (Brush & Vanderwerf, 1992). First, survival indicates whether the new venture was able to overcome the liability of newness. Survival is a binary variable that is calculated based on the existence of employees and of sales after three years. Moreover, the survival variable is corrected for ventures that have been acquired by other firms, because this is considered a positive exit. Second, sales are used as a size indicator, which shows to what degree the new venture is able to create value in the marketplace. The data for the sales variable originates from the FEK database. Because the sales indicator is heavily skewed, it is log transformed.

[7] In Appendix B, a detailed overview of the databases & variables used and data transformation applied, is given.

Independent variables

The longitudinal nature of the data allows for the career paths of individuals to be identified and is used to determine the diversity of the NVT. No restrictions for the experience are used, so new ventures founded by individuals with limited work experience are also included. It is important to note that no career history information is available prior to 1985, limiting the information on early career paths of older individuals. However, because I select new ventures from 2002 onwards, a considerable period of previous experiences is always included for calculating the diversity (Hellerstedt, 2009). The Shannon-Weaver entropy index is used to construct a diversity measure for each NVT based on the past experiences of individuals in a particular diversity dimension (Frenken et al., 1999; Stirling, 2007). The entropy index essentially measures the evenness of the distribution of experiences over different categories, and is based on the following formula:

$$Diversity = \sum_{i=1}^{A} p_i \log\left(\frac{1}{p_i}\right) \qquad (i = 1, \dots, A)$$

In which p is the proportion of measurement units in the particular category i, and A is the amount of categories present in the NVT. The measurement units take the form of individuals, particular experiences or a year of experience and can only be assigned to a single category (Frenken et al., 1999). If the measurement units are concentrated in a limited number of categories and thus overlap in experience exists, the entropy indicates low diversity. When the units are spread over multiple categories, the diversity is high. Hence, the entropy measure takes both balance and variety into account (Stirling, 2007). The resulting diversity values are normalized by calculating z-scores and subsequently the squared terms are calculated. In the following the application of the entropy measure is specified for the different diversity dimensions.

Educational diversity

For educational diversity I take the different educational types of individuals into account. The categories are extracted from RAMS and LISA and based on the International Standard for Educational Classification (ISCED 97). The educational types are considered at the two digit level, resulting in 26 different educational categories (see Appendix A) (Hellerstedt, 2009). Individuals can have experience in multiple educational categories. Therefore, each experience of an individual in a particular educational category is considered a measurement unit (*i*). The category of general high school education is removed from sample, because it does not represent unique educational experiences.

Industrial diversity

The basis for industrial diversity is the different industries individuals have worked in. The industries are grouped into 59 different categories based on the two-digit NACE classification codes (Hellerstedt, 2009). The data ranges from 1985 in RAMS up to the starting year of the venture in LISA. All industry codes are transformed to the 1992 NACE standard to allow for accurate comparison over the years. Similar to educational diversity, the whole career history individuals have worked in is relevant for experience of individuals. Moreover, the length of the experience in a particular industry positively influences the impact of that particular industry on the knowledge, networks and views of the individual. Therefore, I treat every year an individual has worked in a particular industry as a measurement unit (*i*). For example, an individual that has worked 10 years in industry A and 5 years in industry B, appears in the data as 10 units for

industry A and 5 units for industry B. As a result, the entropy measure indicates whether most of the collective years of the NVT are concentrated in limited industries or whether the experience years are spread over multiple industries.

Organizational diversity

For organizational diversity, each different organization where individuals have been employed is a different category, because each organization represents a unique resource base, network and routine set. Similar to industrial diversity, the data ranges from 1985 in RAMS up to the starting year of the venture in LISA. The organizations are identified using their legal identity number. Similar to industrial diversity, the organizational diversity of the NVT is calculated based on the experience years of the individuals in specific organizations.

Occupational diversity

The occupational diversity is based on the previous job functions of individuals in NVTs. The occupational categories used stem from the LISA database and are based on the International Classification of Occupations (SCB, 2015). Unfortunately, the data on occupation is limited from 2001 to 2012. This restricts the availability of historical data on the job functions of individuals. To avoid right-censored bias in the data as a result of differing availability of historical data for ventures founded in different years, I only include the latest occupational status for the individuals of NVTs. Moreover, for 30 percent of the individuals in the selected data, the occupational status is only available at the major level, consisting of one digit. The occupational status for all individuals is analyzed at the one digit level - resulting in nine different occupational categories - to prevent a large amount of missing observations. Individuals are thus assigned to a single occupational category, and therefore the diversity index can be applied using individuals as measurement unit (*i*) (Frenken et al., 1999). Because the diversity value of the NVT is dependent on the number of individuals involved, the diversity value is corrected for the theoretical maximum diversity possible for a particular size (Hellerstedt, 2009). The resulting diversity values lie between 0 (minimal diversity) and 1 (maximal diversity) (see Appendix B).

Control variables

Several control variables are included in this research. First, the founding year of the new venture is added as a variable to adjust for general economic business cycles in the founding year, which may influence the performance of new ventures (Dahl & Reichstein, 2007). Moreover, the results are controlled for initial NVT size, because bigger teams have more power to mobilize resources than smaller teams (Klepper, 2001; Leonard & Sensiper, 1998). I also control for the demographic characteristics of the NVT (Klotz et al., 2013; Williams & O'Reilly, 1998). The mean age of the NVT is included to control for an age effect and the percentage of males in the NVT controls for gender (Chowdhury, 2005). In addition, a control variable is added for the ethnicity of the team, operationalized as the percentage of Swedes within the NVT (Østergaard et al., 2011).

Furthermore, specific types of experience rather than the diversity of individuals' experiences may affect the performance of new ventures (Klepper & Sleeper, 2005). Therefore, I control (1) for the existence of experience in computing science education of the NVT. (2) Prior industry experience is found to be a determinant for success and survival of new ventures (Dahl & Reichstein, 2007; Klepper, 2001). Industry relevant experience can directly be applied in the new venture and may lead to higher quality products (Delmar & Shane, 2006; Klepper, 2001). Thus, I include a variable that indicates the cumulative years of experience of all individuals of

an NVT in the ICT industry. Moreover, (3) the success of the previous employers is a determinant for performance of the new venture, because it indicates better routines and resources (Klepper, 2009). I control for the success of the former employer of the NVT members, measured by the proportion of former organizations that survive after three years. Last (4), I control for specific occupations that are related to ICT because it indicates relevant experience (Agarwal & Echambadi, 2004). The occupational control variable is a dummy based on the existence of experience as physicist & mathematician or computer professional & technician. Table 1 shows a summary of the operationalization.

Table 1: Operationalization table

Variable	Indicator	Calculation of scores	Measurement
		Dependent variables	
Survival	Existence employees & sales after 3 years	1: Sales > 0 & Employees > 0 0: Other cases	Binary
Sales	Sales after 3 years	Log amount of sales (in SEK)	Continuous
		Independent variables	
Educational diversity	Types of education of NVT members	Entropy of experiences	Continuous
Industrial diversity	Previous industries of NVT members	Entropy of experience years	Continuous
Organizational diversity	Previous organizations of NVT members	Entropy of experience years	Continuous
Occupational diversity	Previous functions of NVT members	Entropy of individuals / theoretical maximum	Continuous
		Control variables	
Year	Founding year of the venture	First year of occurrence	Categorical
NVT size	Number of NVT members	Sum of individuals initially involved	Continuous
Age	Age of NVT members	Mean age of individuals	Continuous
Gender	Gender of NVT members	Percentage of males	Continuous
Ethnic	Country of origin of NVT members	Percentage of Swedes	Continuous
Educational experience	Existence of experience computing education in NVT	1: Experience in code 48 0: Other cases	Binary
Industry experience	Years of specific industry experience of NVT members	Sum years of experience of individuals in code 72	Continuous
Organizational experience	Survival of previous employers of NVT members after 3 years	Proportion of employers of individuals that survive	Continuous
Occupational experience	Existence of experience computing profession in NVT	1: Experience in code 21 or 31 0: Other cases	Binary

3.3 Descriptive statistics and correlations

Table 2 presents the mean, standard deviation and correlations of all the studied variables (see Appendix C for additional descriptive statistics). This is based on a total of 891 new ventures that are identified in the software programming industry. Of these new ventures, 500 still survive three years after their initiation. A total of 2364 unique individuals are part of the entrepreneurial teams of the new ventures. In Figure 1, the distribution of cases over the NVT size and starting year of the venture is given. Approximately 50% of the NVTs consist of two persons, where the other 50% is divided over three, four and five person NVTs. In the dispersion over the years, clear plunges in the amount of ventures can be seen around 2002 and 2009. These may result from the negative economic climate following the Dotcom bubble and financial crisis respectively. In Table 2 a few noticeable correlations can be observed. First, the size of the NVT is positively correlated with educational, industrial and organizational diversity with a coefficient of respectively .35, .40 and .37. Moreover, NVT size shows a negative correlation (-0.54) with the squared occupational diversity variable. Second, the organizational and industrial diversity variables also show a significant moderately strong positive correlation of .56. Third, the squared variables of the diversity dimensions show significant negative correlation to their linear counterparts. The correlation is -0.52 for the industrial dimension and .40 for organizational and occupational dimensions.

Table 2: Descriptive statistics and correlations

	Variable	Mean	SD	1	2	3	4	5	6	7	8
1	Survival	0.56	0.50								
2	Sales (log)	14.04	1.92	0.54 ***							
3	Educational Diversity	0.83	0.51	0.05	0.05						
4	Industrial Diversity	1.74	0.43	0.08 *	0.14 ***	0.31 ***					
5	Organizational Diversity	1.76	0.71	0.12 ***	0.25 ***	0.30 ***	0.56 ***				
6	Occupational Diversity	0.54	0.45	-0.04	-0.08 *	0.00	0.11 ***	0.16 ***			
7	Edu Diversity Squared[a]	-	-	0.02	0.05	-0.24 ***	-0.03	-0.07 *	-0.06		
8	Ind Diversity Squared	-	-	-0.02	0.02	-0.11 **	-0.52 ***	-0.27 ***	-0.12 ***	0.08 *	
9	Org Diversity Squared	-	-	-0.08 *	-0.06	-0.12 ***	-0.21 ***	-0.40 ***	-0.16 ***	0.12 ***	0.35 ***
10	Occ Diversity Squared	-	-	-0.07 *	-0.15 ***	-0.20 ***	-0.31 ***	-0.29 ***	-0.40 ***	-0.03	0.06
11	Year	-	1.97	-0.01	-0.02	-0.01	0.00	-0.03	0.05	-0.03	0.00
12	NVT size	2.70	0.95	0.15 ***	0.29 ***	0.35 ***	0.40 ***	0.37 ***	-0.09 **	0.02	-0.01
13	Age	37.10	9.10	0.00	-0.06	-0.03	-0.25 ***	0.17 ***	0.18 ***	-0.08 *	0.07 *
14	Gender	0.80	0.25	0.05	0.14 ***	-0.03	0.10 **	-0.03	-0.13 ***	0.00	-0.03
15	Ethnicity	0.95	0.16	0.06	0.01	0.00	0.18 ***	0.16 ***	0.06	-0.02	-0.21 ***
16	Educational Experience	0.33	0.47	0.02	0.03	0.40 ***	0.13 ***	0.12 ***	-0.11 **	-0.05	-0.05
17	Industrial Experience	8.43	9.61	0.08 *	0.19 ***	0.13 ***	-0.04	0.36 ***	-0.13 ***	-0.04	-0.02
18	Organizational Experience	0.45	0.38	0.02	-0.04	0.07 *	0.05	0.10 **	0.06	0.03	-0.03
19	Occupational Experience	0.21	0.41	0.04	0.08 *	0.05	0.11 ***	0.16 ***	0.32 ***	0.02	-0.08 *

		9	10	11	12	13	14	15	16	17	18
10	Occ Diversity Squared	0.07									
11	Year	0.04	0.02								
12	NVT size	0.04	-0.54 ***	-0.06							
13	Age	-0.19 ***	0.04	0.09 *	-0.17 ***						
14	Gender	0.09 *	-0.01	-0.03	0.13 ***	-0.32 ***					
15	Ethnicity	-0.16 ***	-0.03	-0.13 ***	0.05	0.05	0.00				
16	Educational Experience	-0.03	-0.08 *	0.05	0.21 ***	-0.15 ***	0.11 **	0.06			
17	Industrial Experience	-0.07	-0.07 *	0.05	0.31 ***	0.23 ***	0.04	0.09 **	0.22 ***		
18	Organizational Experience	-0.07 *	0.01	0.15 ***	-0.01	0.08 *	-0.06	0.02	0.04	0.06	
19	Occupational Experience	-0.06	-0.20 ***	0.01	0.11 **	-0.02	0.01	0.02	0.03	0.01	0.02

*Significance codes: $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*', $p < 0.1$ '.'*

[a] The mean and standard deviation are not displayed for the squared diversity variables because they are based on the centered and standardized linear terms

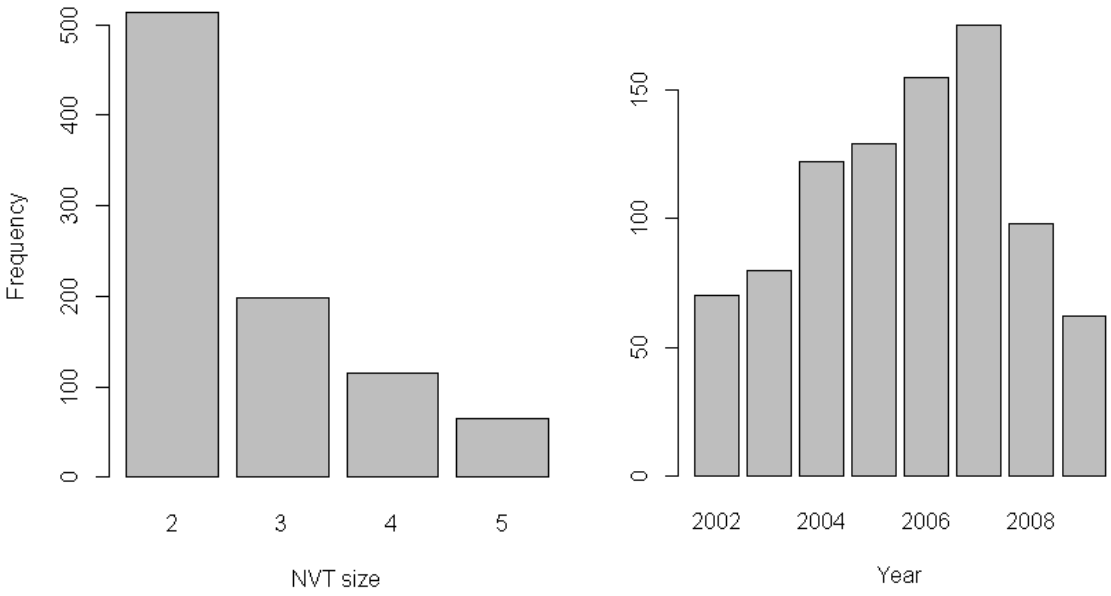


Figure 1: Bar chart of NVT size (left) and Year (right)

3.4 Data analysis

To test the hypotheses formulated in the theory section, I estimate multiple regression analyses. A Binary Logit Model (BLM) is used in case of survival as the dependent variable. An Ordinary Least Squares (OLS) regression is used to analyze the relationship between the NVT diversity and sales. Non-surviving ventures are excluded from the sales models, because they lack sales data and exclusion leads to cleaner results (see Appendix D for alternative analyses). In line with the independent variables, I normalize the control variables and dependent variables as well, with the exception of the dependent binary survival variable and the categorical year variable. Pearson correlations are used to check whether the multicollinearity assumptions are not violated and addition checks are performed by calculating VIF values for the models that are estimated. I estimate three different models for each performance measure: the first model is the control model with only control variables. The second model also includes the linear terms of the diversity dimensions. The third model includes the squared terms of the diversity dimensions next to the linear terms. The latter model is used to test the hypotheses of inverted U-shaped relationships between the different diversity dimensions and performance of new ventures.

4. Results

Table 3 presents the results of the regression models. In the survival models, the McFadden values vary between 0.04 and 0.05. The addition of the diversity dimensions slightly adds to the explanatory power of the model, and is strongest in case of model where squared terms are included as well. However, no significant differences exist between the models. With sales as dependent variable, the control model has an adjusted R^2 of 0.151. The highest adjusted R^2 can be found in the linear model (0.192) and in the squared model the value is slightly lower (0.191). The linear and squared sales models significantly differ from the control model - with $p < 0.01$ - but do not significantly differ from each other.[8] Hence, the diversity dimensions considerably add to the explanatory power of the sales models, but only induce a small insignificant improvement in case of survival models. The VIF values for the regression models are all below 3 and show that the regression assumption of independence between variables is satisfied (Hair et al., 1995).

Educational diversity shows significant results in the sales model, where it has a robust negative relationship to sales. No indications for an inverted U-shaped effect are present in neither the sales nor the survival squared models. Hence, I reject H1. The negative linear effect of educational diversity on sales indicates that teams with overlapping educational experiences perform better and supports the low diversity arguments in literature. The similar experiences of individuals in the educational diversity dimension mainly result from computing and engineering education and to a lesser extent from business & administration. The first two categories are clearly related to the software business, where the latter might indicate higher quality individuals that have done business studies later in their careers.

Industrial diversity does not show any significant effects on either of the performance measures. In contrast to the hypothesis, industrial diversity appears not to influence the performance of new ventures and thus I reject H2.

In case of organizational diversity, one can observe a marginally significant negative squared effect on survival of new ventures. This indicates an inverted U-shaped effect of organizational diversity on survival and is in line with H3. However, for the sales models only a linear positive effect of organizational diversity is found in both the linear and squared model. This supports the high diversity argument made in literature, but is not in line with H3. Hence, H3 is partially confirmed, because there is an inverted U-shaped relationship between organizational diversity and survival, but only a linear positive effect on sales. To get a better insight in the shape of the relationship between organizational diversity and survival, the optimal point of diversity can be calculated based on the regression coefficients. The maximum is reached at a normalized diversity value of 0.4. The right-centered maximum indicates that higher organizational diversity is in general preferable over low organizational diversity.

[8] The adj. R^2 of the squared model is suppressed by the inclusion of insignificant squared terms of educational, industrial and organizational diversity. If only the squared occupational diversity is included, the adj. R2 of the squared sales model is .196 and significantly higher ($p < 0.10$) than the linear model.

Table 3: Regression models results with sales and survival as dependent variables

	Survival			Sales (log)		
	Control	**Linear**	**Squared**	**Control**	**Linear**	**Squared**
Intercept[a]	-0.119	-0.167	-0.184	-0.127	-0.186	-0.043
Educational Diversity		0.001	0.017		-0.125 *	-0.109 *
Educational Diversity Squared			0.068			0.024
Industrial Diversity		-0.04	-0.01		-0.082	-0.073
Industrial Diversity Squared			0.034			0.008
Organizational Diversity		0.165	0.096		0.276 ***	0.271 ***
Organizational Diversity Squared			-0.112 .			0.004
Occupational Diversity		-0.115	-0.111		-0.092 *	-0.119 *
Occupational Diversity Squared			0.017			-0.173 .
NVT size	0.299 ***	0.262 **	0.274 *	0.185 ***	0.172 **	0.107 .
Age	0.07	0.058	0.047	-0.159 ***	-0.192 ***	-0.193 ***
Gender	0.099	0.096	0.102	0.153 ***	0.143 **	0.147 **
Ethnicity	0.083	0.08	0.076	-0.067	-0.081 .	-0.073 .
Educational Experience	-0.04	-0.048	-0.055	-0.059	-0.008	-0.015
Industrial Experience	0.065	0.003	0.02	0.168 ***	0.065	0.077
Organizational Experience	0.05	0.044	0.037	0.028	0.022	0.024
Occupational Experience	0.043	0.062	0.062	0.042	0.042	0.042
Number of observations	886	886	886	484	484	484
McFadden (Survival) & Adj. R^2 (Sales)	0.040	0.044	0.047	0.151	0.192	0.191

*Significance codes: p < 0.001 '***', p < 0.01 '**', p < 0.05 '*', p < 0.1 '.'*
[a] *Year dummies are included in all the regression models as well*

For occupational diversity no significant relationships are found with survival. In case of the linear sales model, a significant negative relationship is found between occupational diversity and sales. However, when the squared term is added as well an inverted U-shaped relationship appears, as reflected by the negative squared occupational diversity. The optimal point of occupational diversity based on the normalized regression coefficients is slightly left-centered with a value of -0.72. Hence, NVTs with some degree of overlapping occupational experiences outperform other NVTs. The overlap between individuals' occupational experiences stems for 80 percent from the management category. However, this category is not fully homogeneous, because different sorts of managers (i.e. financial, operational and general) are included. Other categories with high occurrence of individuals are clerks, which mainly includes financial and accounting functions, service workers and technicians & professionals. Based on the significant findings for sales and insignificant findings for survival from the regression analyses, I partially confirm H4 on the inverted U-shaped relationship between occupational diversity and performance.

When looking at the explanatory power of the diversity dimensions, organizational diversity has the most prominent role. In the survival models it is the only diversity dimension showing significance, and in the sales models the normalized coefficient of organizational diversity is highest (.27). Educational and occupational diversity both have a considerably lower impact on the sales of new ventures, as reflected by their smaller coefficients. As stated before, the industrial dimension does not appear to affect the performance of new ventures, and hence has the lowest relative importance of the diversity dimensions.

Control variables

NVT size has a positive significant effect on the survival of new ventures, which is consistent over the different models. The remainder of the control variables does not show significant relationships with the survival of new ventures. In case of the sales models, five control variables show significance. Again NVT size has a positive effect on performance, but the relationship is not robust over the different models. In squared sales model a large portion of the effect disappears along with the significance. Hence, a large part of the size effect of NVTs on sales is captured by the squared diversity dimensions. Robust results are found for the relationship between the mean age and gender of the NVT and sales. Higher average age is found to have a negative effect on the sales of new ventures and the gender variable is positively related to sales. Hence, male dominated and younger NVTs have a propensity towards higher sales in the new venture. For the ethnicity control variable, the relationship to sales is marginally significant in the linear and squared models. Ventures that have NVTs with a lower percentage of native Swedes may perform better in terms of sales. For industrial experience, the effect is significantly positive in the control model, but disappears when diversity dimensions are added. The other control variables for specific experience do not seem to have any significant relationship to performance.

5. Discussion

The aim of this research was to study the relationship between NVT diversity and new venture performance. Multiple dimensions - educational, industrial, organizational and occupational - of diversity are included and longer career histories of individuals are considered than in existing empirical research on diversity. The results, based on the population of new ventures in Sweden involved in computer programming activities, confirm the effect of diversity of NVTs on performance of new ventures. By allowing for non-linear effects, this study sheds new light on the conflicting explanations of diversity on performance found in literature. Each diversity dimension has its own distinctive influence on the success of the new ventures, and this influence varies for different performance measures. The goldilocks effect of diversity occasionally applies but cannot uniformly explain the effect of different dimensions.

5.1 Theoretical implications

This research shows that industrial diversity is unrelated to the performance of new ventures. This stands in contrast to existing findings in literature that consider industry experience a key determinant for the performance of ventures (Agarwal & Echambadi, 2004; Hellerstedt, 2009; Klepper, 2001). The findings suggest that relevant experiences for individuals for starting a new venture are not bound to industries, i.e. knowledge and skills gained by individuals are not defined by the industries they have worked in. This view is supported by Fern et al. (2012), who find that entrepreneurs exploit knowledge from their own experience rather than using practices from other organizations in their industry. Other diversity dimensions of NVTs tend to be more accurate at explaining the performance of new ventures than industry diversity. Organizational diversity has the most prominent impact on both performance measures. This offers strong support for inheritance theories which state that individuals inherit experience at the organizational level (Agarwal & Echambadi, 2004; Franco & Filson, 2006; Klepper & Sleeper, 2005). The results show that higher levels of organizational diversity enhance performance rather than low organizational diversity. In contrast to these findings, overlap in organizational experience is considered positively related to performance (i.e. survival & growth) in parts of literature (Klepper, 2001), and is confirmed in a large sample representative study by Timmermans (2010). However, this stylized fact arose from studies that only consider recent (up to five years) career histories of individuals. I show that the inclusion of longer career history poses a challenging perspective to the role of overlap in organizational experiences in literature (Beckman, 2006; Eisenhardt & Schoonhoven, 1990).

To a lesser extent educational and occupational diversity affect sales of new ventures. Educational diversity has a negative impact on sales. Hence, NVT members gain technical understanding (e.g. programing languages) in education which serves as a basis for effective cooperation and communication. The knowledge base gained in education does not serve as a basis for novelty and new combinations, because diverse educational backgrounds do not lead to performance advantages. This implies that education is rather a source of general skills than of unique knowledge and perspectives. Occupational diversity shows a left-centered inverted U-shaped effect to sales. Thus, NVTs that show overlap in occupational background but complement is with other occupations perform best. However, one should be careful in deducing implications for the occupational dimension of diversity due to some data limitations. First, only a limited number of occupational categories are included. Experience in some of them – e.g. agriculture, craft or elementary occupations – may not have any relevance in a software

company. Second, the high level occupational categories used might have endogenous heterogeneity and individuals with the same category might not be as similar in experience.

The effects of the diversity dimensions are substantially different for survival and sales, as is most pronounced for organizational diversity. In essence, measuring the performance of new venture is relative because the goals of the individuals involved might differ (Kamm et al., 1990). NVTs with high organizational diversity might use their broad insights and knowledge to pursue more novel and risky business opportunities (Beckman, 2006). This results in higher hazard rates, but also faster growth when ventures manage to survive because they bring a new and innovative product to the marketplace (Helfat & Lieberman, 2002; Timmermans & Coad, 2013). NVTs that consist of individuals that have worked for the same organization might continue in the same market as their former employer or enter a specific niche (Anton & Yao, 1995). These ventures will most likely have a clearly defined market and value proposition and can built on copied routines (Klepper, 2001). This lowers the risk for the venture, increasing its chances on survival but tempering growth. These differences might also explain the absence of an effect of educational and occupational diversity, because the continuation of existing business might be feasible with less complete teams and cooperation might be based on shared routines rather than shared education.

5.2 Societal implications

Some interesting implications for the composition of NVTs by entrepreneurs, incubators and venture capitalists, mainly applicable to new software ventures, follow from this research. Optimal team configuration seems to be dependent on the goals of the venture, i.e. survival or sales. In case of survival, NVTs should complement diverse with overlapping organizational experiences. The former allows for new combinations of perspectives and resources, where the latter enables effective cooperation. Other diversity dimensions do not significantly impact the survival of new ventures. In case of sales, organizational diversity can be maximized. In addition, NVT members should have an overlap in educational experience, preferably in a study field related to the core business of the new venture. Moreover, NVTs aiming to maximize sales should combine some members with overlapping occupational experience with individuals with a deviant occupational background.

The results also have implications for employee mobility and educational policies. In case of the former, employee mobility between organizations increases the organizational diversity of individuals which positively impacts the performance of new ventures. Note that employee mobility does not have to occur between industries per se, because industrial diversity is not related to new venture performance. Policy makers could, for example, give tax exemptions in the first few years of employment at new organization to incentivize individuals to switch jobs. Second, policy makers should keep the long-lasting impact of education in mind. The results suggest that all members of an NVT should have similar education in a relevant field. Policy makers can stimulate networking activities between individuals with similar education to strengthen personal relationships, which are the basis for NVT formation.

5.3 Limitations

This research has a number of limitations. First, the data used in this research poses challenges to the validity. Official classes are used as an indicator for experiences in the cases of educational, industrial and occupational diversity. However, these classes are constructed for administrative purposes rather than the reflection of individuals' experience. Hence, the data

used is only an indication of the experiences of individuals. Second, the diversity indexes are useful for determining diversity in a broad set of options and comparing different NVTs. However, the diversity of a team is brought back to a single diversity value and is only a simplification of reality (Timmermans & Coad, 2013). Insight in the content of experiences and the distribution of categories within the NVT is lost. The third limitation is generalizability. Only new ventures that are classified as software programming ventures are included in this research. Moreover, the cases selected are subject to the Swedish context. Although the Swedish startup characteristics are found to be similar to other western economies (Andersson & Klepper, 2013), scholars should take care in generalizing findings to regions with other institutional configurations.

5.4 Future research

Future research can further disentangle the sources diversity by looking at other diversity dimensions, for example sort of industries of type of organizations. More qualitative research may offer insights in the specific skills individuals gain by education or occupation. Furthermore, interaction between diversity dimensions is in need of further research as well. Can overlap in certain dimensions satisfy similarity needs and allow for higher diversity in other dimensions? Scholars should take care not to simplify by aggregating multiple dimensions into one diversity index (Harrison & Klein, 2007), but rather compare NVTs with different configurations. This research has made a first attempt to include longer career histories of individuals than before. More research is needed to corroborate the findings from this study, include longer occupational history as well and delve deeper in the mechanisms that play a role. For example, do recent experiences have a more profound effect than older experiences? Other interesting ways to take changes over time into account relate to membership changes in the NVT and further development of the new venture (Vanaelst et al., 2006). Where some initial research is performed on the determinants of changes within the NVT (Chandler et al., 2005; Hellerstedt, 2009; Ucbasaran et al., 2003), its consequences for diversity still remain to be understood. Moreover, over time new ventures tend to develop their own routines which might reduce the impact of diversity of the original NVT. How long does the influence of the NVT on the new venture last (see Bamford, Dean, & McDougall, 2000)? Finally, this research emphasizes that instead of searching for general truths on the effect of diversity, the future challenge for scholars of diversity is to fine-tune diversity mechanisms to specific diversity dimensions.

References

Agarwal, R., & Echambadi, R. (2004). Knowledge transfer through inheritance: Spin-out generation, development, and survival. *… of Management Journal*, *47*(4), 501–522. Retrieved from http://amj.aom.org/content/47/4/501.short

Ali, M., Ng, Y. L., & Kulik, C. T. (2013). Board Age and Gender Diversity: A Test of Competing Linear and Curvilinear Predictions. *Journal of Business Ethics*, *125*(3), 497–512. http://doi.org/10.1007/s10551-013-1930-9

Amason, A. C., Shrader, R. C., & Tompson, G. H. (2006). Newness and novelty: Relating top management team composition to new venture performance. *Journal of Business Venturing*, *21*(1), 125–148. http://doi.org/10.1016/j.jbusvent.2005.04.008

Andersson, M., Baltzopoulos, A., & Lööf, H. (2010). R&D strategies and entrepreneurial spawning. *Research Policy*, *41*(228), 54–68. http://doi.org/10.1016/j.respol.2011.08.005

Andersson, M., & Klepper, S. (2013). Characteristics and performance of new firms and spinoffs in Sweden. *Industrial and Corporate Change*, *22*(1), 245–280. http://doi.org/10.1093/icc/dts046

Anton, J., & Yao, D. (1995). Start-ups, spin-offs, and internal projects. *Journal of Law, Economics, & Organization*. Retrieved from http://www.jstor.org/stable/765002

Bamford, C. E., Dean, T. J., & McDougall, P. P. (2000). An examination of the impact of initial founding conditions and decisions upon the performance of new bank start-ups. *Journal of Business Venturing*, *15*(3), 253–277. http://doi.org/10.1016/S0883-9026(98)00011-1

Beckman, C. M. (2006). The influence of founding team company affiliations on firm behavior. *Academy of Management Journal*, *49*(4), 741–758. http://doi.org/10.5465/AMJ.2006.22083030

Beckman, C. M., Burton, M. D., & O'Reilly, C. (2007). Early teams: The impact of team demography on VC financing and going public. *Journal of Business Venturing*, *22*(2), 147–173. http://doi.org/10.1016/j.jbusvent.2006.02.001

Brüderl, J., Preisendörfer, P., & Ziegler, R. (1992). Survival chances of newly founded business organizations. *American Sociological Review*, *57*(2), 227–242. Retrieved from http://www.jstor.org/stable/2096207

Brush, C., & Vanderwerf, P. (1992). A comparison of methods and sources for obtaining estimates of new venture performance. *Journal of Business Venturing*, 157–170. Retrieved from http://www.sciencedirect.com/science/article/pii/0883902692900100

Burt, R. (2004). Structural holes and good ideas. *American Journal of Sociology*, *110*(2), 349–399. Retrieved from http://www.jstor.org/stable/10.1086/421787

Burton, M. D., Sørensen, J. B., & Beckman, C. M. (2002). Coming from good stock: Career histories and new venture formation. *Research in the Sociology of Organizations Volume 19*, 229–262. http://doi.org/10.1016/S0733-558X(02)19007-0

Cantner, U., Goethner, M., & Stuetzer, M. (2010). Disentangling the effects of new venture team functional heterogeneity on new venture performance, *No. 2010,0*(Jena economic research papers). Retrieved from http://www.econstor.eu/handle/10419/32628

Chandler, G., Honig, B., & Wiklund, J. (2005). Antecedents, moderators, and performance consequences of membership change in new venture teams. *Journal of Business Venturing*, *20*(5), 705–725. http://doi.org/10.1016/j.jbusvent.2004.09.001

Chandler, G., & Jansen, E. (1992). The founder's self-assessed competence and venture performance. *Journal of Business Venturing*, 223–236. Retrieved from http://www.sciencedirect.com/science/article/pii/088390269290028P

Chatterji, A. (2009). Spawned with a silver spoon? Entrepreneurial performance and innovation in the medical device industry. *Strategic Management Journal*, *206*(September 2008), 185–206. http://doi.org/10.1002/smj

Chi, N. W., Huang, Y. M., & Lin, S. C. (2009). A Double-Edged Sword? Exploring the Curvilinear Relationship Between Organizational Tenure Diversity and Team Innovation: The Moderating Role of Team-Oriented HR Practices. *Group & Organization Management*, *34*(6), 698–726. http://doi.org/10.1177/1059601109350985

Chowdhury, S. (2005). Demographic diversity for building an effective entrepreneurial team: Is it important? *Journal of Business Venturing*, *20*(6), 727–746. http://doi.org/10.1016/j.jbusvent.2004.07.001

Clarysse, B., & Moray, N. (2004). A process study of entrepreneurial team formation: The case of a research-based spin-off. *Journal of Business Venturing*, *19*(1), 55–79. http://doi.org/10.1016/S0883-9026(02)00113-1

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity : A New Perspective on and Innovation Learning. *Administrative Science Quarterly*, *35*(1), 128–152. Retrieved from http://www.jstor.org/stable/2393553

Colombo, M. G., & Grilli, L. (2005). Founders' human capital and the growth of new technology-based firms: A competence-based view. *Research Policy*, *34*(6), 795–816. http://doi.org/10.1016/j.respol.2005.03.010

Cooper, A., Gimeno-Gascon, F., & Woo, C. (1994). Initial human and financial capital as predictors of new venture performance. *Journal of Business Venturing*, (August 1991), 371–395. Retrieved from http://www.sciencedirect.com/science/article/pii/0883902694900132

Dahl, M. S., & Reichstein, T. (2007). Are You Experienced? Prior Experience and the Survival of New Organizations. *Industry & Innovation*, *14*(5), 497–511. http://doi.org/10.1080/13662710701711414

Davis, C. H., & Schaefer, N. V. (2003). Development Dynamics of a Startup Innovation Cluster : the ICT Sector in New Brunswick. In *Clusters old and new: the transition to a knowledge economy in Canada's regions* (pp. 1–64).

Delmar, F., & Shane, S. (2006). Does experience matter? The effect of founding team experience on the survival and sales of newly founded ventures. *Strategic Organization*, *4*(3), 215–247. http://doi.org/10.1177/1476127006066596

Dencker, J. C., Gruber, M., & Shah, S. K. (2009). Pre-Entry Knowledge, Learning, and the Survival of New Firms. *Organization Science*, *20*(3), 516–537. http://doi.org/10.1287/orsc.1080.0387

Eisenhardt, K. M., & Schoonhoven, C. B. (1990). Growth : Organizational Linking Founding Team , Strategy , Environment , and Growth among U . S . Semiconductor Claudia Bird Schoonhoven. *Administrative Science Quarterly*, *35*(3), 504–529.

Ensley, M. D., Carland, J. W., & Carland, J. A. (1999). The effect of entrepreneurial team skill heterogeneity and functional diversity on new venture performance. *Journal of Business and Entrepreneurship*, *10.1*(July), 1–19.

Ensley, M. D., Pearson, A. W., & Amason, A. C. (2002). Understanding the dynamics of new venture top management teams: cohesion, conflict, and new venture performance. *Journal of Business Venturing*, *17*(4), 365–386. http://doi.org/10.1016/S0883-9026(00)00065-3

Eriksson, T., & Kuhn, J. M. (2006). Firm spin-offs in Denmark 1981–2000—patterns of entry and exit. *International Journal of Industrial Organization*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167718706000038

Fern, M. J., Cardinal, L. B., & O'Neill, H. M. (2012). The genesis of strategy in new ventures: Escaping the constraints of founder and team knowledge. *Strategic Management Journal*, *33*(4), 427–447. http://doi.org/10.1002/smj.1944

Franco, A., & Filson, D. (2006). Spin-outs: knowledge diffusion through employee mobility. *The RAND Journal of Economics*, *37*(4), 841–860. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1756-2171.2006.tb00060.x/abstract

Frederiksen, L., & Wennberg, K. (2011). Mobility and Entrepreneurship: How career history influences entrepreneurial entry and survival. *DRUID 2011*. Retrieved from http://druid8.sit.aau.dk/acc_papers/335av3xp0ym18ndcifo01hn3ynq3.pdf

Frenken, K., Saviotti, P. P., & Trommetter, M. (1999). Variety and niche creation in aircraft, helicopters, motorcycles and microcomputers. *Research Policy*, *28*(5), 469–488. http://doi.org/10.1016/S0048-7333(99)00008-6

Furlan, A., & Grandinetti, R. (2014). Spin-off performance in the start-up phase – a conceptual framework. *Journal of Small Business and Enterprise Development*, *21*(3), 528–544. http://doi.org/10.1108/JSBED-04-2014-0055

Hair, J. F., Black, B., Anderson, R., & Tatham, R. (1995). *Multivariate Data Analysis: Text and Readings.* MacMillan, International Edition. Retrieved from http://works.bepress.com/joe_hair/188/

Harrison, D. A., & Klein, K. J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review*, *32*(4), 1199–1228. http://doi.org/10.5465/AMR.2007.26586096

Helfat, C., & Lieberman, M. (2002). The birth of capabilities: market entry and the importance of pre-history. *Industrial and Corporate Change*, *11*(4), 725–760. Retrieved from http://icc.oxfordjournals.org/content/11/4/725.short

Hellerstedt, K. (2009). The Composition of New Venture Teams. *JIBS Dissertation Series*, *No. 056*.

Horwitz, S. K., & Horwitz, I. B. (2007). The Effects of Team Diversity on Team Outcomes: A Meta-Analytic Review of Team Demography. *Journal of Management*, *33*(6), 987–1015. http://doi.org/10.1177/0149206307308587

Jackson, S. E., Joshi, A., & Erhardt, N. L. (2003). Recent Research on Team and Organizational Diversity: SWOT Analysis and Implications. *Journal of Management*, *29*(6), 801–830. http://doi.org/10.1016/S0149-2063(03)00080-1

Jones-Evans, D. (1996). Experience and entrepreneurship: Technology-based owner-managers in the UK. *New Technology, Work and Employment*, *11*, 39–54. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-0029906256&partnerID=tZOtx3y1

Kakarika, M. (2013). Staffing an entrepreneurial team: diversity breeds success. *Journal of Business Strategy*, *34*(4), 31–38. http://doi.org/10.1108/JBS-06-2012-0020

Kakarika, M., & Biniari, M. G. (2011). *Unfolding the role of team diversity in the entrepreneurial process*. *4th Annual EuroMed Conference of the EuroMed Academy of Business Business*.

Kamm, J. B., Shuman, J. C., Seeger, J. a, & Nurick, A. J. (1990). Entrepreneurial teams in new venture creation: A research agenda. *Entrepreneurship: Theory and Practice*, *14*(4), 7–17. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=9603274990&site=ehost-live

Klepper, S. (2001). Employee startups in high-tech industries. *Industrial and Corporate Change*, 639–674. Retrieved from http://icc.oxfordjournals.org/content/10/3/639.short

Klepper, S. (2009). Spinoffs: A review and synthesis. *European Management Review*, *6*(3), 159–171. http://doi.org/10.1057/emr.2009.18

Klepper, S., & Sleeper, S. (2005). Entry by spinoffs. *Management Science*, *51*(8), 1291–1306. Retrieved from http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1050.0411

Klotz, C., Hmieleski, K. M., Bradley, B. H., & Busenitz, L. W. (2013). New Venture Teams: A Review of the Literature and Roadmap for Future Research. *Journal of Management*, *40*(1), 226–255. http://doi.org/10.1177/0149206313493325

Lasch, F., Le Roy, F., & Yami, S. (2007). Critical growth factors of ICT start-ups. *Management Decision*, *45*(1), 62–75. http://doi.org/10.1108/00251740710718962

Leary, M. M., & DeVaughn, M. L. (2009). Entrepreneurial team characteristics that influence the successful launch of a new venture. *Management Research News*, *32*(6), 567–579. http://doi.org/10.1108/01409170910962993

Leonard, D., & Sensiper, S. (1998). The Role of Tacit Knowledge in Group Innovation. *California Management Review*, *40*(3), 112–132. http://doi.org/10.2307/41165946

Martinez, M. a., & Aldrich, H. E. (2011). *Networking strategies for entrepreneurs: balancing cohesion and diversity*. *International Journal of Entrepreneurial Behavior & Research* (Vol. 17). http://doi.org/10.1108/13552551111107499

Melillo, F., Folta, T., & Delmar, F. (2012). What determines the initial size of new ventures. *DRUID Celebration Conference 2013*, *35th*(2). Retrieved from http://druid8.sit.aau.dk/druid/acc_papers/f1sscrlqbgj0fkbk4rmxm8vlntj6.pdf

Milliken, F., & Martins, L. (1996). Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *Academy of Management Review*, *21*(2), 402–433. Retrieved from http://amr.aom.org/content/21/2/402.short

Murphy, G. B., Trailer, J. W., & Hill, R. C. (1996). Measuring performance in entrepreneurship research. *Journal of Business Research*, *36*(1), 15–23. http://doi.org/10.1016/0148-2963(95)00159-X

Østergaard, C. R., Timmermans, B., & Kristinsson, K. (2011). Does a different view create something new? The effect of employee diversity on innovation. *Research Policy*, *40*(3), 500–509. http://doi.org/10.1016/j.respol.2010.11.004

Reich, R. B. (1987). Entrepreneurship reconsidered: the team as a hero. *Harvard Business Review*, *65*(3), 77–83. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bsh&AN=4127975&site=ehost-live

Roure, J. B., & Maidique, M. a. (1986). Linking prefunding factors and high-technology venture success: An exploratory study. *Journal of Business Venturing*, *1*, 295–306. http://doi.org/10.1016/0883-9026(86)90006-6

Ruef, M. (2002). A structural event approach to the analysis of group composition. *Social Networks*, *24*(2), 135–160. http://doi.org/10.1016/S0378-8733(01)00054-5

Ruef, M., Aldrich, H. E., & Carter, N. M. (2003). The Structure of Founding Teams: Homophily, Strong Ties, and Isolation among U.S. Entrepreneurs. *American Sociological Review*, *68*(2), 195. http://doi.org/10.2307/1519766

SCB. (2015). MONA documentation. Retrieved from http://www.scb.se/sv_/Dokumentation/Dokumentationer/

Schlosser, F. (2013). Identifying and Differentiating Key Employees from Owners and Other Employees in SMEs. *Journal of Small Business Management*, *53*(1), 37–53. http://doi.org/10.1111/jsbm.12066

Shane, S. (2000). Prior knowledge and the discovery of entrepreneurial opportunities. *Organization Science*, *11*(4), 448–469. Retrieved from http://pubsonline.informs.org/doi/abs/10.1287/orsc.11.4.448.14602

Shane, S., & Khurana, R. (2003). Bringing individuals back in: the effects of career experience on new firm founding. *Industrial and Corporate Change*, *12*(3), 519–543. Retrieved from http://icc.oxfordjournals.org/content/12/3/519.short

Stam, W., & Elfring, T. (2008). Entrepreneurial orientation and new venture performance: The moderating role of intra-and extraindustry social capital. *Academy of Management Journal*, *51*(1), 97–111. http://doi.org/10.5465/AMJ.2008.30744031

Steffens, P., Terjesen, S., & Davidsson, P. (2012). Birds of a feather get lost together: New venture team composition and performance. *Small Business Economics*, *39*(3), 727–743. http://doi.org/10.1007/s11187-011-9358-z

Stinchcombe, A. (1965). Social structure and organizations. *Advances in Strategic Management, Volume 17*, pp.229 – 259. Retrieved from http://www.emeraldinsight.com/doi/pdf/10.1016/S0742-3322(00)17019-6

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society, Interface / the Royal Society*, *4*(15), 707–19. http://doi.org/10.1098/rsif.2007.0213

Talaulicar, T., Grundei, J., & Werder, A. V. (2005). Strategic decision making in start-ups: The effect of top management team organization and processes on speed and comprehensiveness. *Journal of Business Venturing*, *20*(4), 519–541. http://doi.org/10.1016/j.jbusvent.2004.02.001

Timmermans, B. (2010). The Effect of Prior Joint Work Experience on New Venture Performance, 1–26.

Timmermans, B., & Coad, A. (2013). Two's Company: Composition and Performance of Entrepreneurial Pairs. *35th DRUID Celebration Conference 2013, Barcelona, Spain, June 17-19*, *11*, 1–23. http://doi.org/10.2139/ssrn.2169547

Ucbasaran, D., Lockett, A., Wright, M., & Westhead, P. (2003). Entrepreneurial founder teams : factors associated with member entry and exit. *Entrepreneurship Theory and Practice*, *28*(2), 107–128. http://doi.org/10.1046/j.1540-6520.2003.00034.x

Van Knippenberg, D., De Dreu, C. K. W., & Homan, A. C. (2004). Work group diversity and group performance: an integrative model and research agenda. *The Journal of Applied Psychology*, *89*(6), 1008–22. http://doi.org/10.1037/0021-9010.89.6.1008

Van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annual Review of Psychology*, *58*, 515–41. http://doi.org/10.1146/annurev.psych.58.110405.085546

Vanaelst, I., Clarysse, B., Wright, M., Lockett, A., Moray, N., & S'Jegers, R. (2006). Entrepreneurial team development in academic spinouts: An examination of team heterogeneity. *Entrepreneurship: Theory and Practice*, *30*, 249–271. http://doi.org/10.1111/j.1540-6520.2006.00120.x

Vegt, G. S. Van Der, & Bunderson, J. S. (2005). Learning and performance in multidisciplinary teams: the importance of collective team identification, *48*(3), 532–547.

Webber, S. S., & Donahue, L. M. (2001). Impact of highly and less job-related diversity on work group cohesion and performance: a meta-analysis. *Journal of Management*, *27*(2), 141–162. http://doi.org/10.1177/014920630102700202

Williams, K., & O'Reilly, C. (1998). Demography and diversity in organizations: A review of 40 years of research. *Research in Organizational Behavior*. Retrieved from http://ils.unc.edu/courses/2013_spring/inls285_001/materials/WIlliams.OReilly.1996.Diversity%26demography.pdf

Wright, M., Hmieleski, K. M., Siegel, D. S., & Ensley, M. D. (2007). The role of human capital in technological entrepreneurship. *Entrepreneurship: Theory and Practice*, *31*(6), 791–806. http://doi.org/10.1111/j.1540-6520.2007.00202.x

Appendix A: List of categories of educational, industrial and occupational dimensions

Educational categories: International Standard for Educational Classification (ISCED 97)

Source: *http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx*

0	**General Programmes**
01	Basic/broad, general programmes
08	Literacy and numeracy
09	Personal skills
1	**Education**
14	Teacher training and education science
2	**Humanities and Arts**
21	Arts
22	Humanities
3	**Social sciences, Business and Law**
31	Social and behavioural science
32	Journalism and information
34	Business and administration
38	Law
4	**Science, Mathematics and Computing**
42	Life science
44	Physical science
46	Mathematics and statistics
48	Computing
5	**Engineering, Manufacturing and Construction**
52	Engineering and engineering trades
54	Manufacturing and processing
58	Architecture and building
6	**Agriculture and Veterinary**
62	Agriculture, forestry and fishery
64	Veterinary
7	**Health and Welfare**
72	Health
76	Social services
8	**Services**
81	Personal services
84	Transport services
85	Environmental protection
86	Security services
99	**Not known or unspecified**

Industrial categories: NACE industry code (2002)

Source: http://www.scb.se/SNI2007/

Division (Official code structure)	Description
01	Agriculture, hunting and related service activities
02	Forestry, logging and related service activities
05	Fishing, operation of fish hatcheries and fish farms; service activities incidental to fishing
10	Mining of coal and lignite; extraction of peat
11	Extraction of crude petroleum and natural gas; service activities incidental to oil and gas extraction excluding surveying
12	Mining of uranium and thorium ores
13	Mining of metal ores
14	Other mining and quarrying
15	Manufacture of food products and beverages
16	Manufacture of tobacco products
17	Manufacture of textiles
18	Manufacture of wearing apparel; dressing and dyeing of fur
19	Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear
20	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
21	Manufacture of pulp, paper and paper products
22	Publishing, printing and reproduction of recorded media
23	Manufacture of coke, refined petroleum products and nuclear fuel
24	Manufacture of chemicals and chemical products
25	Manufacture of rubber and plastic products
26	Manufacture of other non-metallic mineral products
27	Manufacture of basic metals
28	Manufacture of fabricated metal products, except machinery and equipment
29	Manufacture of machinery and equipment n.e.c.
30	Manufacture of office machinery and computers
31	Manufacture of electrical machinery and apparatus n.e.c.
32	Manufacture of radio, television and communication equipment and apparatus
33	Manufacture of medical, precision and optical instruments, watches and clocks
34	Manufacture of motor vehicles, trailers and semi-trailers
35	Manufacture of other transport equipment
36	Manufacture of furniture; manufacturing n.e.c.
37	Recycling
40	Electricity, gas, steam and hot water supply
41	Collection, purification and distribution of water
45	Construction
50	Sale, maintenance and repair of motor vehicles and motorcycles; retail sale of automotive fuel
51	Wholesale trade and commission trade, except of motor vehicles and motorcycles
52	Retail trade, except of motor vehicles and motorcycles; repair of personal and household goods
55	Hotels and restaurants

60 Land transport; transport via pipelines
61 Water transport
62 Air transport
63 Supporting and auxiliary transport activities; activities of travel agencies
64 Post and telecommunications
65 Financial intermediation, except insurance and pension funding
66 Insurance and pension funding, except compulsory social security
67 Activities auxiliary to financial intermediation
70 Real estate activities
71 Renting of machinery and equipment without operator and of personal and household goods
72 Computer and related activities
73 Research and development
74 Other business activities
75 Public administration and defence; compulsory social security
80 Education
85 Health and social work
90 Sewage and refuse disposal, sanitation and similar activities
91 Activities of membership organizations n.e.c.
92 Recreational, cultural and sporting activities
93 Other service activities
95 Activities of households as employers of domestic staff
99 Extra-territorial organizations and bodies

Occupational categories: International Classification of Occupation 1997

Source: http://laborsta.ilo.org/applv8/data/isco88e.html; Only the major groups are used as categories to calculate occupational diversity, due to a lack of lower level data

Major Group 1 Legislators, senior officials and managers
11. Legislators and senior officials
12. Corporate managers 1
13. General managers 2
Major Group 2 Professionals
21. Physical, mathematical and engineering science professionals
22. Life science and health professional
23. Teaching professionals
24. Other professionals
Major Group 3 Technicians and associate professionals
31. Physical and engineering science associate professionals
32. Life science and health associate professionals
33. Teaching associate professionals
34. Other associate professionals
Major Group 4 Clerks
41. Office clerks
42. Customer service clerks
Major Group 5 Service workers and shop and market sales workers
51. Personal and protective services workers

52. Models, salespersons and demonstrators

Major Group 6 Skilled agricultural and fishery workers

61. Market-oriented skilled agricultural and fishery workers
62. Subsistence agricultural and fishery workers

Major Group 7 Craft and related trade workers

71. Extraction and building trade workers
72. Metal, machinery and related trades workers
73. Precision, handicraft, printing and related trades workers
74. Other craft and related trades workers

Major Group 8 Plant and machine operators and assemblers

81. Stationary plant and related operators
82. Machine operators and assemblers
83. Drivers and mobile plant operators

Major Group 9 Elementary occupations

91. Sales and services elementary occupations
92. Agricultural, fishery and related labourers
93. Labourers in mining, construction, manufacturing and transport

Major Group 0 Armed forces

01 Armed forces

Appendix B: Methodological details

The name of final variables constructed in the dataset is given between [brackets]

Data restriction

Step	Database	Variable used	Description
Year selection [Year]	All databases	Year indication of database	Only select ventures from years 2002 (year after start occupational data) and 2009 (the dataset stops at 2012, I need performance measure after 3 years).
New firms	FAD	KVar	New ventures are identified by the Kvar codes: 01, 02 and 03
New establishment	FAD	AKvar	New establishments are identified by the AKvar codes: 01, 02 and 03
Max establishments	FAD	Count AKvar	Based on the count of the number of AKvar codes, only ventures with 1 are selected
Legal number	FAD	LopNr_PeOrgNr1 &LopNr_PeOrgNr	If x1 is the same as normal legal code, than the venture has same legal number in subsequent years and is removed
Double firms	FAD	LopNr_PeOrgNr	If legal number occurs more than once in different years, venture is removed
Industry	FAD	AstSNI2002/92	Select 72.210 in SNI2002 and 72.202 in SNI92. This subset involves the programming of software.
Shareholders	FRIDA	PersonLopNr	Connect individual codes to legal code of the venture
Employees	FAD	PersonLopNr	Connect individual codes to legal code of the venture
NVT size [NVT]	-	PersonLopNr	Count the unique legal individual number that are involved in the venture -> NVT
NVT selection	-	NVT	Select only ventures with 2 – 5 individuals involved

Dependent variables

New ventures are followed over time by identifier: LopNr_PeOrgNr (legal number)

Step	Database	Variable used	Description
Employees [EmpY3]	FAD	T2	Take the amount of employees at t+3
Sales [Sales]	FEK	Omsättning	Take the amount of Sales in SEK at t+3 -> Sales
Exit [Exit]	FAD	Kvar	Exit code "10" is used to identify firms take experienced a takeover -> Exit
Survival [Sur]	-	Sales, Exit, EmpY3	1: IF (EmpY3 > 0 & Sales > 0) OR Exit = "10" 0: Other cases
Saleslog [Saleslog]	-	Sales	Take the log of the sales because the variable is skewed

Independent variables

Step	Database	Variable used	Description
General method	• See appendix A for the list of categories • Educational and Industrial categories are transformed to the 2-digit level, 1-digit in case of occupational data • The status of the individuals over the years is transformed to a table that shows the number of years of each individual per category • All the years in the categories are transformed to occurrence for educational & occupational • In case of Industrial and Organizational the sum of the years is taken • Duplicate individuals are removed (to prevent shareholders and employees being counted twice) • The data is aggregated of the new venture, now the data is ready for calculating diversity measures		
Educational status	RAMS/LISA	Sun2000Inr	
Industrial status	RAMS/LISA	AstSNI92	AstSNI2002 is transformed to AstSNI92 using a key present in MONA
Organizational status	RAMS/LISA	LopNr_PeOrgNr	
Occupational status	LISA	Ssyk3	

Diversity measures

- To calculate the diversity, the Shannon-Weaver (or Frenken) entropy index is used: $Diversity = \sum_{i=1}^{A} p_i \log\left(\frac{1}{p_i}\right) \qquad (i = 1, \dots, A)$.
- Based on the aggregated information per diversity dimensions, first a total column is calculated. This is needed to calculate the proportions (p) for the entropy index. Finally, the proportions are the input for the calculation of the diversity.
- [EduDivF] is based on the last educational status of each individual.
- [EduDivA] is based on all the educational categories the individuals have experience in
- [OrgDivF] is based on the dispersion of all the years of the individuals over unique organizations
- [IndDivF] is based on the dispersion of all the years of individuals over industrial categories
- [OccDivF] is based on the last occupational status of each individual

Correcting the diversity measures [EduDivF] and [OccDivF]

- The absolute Frenken measures for Educational [EduDivF] and Occupational [OccDivF] are dependent on the size of the NVT because each individual can only be assigned a single value (see table below)
- To correct for this effect, the absolute values are corrected for the theoretical max of the NVT size. First, a dataframe is made that connects the size of the NVT with the theoretical maximum value of the diversity measure. Second, the absolute values are divided by their theoretical max. This results in relative diversity measures [EduDivF2] and [OccDivF2]
- NOT USED IN FINAL THESIS, BUT IN APPENDIXES: In the table below all the theoretical possibilities are given for different configurations of teams with different NVT size. This

is based on an increasing amount of individuals that deviate from each other (see example of team with 3 individuals). Because of the categorical nature of this data, the relative measures are transformed to a categorical variable, namely [EduDivF2C] and [OccDivF2C]. The coding of categories is presented in the table below. The variables are releveled to make category 5 the reference category.

Frenken								
	NVT size	**Min**			**Average**			**Max**
Absolute	2	0						0,69
	3	0			0,63			1,09
	4	0		0,56	0,69	1,03		1,38
	5	0	0,5	0,67	0,95	1,05	1,33	1,6
Relative	2	0						1
	3	0			0,577982			1
	4	0		0,405797	0,5	0,746377		1
	5	0	0,3125	0,41875	0,59375	0,65625	0,83125	1
Categorical	**Categories**	1	2	2	3	4	4	5

Possibilities of team with 3 individuals

	Minimal diversity			Average diversity			Maximal diversity		
	Cat 1	**Cat 2**	**Cat 3**	**Cat 1**	**Cat 2**	**Cat 3**	**Cat 1**	**Cat 2**	**Cat 3**
Indiv. 1	1			1			1		
Indiv. 2	1			1				1	
Indid. 3	1				1				1

Control variables

Step	Database	Variable used	Description
Age [Age]	LISA	Alder	Connect age to all individuals, aggregate over new venture and take the mean age
Gender [Gen]	LISA	Kon	Recode to 1: male; 0: female. Take the mean by aggregating over new venture.
Ethnicity [Eth]	LISA	MedbGrEg2	Make dummy for each individual (1: Sweden, 0: Other). Aggregate by mean over new venture
Educational experience [EduExp]	RAMS/LISA	Sun2000Inr	Make a dummy based on the occurrence of category "48" in the NVT (use transformed data from independent variables)
Industrial experience [IndExp]	RAMS/LISA	AstSNI92	Sum the years of experience of individuals in the team of new venture in category "72"
Organizational experience [OrgExp]	FAD	LopNr_PeOrgNr, T2	Find last employer of each individual, make survival variable for employers based on existence employees (T2) at t+3, aggregate by mean to new venture to obtain proportion of surviving old employers
Occupational experience	LISA	Ssyk3	Make a dummy based on the occurrence of category "21" or "31" in

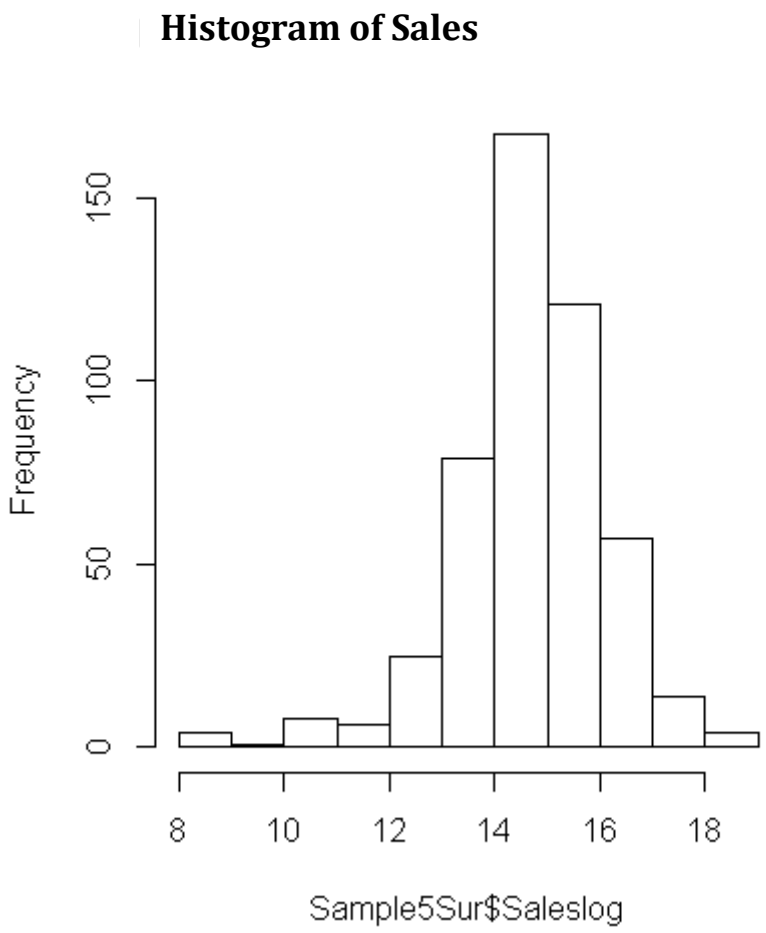
[OccExp]	the NVT (use transformed data from independent variables)

Analyses

- **Missing data:** in case of NA's in the data, remove the case for particular regression analysis (na.action = na.omit)
- **Survival as dependent variable:** use generalized linear model with function "binomial" (function = glm, family = "binomial")
- **Saleslog as dependent variable:** use ordinary least squares (linear regression) (function = lm)
- **Significance of categorical variables:** estimate model without the categorical variable, use Chisq test to determine significance of the model without the variable with the complete model (function = drop1, test = "Chisq")
- **Model fit:** in case of the linear regression R squared adjusted is used; in case of binary logit model, McFadden's model fit is used (function = pR2)
- **Multicollinearity:** correlation matrix is based on Pearson, VIF values are calculated as well (function = rcorr & vif)
- **Two step Heckman:** use survival variable that is not corrected for takeover exits, method used is maximum likelihood

Appendix C: Additional descriptive statistics

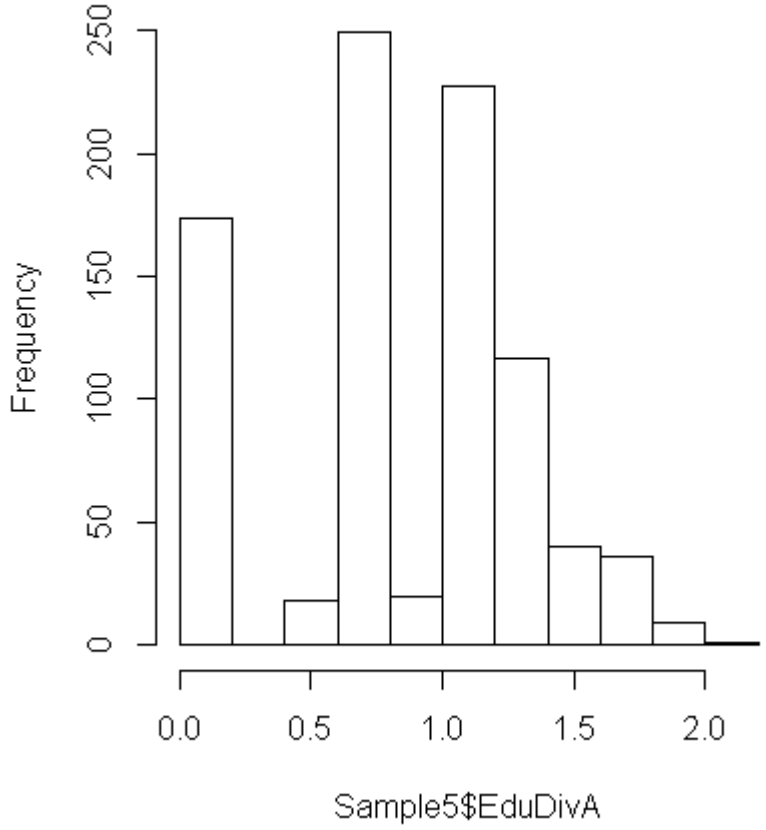
Histograms



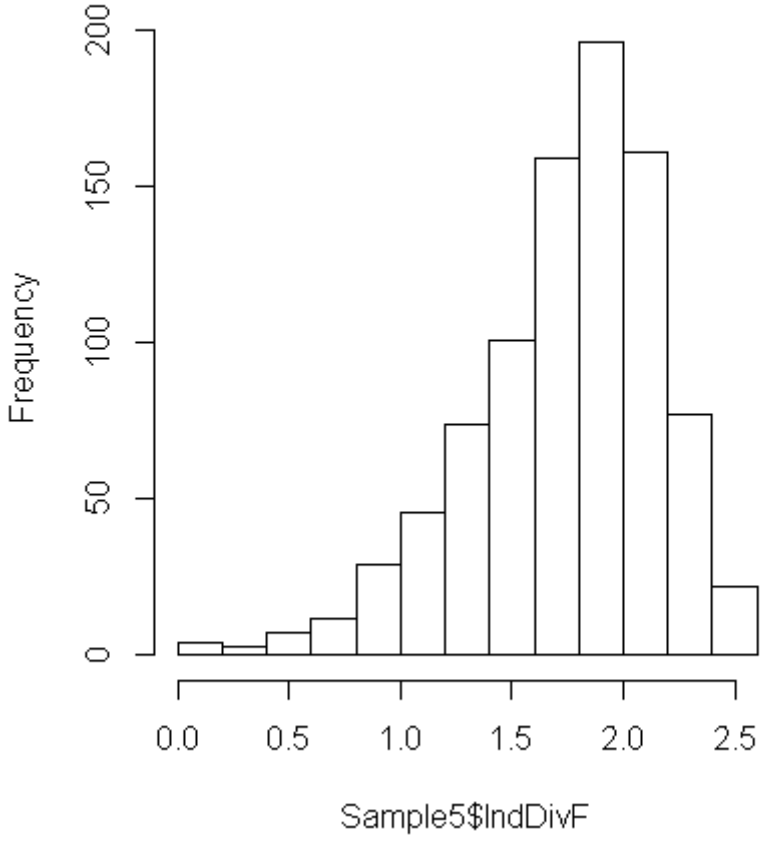
This Appendix includes histograms and scatter plots of all variables used in the regression analyses. In the title of the histograms the reference relates to the terminology in the Methodology and Results, where the bottom reference relates to the construction of variables discussed in Appendix B. The histograms are based on the non-normalized variables. The scatter plots are based on the same underlying variables, but normalized.

Independent variables

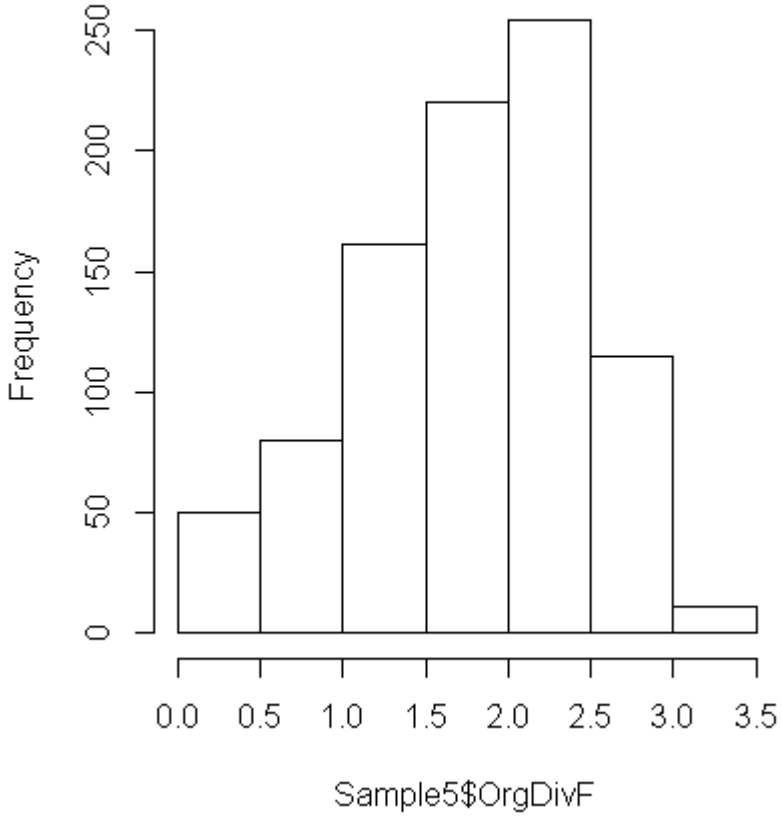
Histogram of Educational diversity



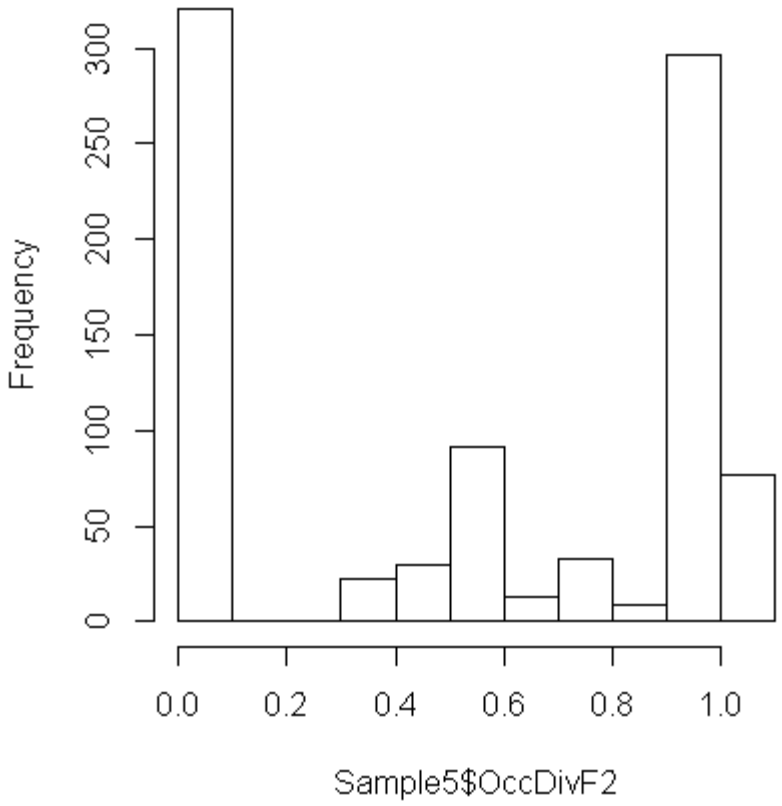
Histogram of Industrial diversity



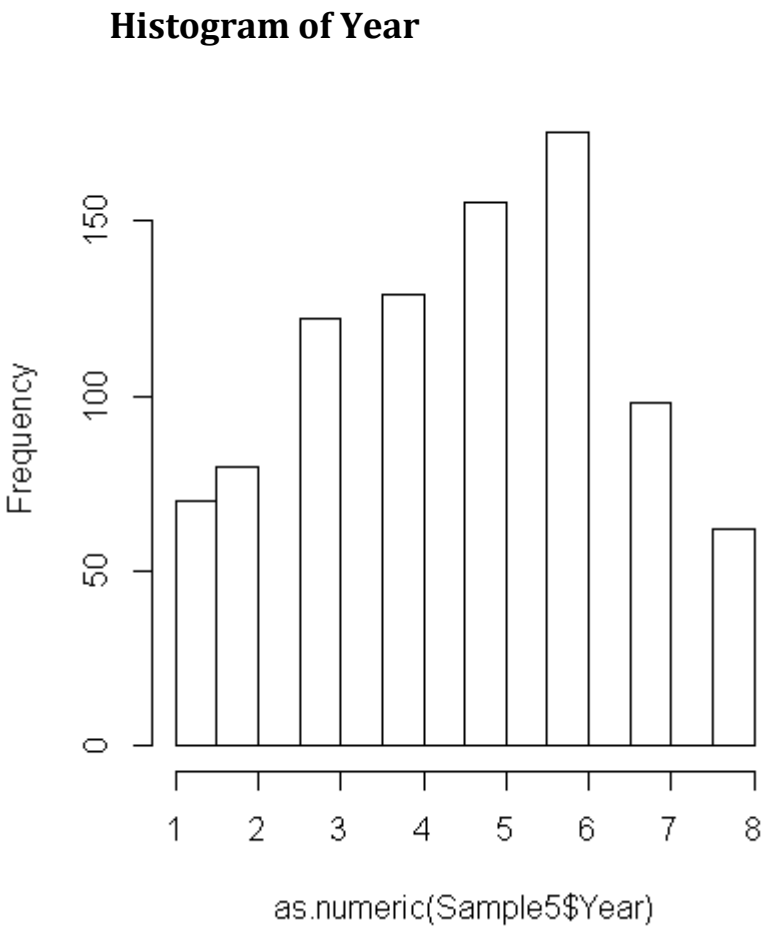
Histogram of Organizational diversity

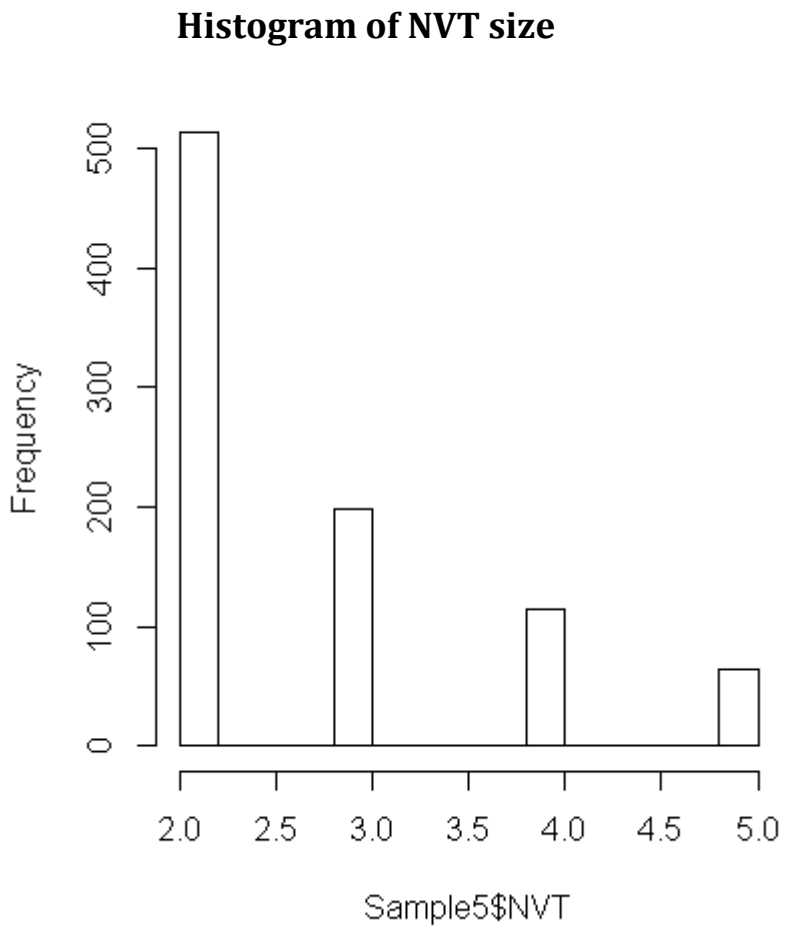


Histogram of Occupational diversity

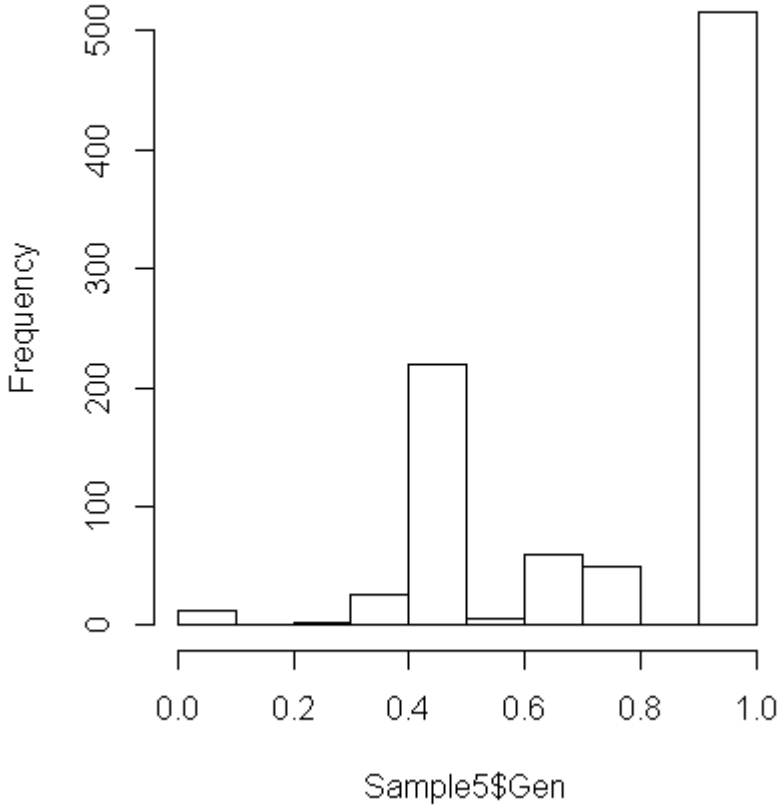


Control variables

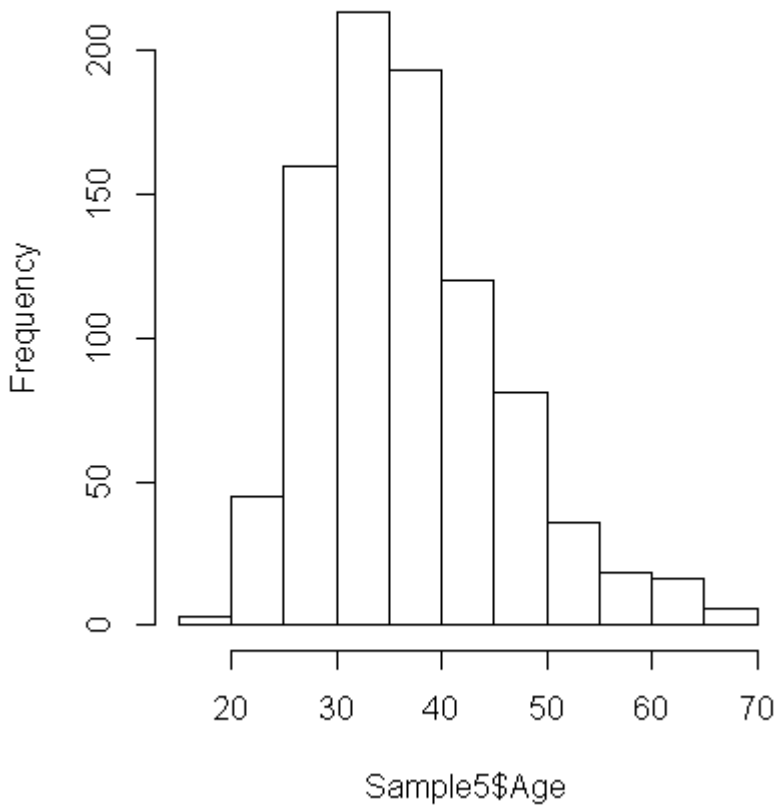




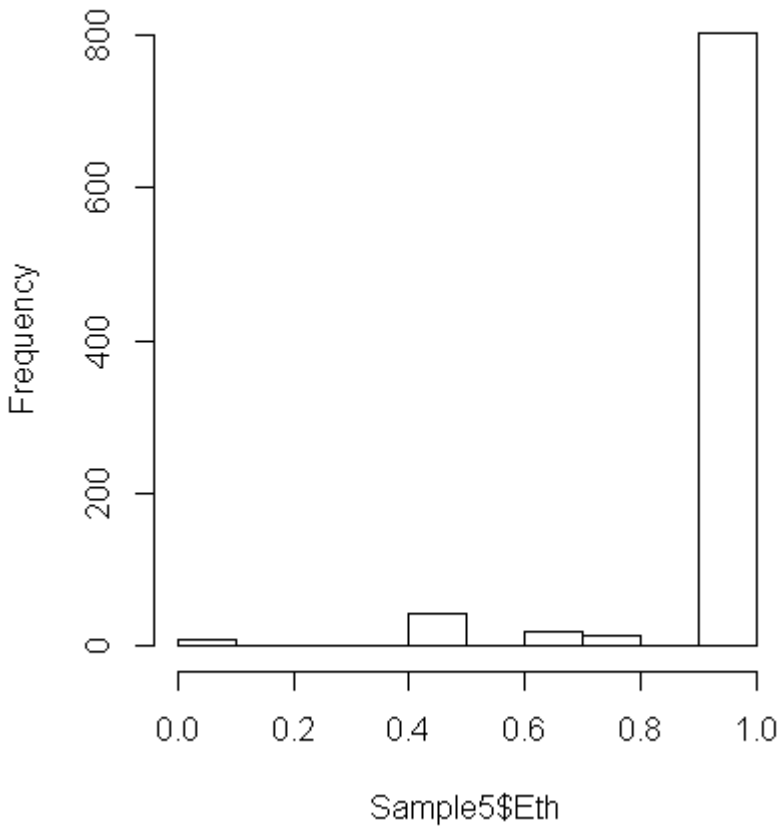
Histogram of Gender



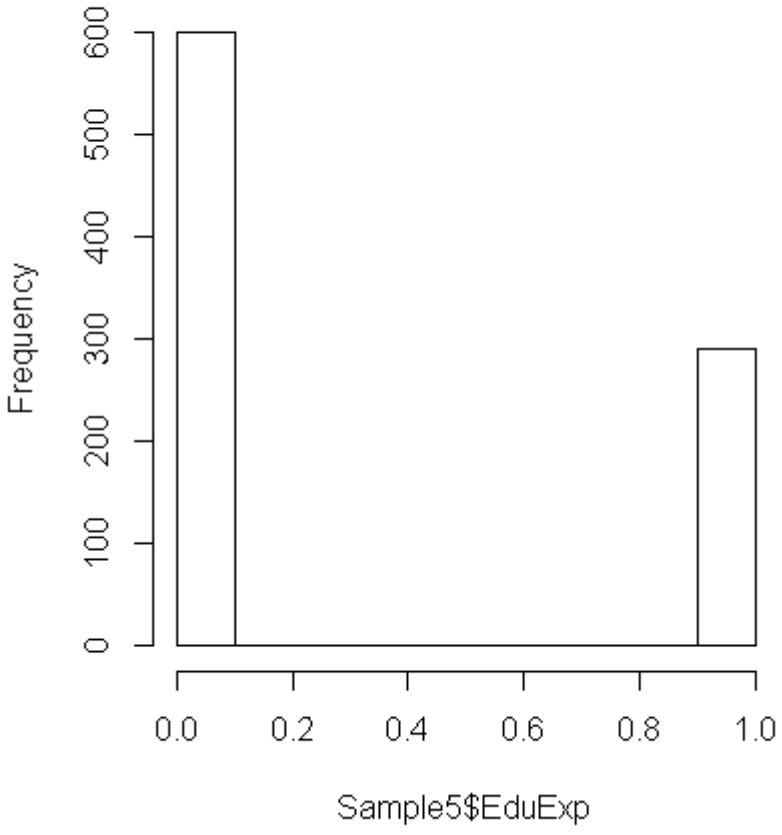
Histogram of Age



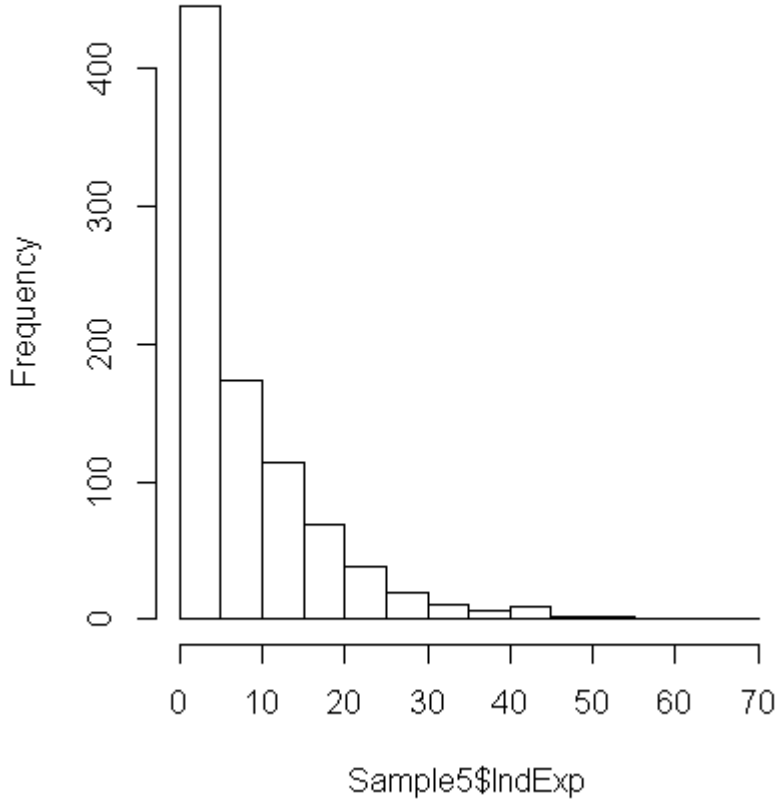
Histogram of Ethnicity



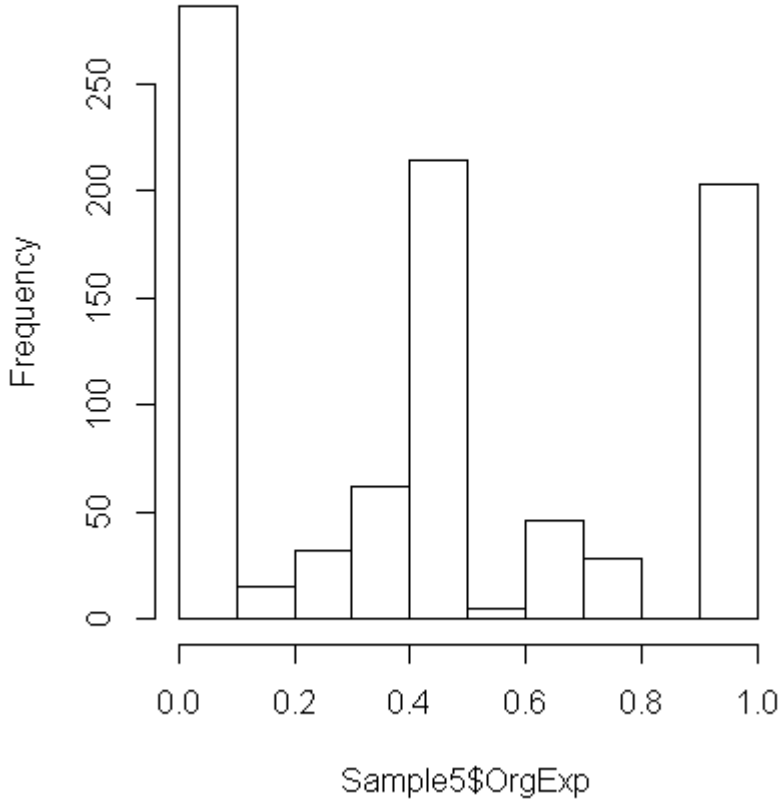
Histogram of Educational experience



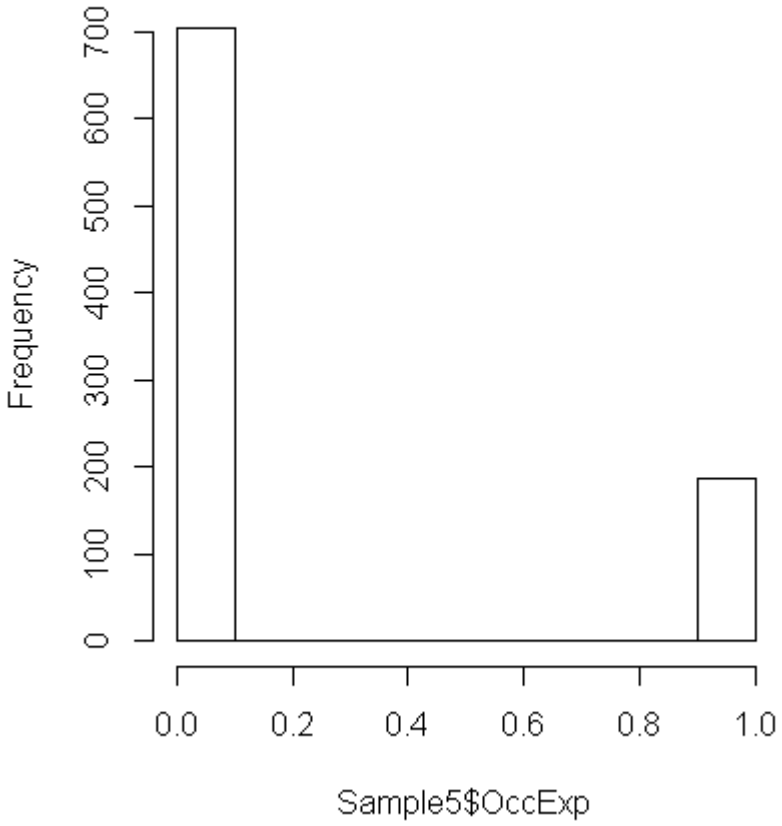
Histogram of Industrial experience



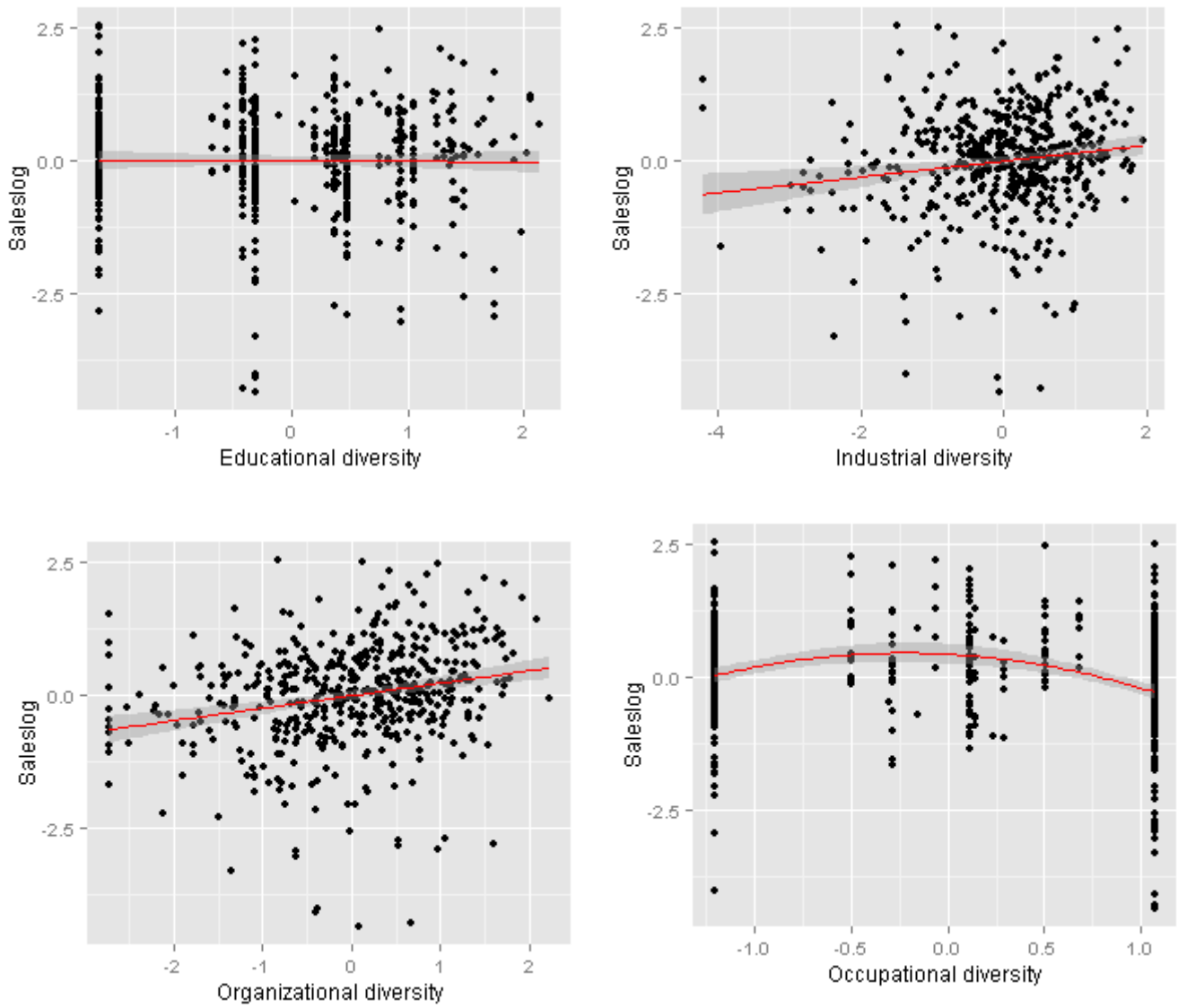
Histogram of Organizational experience



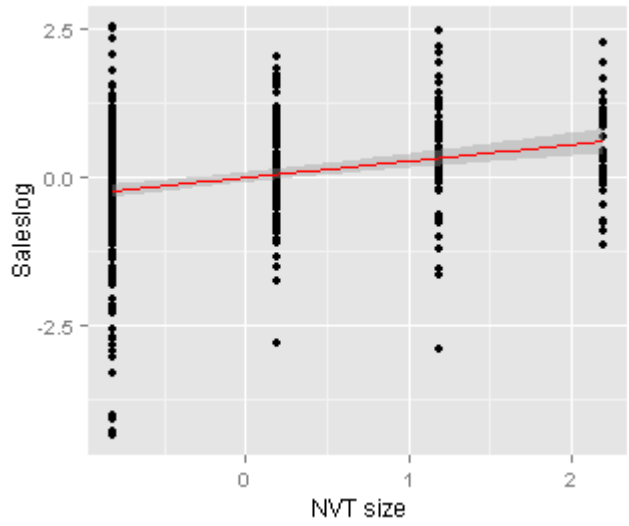
Histogram of Occupational experience

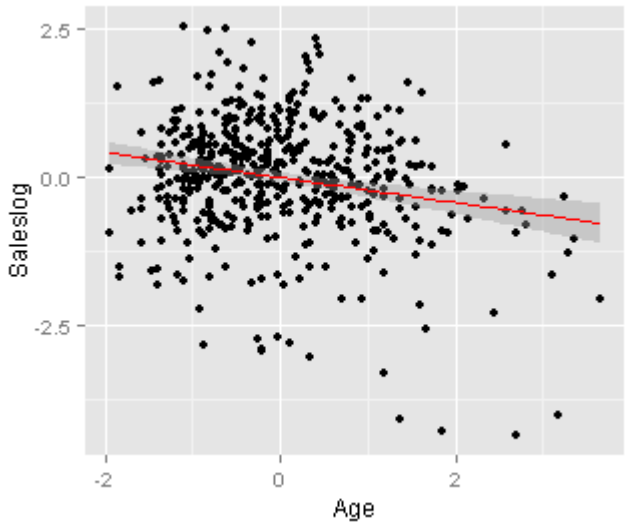


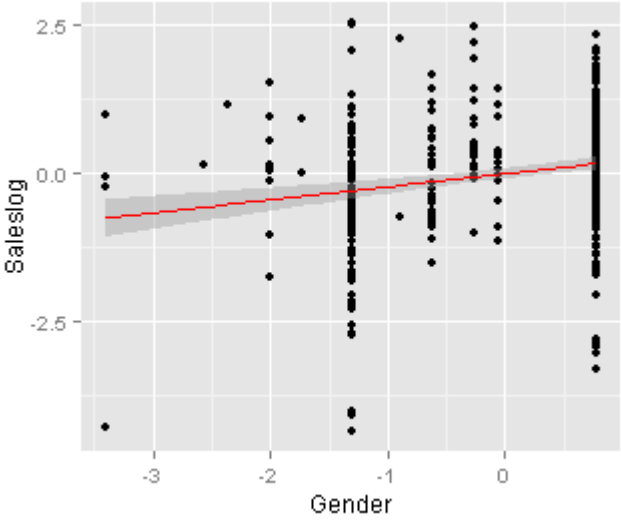
Sales regression descriptives

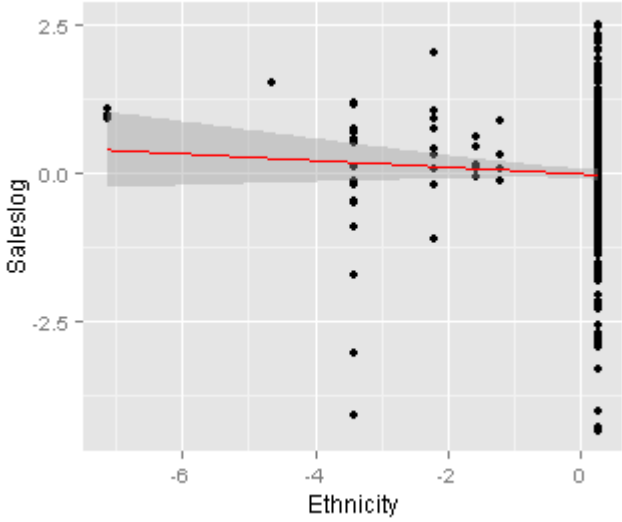


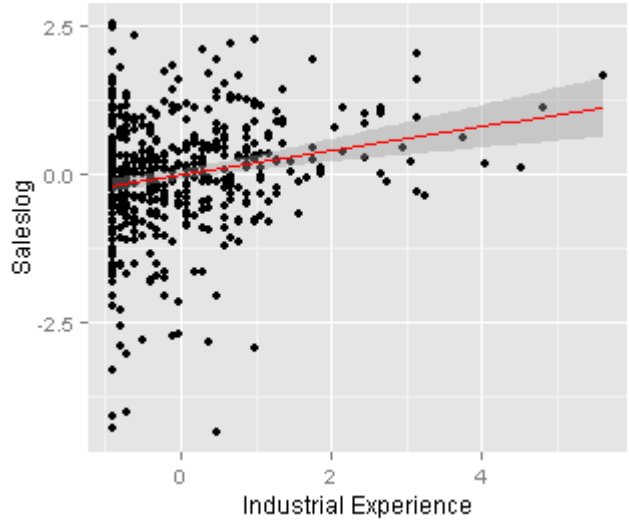
The scatter plots of the diversity dimensions with sales shows no large issues of heteroskedasticity. The clouds have a rather symmetrical shape and the SE of the regression line is rather consistent. Only in case of industrial diversity, some cases in the lower end behave strangely.





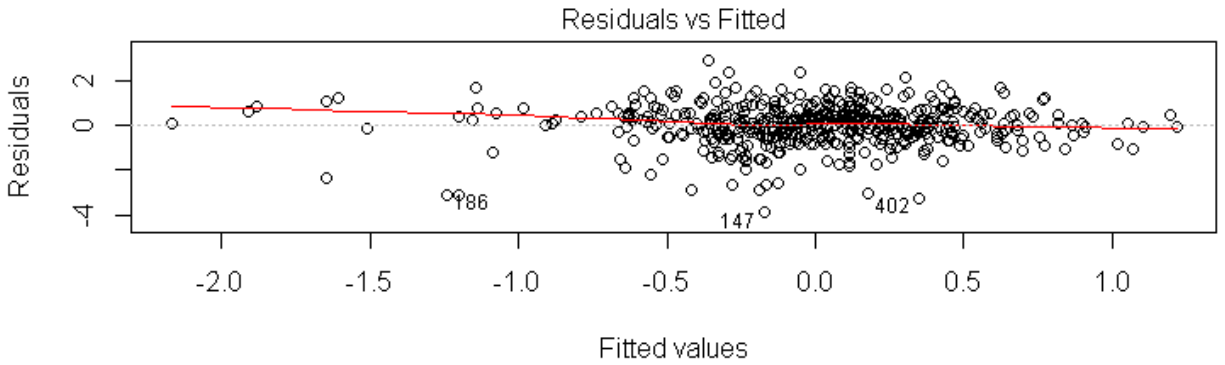


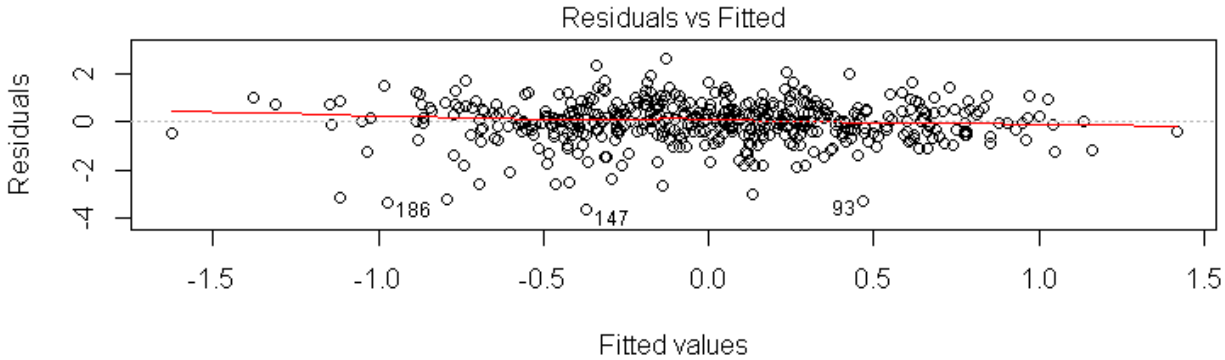


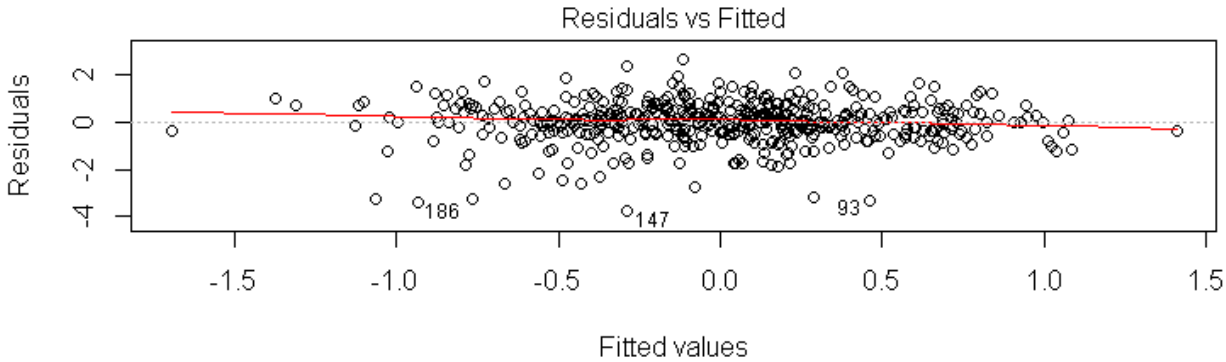


The scatter plots of the significant control variables show some reason for concern. Especially in the case of Ethnicity, Gender and Industrial Experience heteroskadasticity can be observed. This is the results of some cases that have extreme values for these variables. Because these variables are not the main focus of this research it is not a large concern.

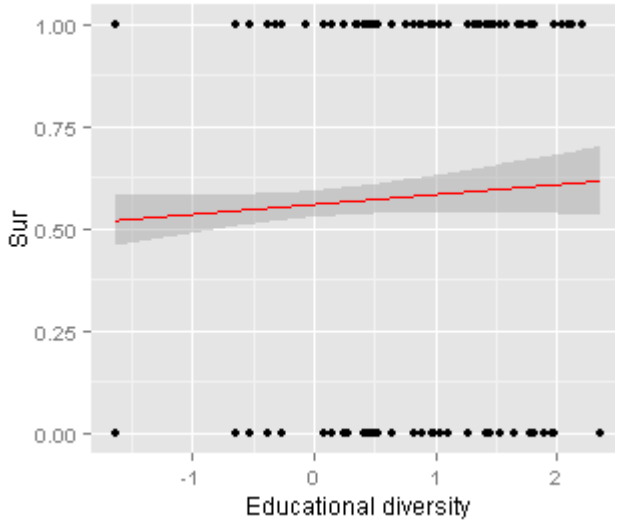
Below the residual plots are shown for the linear regression, first of the control model, second of the linear model and third of the squared model. All plots show a rather horizontal and vertical even spread. This indicates that the models are unbiased and homoscedastic. Some outliers are indicated at the low end of the residual scale. The regressions are estimated again without these estimators, but show largely similar results as before. The only change is more significance for Ethnicity, because the concerned cases are the outliers in the ethnicity variable.

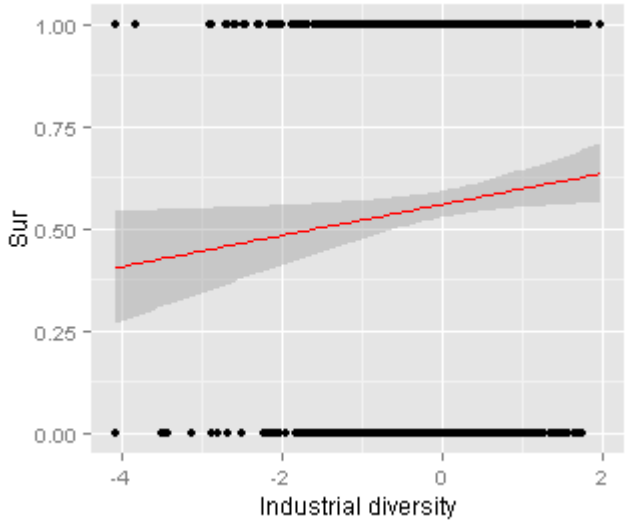


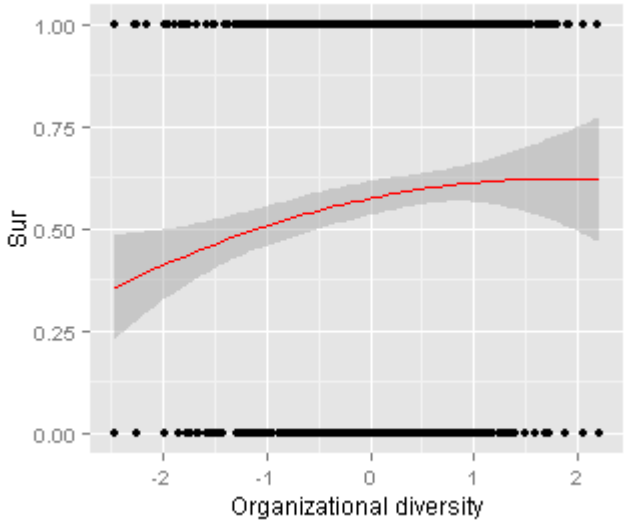


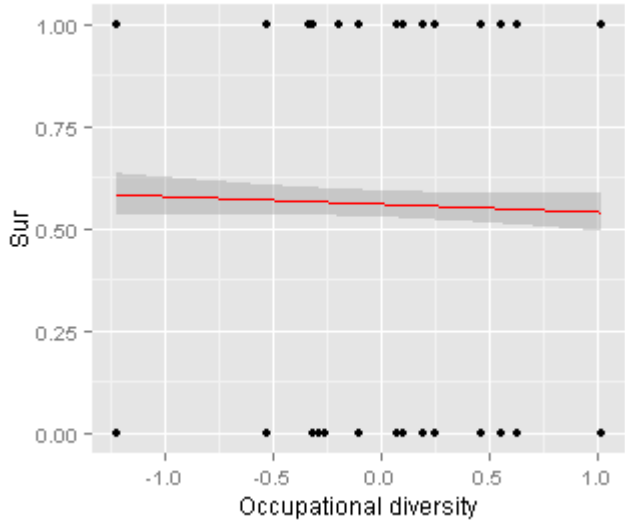


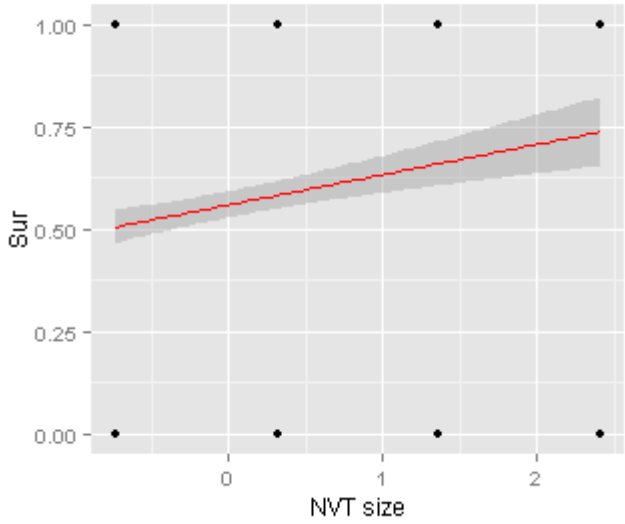
Survival regression descriptives





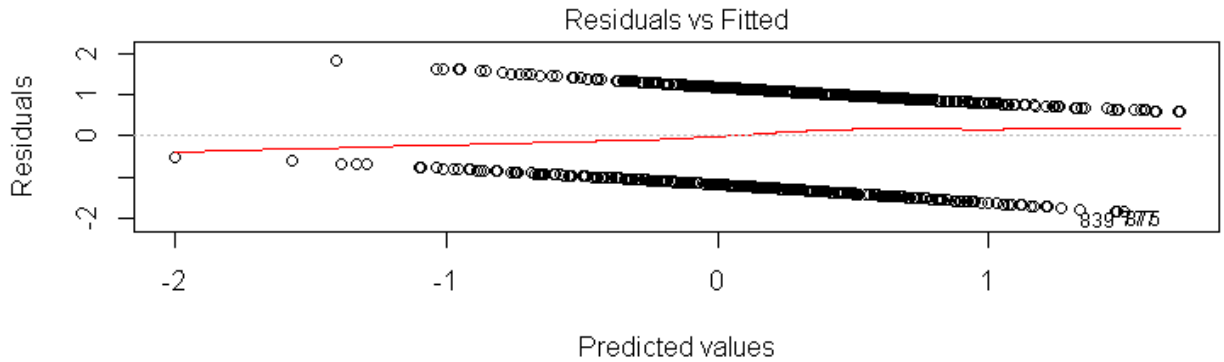


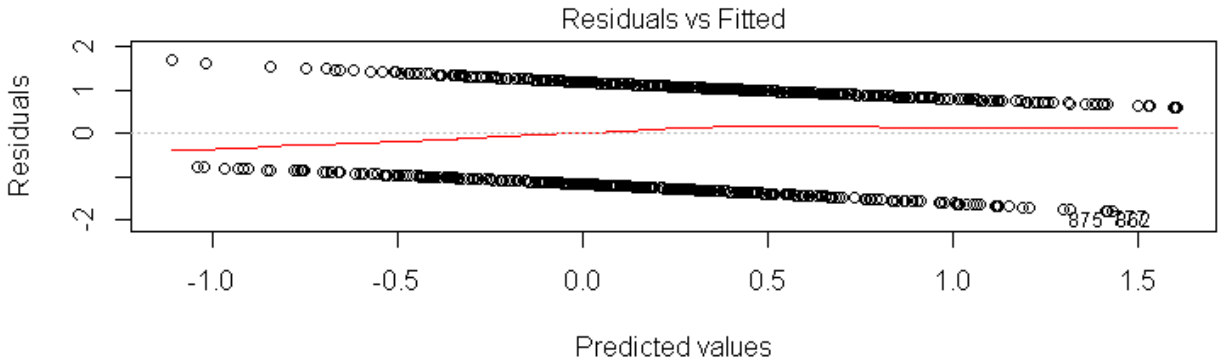


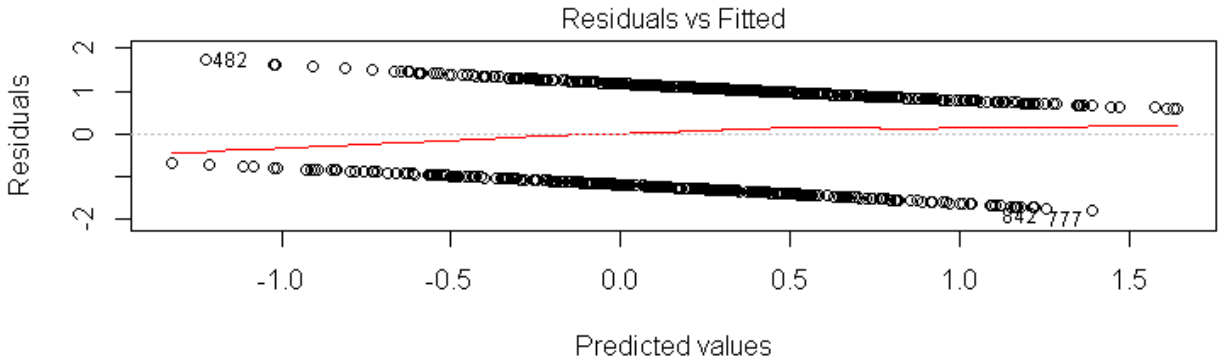


The scatter plots of independent variables and NVT with survival show somewhat higher SE than with sales. In case of industrial diversity the low amount of cases at the lower end results in higher SE. The same is the case for organizational diversity and NVT size.

The residual plots for the survival models are shown below, first the control model, second the linear model and third the squared model. The residuals are not evenly spread over the predicted values. This indicates a bias in the results. The low explanatory power of the Survival models mean that a large part is explained by the error term. Hence, I observe a correlation between the residuals and predicted values.







Appendix D: Additional analyses

Absolute continuous Entropy diversities

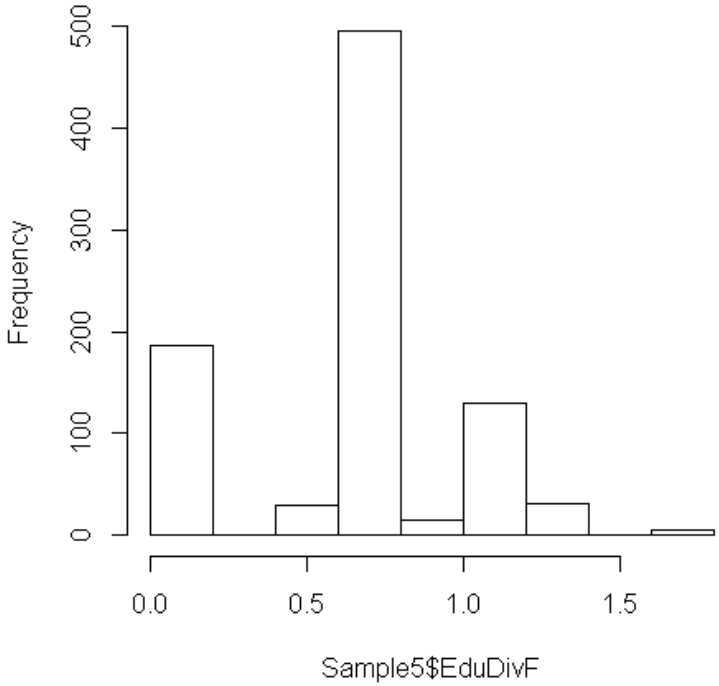
Discussion: These analyses are based on the uncorrected diversity measures (see Appendix B). Correlation between Educational and Occupational dimension and NVT is .43 and .29 respectively, the VIF values for the different models that are estimated however remain acceptable. The Organizational and Industrial show similar results as with the relative Educational and Occupational measures. Educational has a negative linear relationship with sales, and Industrial diversity shows a negative marginally significant effect in one model. Note that the effect of NVT is highly significant in all different models. The histograms of Educational and Occupational show a strange pattern. The peak in the middle is largely due to two person NVTs that are completely dissimilar. In these teams, no overlap exists as a basis for communication. However, currently they have the same level of diversity as rather homogeneous bigger teams. The diversity values that can be reached are heavily dependent on the size of the NVT because an individual is limited to a single Educational and Occupational status. Hence, in order to compare the different teams accurately, I control for the theoretical maximum diverisity a team can reach and calculate a relative Entropy measure.

Correlation matrix of Educational and Occupational absolute measures and other studied variables

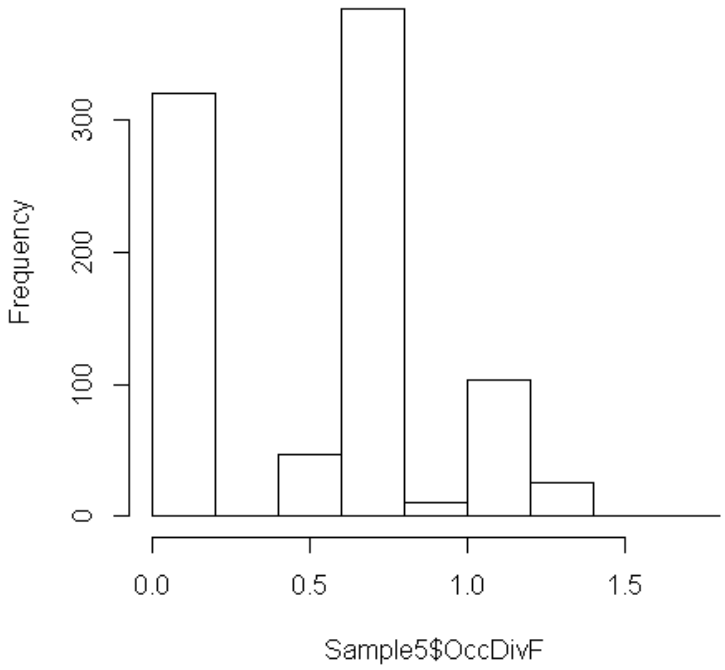
	Sur	Saleslog	Year	NVT	Age	Gen	Eth	EduExp	IndExp	OrgExp	OccExp	EduDivF	IndDivF	OrgDivF	OccDivF
Sur	1.00	0.52	0.02	0.14	0.00	0.06	0.05	0.02	0.06	0.01	0.02	0.05	0.07	0.11	0.01
Saleslog	0.52	1.00	-0.02	0.29	-0.06	0.14	0.01	0.03	0.19	-0.04	0.08	0.03	0.14	0.25	0.03
Year	0.02	-0.02	1.00	-0.06	0.09	-0.03	-0.13	0.05	0.05	0.15	0.01	0.00	0.00	-0.03	0.04
NVT	0.14	0.29	-0.06	1.00	-0.17	0.13	0.05	0.21	0.31	-0.01	0.11	0.43	0.40	0.37	0.29
Age	0.00	-0.06	0.09	-0.17	1.00	-0.32	0.05	-0.15	0.23	0.08	-0.02	0.09	-0.25	0.17	0.10
Gen	0.06	0.14	-0.03	0.13	-0.32	1.00	0.00	0.11	0.04	-0.06	0.01	-0.13	0.10	-0.03	-0.09
Eth	0.05	0.01	-0.13	0.05	0.05	0.00	1.00	0.06	0.09	0.02	0.02	0.04	0.18	0.16	0.07
EduExp	0.02	0.03	0.05	0.21	-0.15	0.11	0.06	1.00	0.22	0.04	0.03	0.20	0.13	0.12	-0.02
IndExp	0.06	0.19	0.05	0.31	0.23	0.04	0.09	0.22	1.00	0.06	0.01	0.14	-0.04	0.36	-0.04
OrgExp	0.01	-0.04	0.15	-0.01	0.08	-0.06	0.02	0.04	0.06	1.00	0.02	0.01	0.05	0.10	0.03
OccExp	0.02	0.08	0.01	0.11	-0.02	0.01	0.02	0.03	0.01	0.02	1.00	0.15	0.11	0.16	0.38
EduDivF	0.05	0.03	0.00	0.43	0.09	-0.13	0.04	0.20	0.14	0.01	0.15	1.00	0.20	0.29	0.35
IndDivF	0.07	0.14	0.00	0.40	-0.25	0.10	0.18	0.13	-0.04	0.05	0.11	0.20	1.00	0.56	0.25
OrgDivF	0.11	0.25	-0.03	0.37	0.17	-0.03	0.16	0.12	0.36	0.10	0.16	0.29	0.56	1.00	0.27
OccDivF	0.01	0.03	0.04	0.29	0.10	-0.09	0.07	-0.02	-0.04	0.03	0.38	0.35	0.25	0.27	1.00

	Sur	Saleslog	Year	NVT	Age	Gen	Eth	EduExp	IndExp	OrgExp	OccExp	EduDivF	IndDivF	OrgDivF	OccDivF
Sur		0.0000	0.6518	0.0000	0.9960	0.0744	0.1352	0.5463	0.0560	0.6590	0.5853	0.1249	0.0321	0.0007	0.6945
Saleslog	0.0000		0.5474	0.0000	0.1190	0.0003	0.7873	0.3924	0.0000	0.2942	0.0443	0.3777	0.0002	0.0000	0.3955
Year	0.6518	0.5474		0.1001	0.0104	0.4513	0.0000	0.1398	0.1379	0.0000	0.7111	0.9028	0.9145	0.3199	0.2825
NVT	0.0000	0.0000	0.1001		0.0000	0.0002	0.1679	0.0000	0.0000	0.8182	0.0012	0.0000	0.0000	0.0000	0.0000
Age	0.9960	0.1190	0.0104	0.0000		0.0000	0.1752	0.0000	0.0000	0.0139	0.6110	0.0092	0.0000	0.0000	0.0019
Gen	0.0744	0.0003	0.4513	0.0002	0.0000		0.9713	0.0017	0.2108	0.0969	0.7845	0.0002	0.0026	0.3968	0.0076
Eth	0.1352	0.7873	0.0000	0.1679	0.1752	0.9713		0.0652	0.0049	0.4858	0.4721	0.1870	0.0000	0.0000	0.0475
EduExp	0.5463	0.3924	0.1398	0.0000	0.0000	0.0017	0.0652		0.0000	0.2509	0.4212	0.0000	0.0000	0.0003	0.4677
IndExp	0.0560	0.0000	0.1379	0.0000	0.0000	0.2108	0.0049	0.0000		0.0535	0.8216	0.0000	0.2105	0.0000	0.2714
OrgExp	0.6590	0.2942	0.0000	0.8182	0.0139	0.0969	0.4858	0.2509	0.0535		0.4996	0.7014	0.1379	0.0024	0.2977
OccExp	0.5853	0.0443	0.7111	0.0012	0.6110	0.7845	0.4721	0.4212	0.8216	0.4996		0.0000	0.0007	0.0000	0.0000
EduDivF	0.1249	0.3777	0.9028	0.0000	0.0092	0.0002	0.1870	0.0000	0.0000	0.7014	0.0000		0.0000	0.0000	0.0000
IndDivF	0.0321	0.0002	0.9145	0.0000	0.0000	0.0026	0.0000	0.0000	0.2105	0.1379	0.0007	0.0000		0.0000	0.0000
OrgDivF	0.0007	0.0000	0.3199	0.0000	0.0000	0.3968	0.0000	0.0003	0.0000	0.0024	0.0000	0.0000	0.0000		0.0000
OccDivF	0.6945	0.3955	0.2825	0.0000	0.0019	0.0076	0.0475	0.4677	0.2714	0.2977	0.0000	0.0000	0.0000	0.0000	

Histogram of Educational diversity



Histogram of Occupational diversity



Linear Survival Model

```
Call:
glm(formula = Sur ~ Year + NVT + Age + Gen + Eth + EduExp + IndExp +
    OrgExp + OccExp + EduDivF + IndDivF + OrgDivF + OccDivF,
    family = "binomial", data = Sample5, na.action = na.omit)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8685  -1.2202   0.8204   1.0330   1.5030

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.203515   0.755543  -2.916 0.003540 **
Year2003     0.554651   0.341351   1.625 0.104190
Year2004     0.348495   0.310979   1.121 0.262442
Year2005     0.392155   0.310242   1.264 0.206220
Year2006     0.655950   0.303492   2.161 0.030669 *
Year2007     0.809975   0.298847   2.710 0.006722 **
Year2008     0.381108   0.328622   1.160 0.246164
Year2009     0.134344   0.367567   0.365 0.714742
NVT          0.334406   0.100572   3.325 0.000884 ***
Age          0.008084   0.009470   0.854 0.393261
Gen          0.458202   0.298221   1.536 0.124428
Eth          0.495978   0.464298   1.068 0.285416
EduExp      -0.069957   0.161064  -0.434 0.664038
IndExp      -0.004226   0.009518  -0.444 0.657040
OrgExp       0.059258   0.188020   0.315 0.752635
OccExp       0.053860   0.187243   0.288 0.773615
EduDivF     -0.023471   0.225779  -0.104 0.917204
IndDivF     -0.119871   0.240351  -0.499 0.617967
OrgDivF      0.237218   0.145469   1.631 0.102950
OccDivF     -0.271308   0.207543  -1.307 0.191131
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1202.7  on 885  degrees of freedom
Residual deviance: 1163.1  on 866  degrees of freedom
  (5 observations deleted due to missingness)
AIC: 1203.1

Number of Fisher Scoring iterations: 4
```

Linear Sales Model

```
Call:
lm(formula = Saleslog ~ Year + NVT + Age + Gen + Eth + EduExp +
    IndExp + OrgExp + OccExp + EduDivF + IndDivF + OrgDivF +
    OccDivF, data = Sample5Sur, na.action = na.omit)

Residuals:
    Min      1Q  Median      3Q     Max
-5.4586 -0.6307  0.0808  0.7881  3.9933

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.623725   0.715601  20.436  < 2e-16 ***
Year2003     0.064496   0.299723   0.215 0.829717
Year2004     0.512344   0.283778   1.805 0.071654 .
Year2005     0.295628   0.284768   1.038 0.299748
Year2006     0.567298   0.275586   2.059 0.040099 *
Year2007     0.230656   0.265929   0.867 0.386192
Year2008     0.452445   0.301099   1.503 0.133611
Year2009    -0.112293   0.360387  -0.312 0.755493
NVT          0.340320   0.081134   4.195 3.28e-05 ***
Age         -0.033194   0.008575  -3.871 0.000124 ***
Gen          0.833255   0.269939   3.087 0.002144 **
Eth         -0.775651   0.458909  -1.690 0.091660 .
EduExp      -0.085958   0.135328  -0.635 0.525622
IndExp       0.008379   0.007635   1.097 0.273026
OrgExp       0.050897   0.158012   0.322 0.747517
OccExp       0.186279   0.153745   1.212 0.226279
EduDivF     -0.562655   0.189275  -2.973 0.003106 **
IndDivF     -0.394767   0.212514  -1.858 0.063859 .
OrgDivF      0.596802   0.132370   4.509 8.27e-06 ***
OccDivF     -0.172422   0.172120  -1.002 0.316985
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.292 on 464 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.2261,    Adjusted R-squared:  0.1944
F-statistic: 7.134 on 19 and 464 DF,  p-value: < 2.2e-16
```

Squared Survival Model

```
Call:
glm(formula = Sur ~ Year + NVT + Age + Gen + Eth + EduExp + IndExp +
    OrgExp + OccExp + EduDivF + I(EduDivF^2) + IndDivF + I(IndDivF^2) +
    OrgDivF + I(OrgDivF^2) + OccDivF + I(OccDivF^2), family = "binomial",
    data = Sample5, na.action = na.omit)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8893  -1.2347   0.8098   1.0304   1.5763

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.095744   0.948686  -2.209 0.027168 *
Year2003      0.517158   0.343284   1.507 0.131938
Year2004      0.311245   0.313160   0.994 0.320279
Year2005      0.390844   0.312156   1.252 0.210541
Year2006      0.619301   0.305608   2.026 0.042718 *
Year2007      0.792916   0.300171   2.642 0.008253 **
Year2008      0.379658   0.330884   1.147 0.251214
Year2009      0.116892   0.370810   0.315 0.752584
NVT           0.404168   0.118307   3.416 0.000635 ***
Age           0.006212   0.009581   0.648 0.516713
Gen           0.484743   0.299304   1.620 0.105326
Eth           0.501218   0.473413   1.059 0.289721
EduExp       -0.099287   0.162323  -0.612 0.540762
IndExp       -0.001297   0.009718  -0.134 0.893793
OrgExp        0.047093   0.189082   0.249 0.803315
OccExp        0.069223   0.188141   0.368 0.712923
EduDivF       0.710965   0.511483   1.390 0.164527
I(EduDivF^2) -0.707113   0.451926  -1.565 0.117662
IndDivF      -1.224062   0.892578  -1.371 0.170257
I(IndDivF^2)  0.361664   0.286150   1.264 0.206268
OrgDivF       1.031549   0.399856   2.580 0.009886 **
I(OrgDivF^2) -0.265250   0.120188  -2.207 0.027317 *
OccDivF      -0.401388   0.505548  -0.794 0.427215
I(OccDivF^2)  0.134480   0.479461   0.280 0.779108
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 1202.7  on 885  degrees of freedom
Residual deviance: 1155.2  on 862  degrees of freedom
  (5 observations deleted due to missingness)
AIC: 1203.2
Number of Fisher Scoring iterations: 4
```

Squared Sales Model

```
Call:
lm(formula = Saleslog ~ Year + NVT + Age + Gen + Eth + EduExp +
    IndExp + OrgExp + OccExp + +EduDivF + I(EduDivF^2) + IndDivF +
    I(IndDivF^2) + OrgDivF + I(OrgDivF^2) + OccDivF + I(OccDivF^2),
    data = Sample5Sur, na.action = na.omit)
Residuals:
    Min      1Q  Median      3Q     Max
-5.4389 -0.6262  0.0605  0.7763  3.9997

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  14.738706   0.844156  17.460  < 2e-16 ***
Year2003      0.095930   0.302399   0.317 0.751213
Year2004      0.551444   0.287403   1.919 0.055639 .
Year2005      0.319495   0.287008   1.113 0.266207
Year2006      0.609794   0.279306   2.183 0.029523 *
Year2007      0.258197   0.268056   0.963 0.335942
Year2008      0.477470   0.304185   1.570 0.117179
Year2009     -0.089833   0.364157  -0.247 0.805261
NVT           0.319959   0.092306   3.466 0.000577 ***
Age          -0.033482   0.008643  -3.874 0.000123 ***
Gen           0.822830   0.271348   3.032 0.002563 **
Eth          -0.767092   0.467808  -1.640 0.101739
EduExp       -0.073574   0.136872  -0.538 0.591153
IndExp        0.008205   0.007892   1.040 0.299032
OrgExp        0.052187   0.158816   0.329 0.742608
OccExp        0.170181   0.156360   1.088 0.276991
EduDivF      -0.948975   0.450369  -2.107 0.035649 *
I(EduDivF^2)  0.352691   0.375084   0.940 0.347559
IndDivF      -0.344623   0.771529  -0.447 0.655320
I(IndDivF^2) -0.016046   0.244461  -0.066 0.947693
OrgDivF       0.459869   0.385002   1.194 0.232914
I(OrgDivF^2)  0.044775   0.109724   0.408 0.683410
OccDivF       0.082810   0.423645   0.195 0.845112
I(OccDivF^2) -0.247155   0.384172  -0.643 0.520321
---
Residual standard error: 1.296 on 460 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.2282,    Adjusted R-squared:  0.1896
F-statistic: 5.913 on 23 and 460 DF,  p-value: 1.981e-15
```

Relative continuous Educational & Occupational Entropy diversities

Discussion: In the relative Educational & Occupational dimensions, the absolute is divided by the theoretical maximum (see Appendix B). Hence, I can more accurately compare teams of different sizes. The correlation with NVT turns from a large positive value to rather small negative values. The histograms have a spread between 0 and 1 (due to small rounding errors some are displayed above 1). In the regression analyses we observe a similar pattern as with the absolute measures. However, the expained variance is higher and the effect of NVT is reduced in some of the models, as Educational and Occupational relative measures are introduced. The scatter plots show a clear inverted U-shaped pattern for Occupational diversity (and to some extend also for Educational). This effect cannot clearly be seen in the regression results, but some indication for effects are there (significant Occ squared and linear term Edu on sales). We observe clear clusters in the data. Hence, Edu and Occ are also made categorical to improve the results.

Correlation matrix of Educational and Occupational relative measures and other studied variables

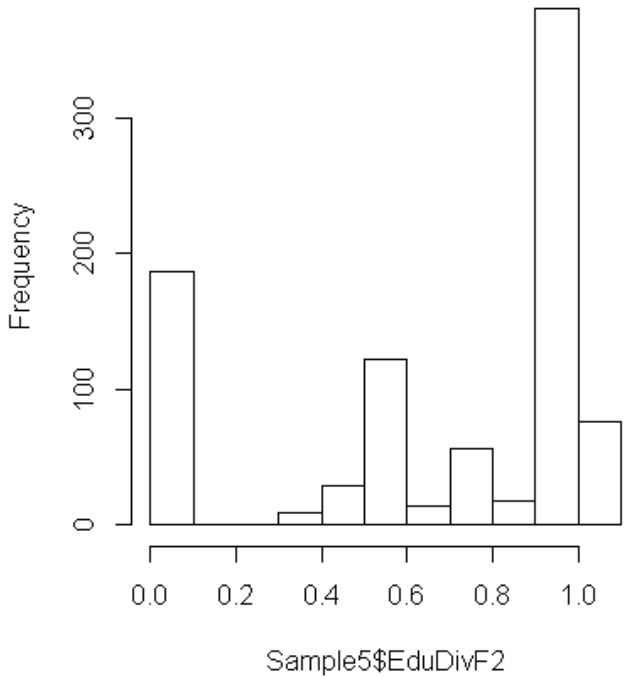
	Sur	Saleslog	Year	NVT	Age	Gen	Eth	EduExp	IndExp	OrgExp	OccExp	EduDivF2	IndDivF	OrgDivF	OccDivF2
Sur	1.00	0.52	0.02	0.14	0.00	0.06	0.05	0.02	0.06	0.01	0.02	-0.01	0.07	0.11	-0.04
Saleslog	0.52	1.00	-0.02	0.29	-0.06	0.14	0.01	0.03	0.19	-0.04	0.08	-0.13	0.14	0.25	-0.08
Year	0.02	-0.02	1.00	-0.06	0.09	-0.03	-0.13	0.05	0.05	0.15	0.01	0.04	0.00	-0.03	0.05
NVT	0.14	0.29	-0.06	1.00	-0.17	0.13	0.05	0.21	0.31	-0.01	0.11	-0.12	0.40	0.37	-0.09
Age	0.00	-0.06	0.09	-0.17	1.00	-0.32	0.05	-0.15	0.23	0.08	-0.02	0.20	-0.25	0.17	0.18
Gen	0.06	0.14	-0.03	0.13	-0.32	1.00	0.00	0.11	0.04	-0.06	0.01	-0.21	0.10	-0.03	-0.13
Eth	0.05	0.01	-0.13	0.05	0.05	0.00	1.00	0.06	0.09	0.02	0.02	0.02	0.18	0.16	0.06
EduExp	0.02	0.03	0.05	0.21	-0.15	0.11	0.06	1.00	0.22	0.04	0.03	0.09	0.13	0.12	-0.11
IndExp	0.06	0.19	0.05	0.31	0.23	0.04	0.09	0.22	1.00	0.06	0.01	-0.04	-0.04	0.36	-0.13
OrgExp	0.01	-0.04	0.15	-0.01	0.08	-0.06	0.02	0.04	0.06	1.00	0.02	0.03	0.05	0.10	0.06
OccExp	0.02	0.08	0.01	0.11	-0.02	0.01	0.02	0.03	0.01	0.02	1.00	0.08	0.11	0.16	0.32
EduDivF2	-0.01	-0.13	0.04	-0.12	0.20	-0.21	0.02	0.09	-0.04	0.03	0.08	1.00	-0.02	0.11	0.25
IndDivF	0.07	0.14	0.00	0.40	-0.25	0.10	0.18	0.13	-0.04	0.05	0.11	-0.02	1.00	0.56	0.11
OrgDivF	0.11	0.25	-0.03	0.37	0.17	-0.03	0.16	0.12	0.36	0.10	0.16	0.11	0.56	1.00	0.16
OccDivF2	-0.04	-0.08	0.05	-0.09	0.18	-0.13	0.06	-0.11	-0.13	0.06	0.32	0.25	0.11	0.16	1.00

P

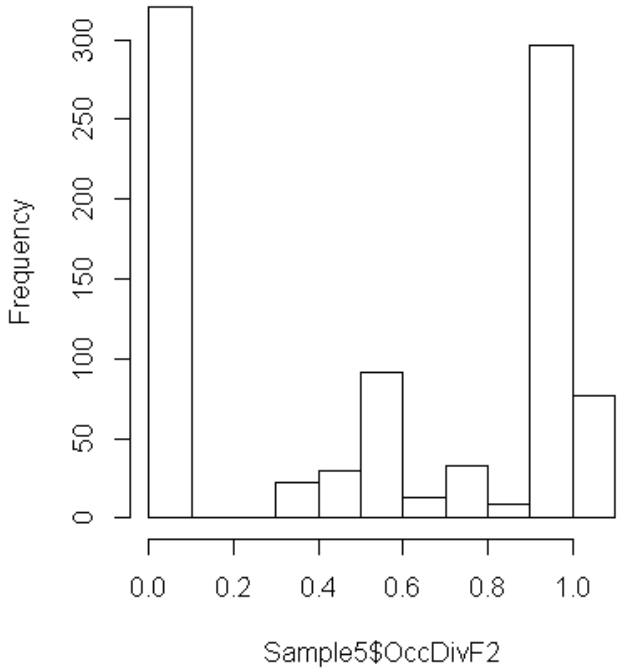
	Sur	Saleslog	Year	NVT	Age	Gen	Eth	EduExp	IndExp	OrgExp	OccExp	EduDivF2	IndDivF	OrgDivF	OccDivF2
Sur		0.0000	0.6518	0.0000	0.9960	0.0744	0.1352	0.5463	0.0560	0.6590	0.5853	0.7129	0.0321	0.0007	0.2775
Saleslog	0.0000		0.5474	0.0000	0.1190	0.0003	0.7873	0.3924	0.0000	0.2942	0.0443	0.0008	0.0002	0.0000	0.0305
Year	0.6518	0.5474		0.1001	0.0104	0.4513	0.0000	0.1398	0.1379	0.0000	0.7111	0.2204	0.9145	0.3199	0.1569

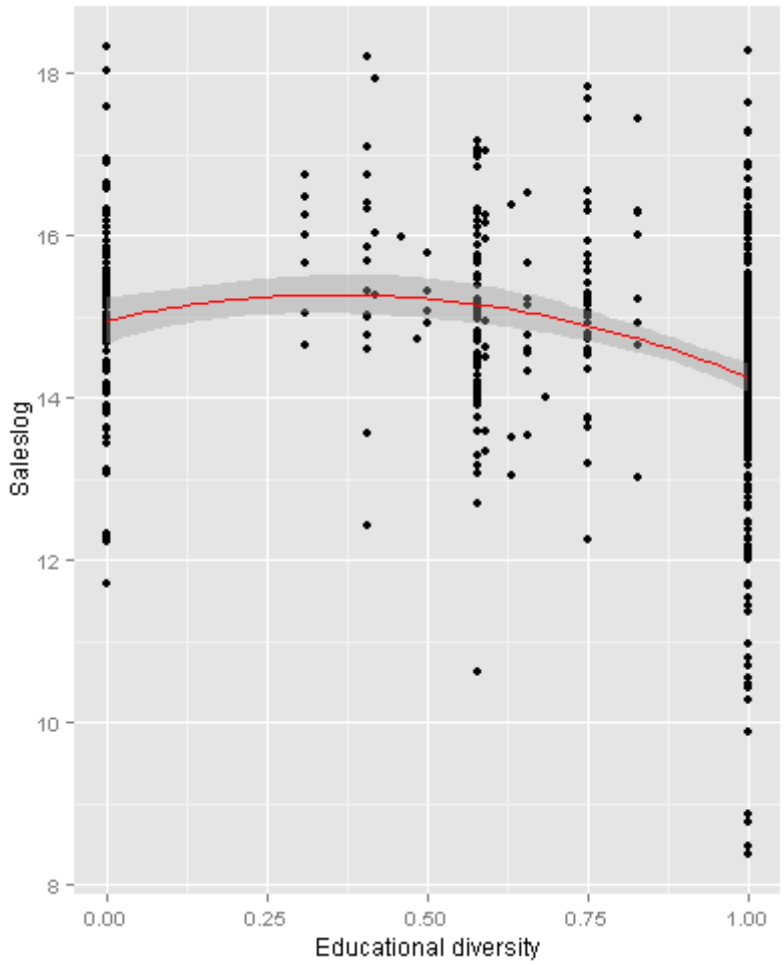
```
NVT      0.0000 0.0000   0.1001        0.0000 0.0002 0.1679 0.0000 0.0000 0.8182 0.0012 0.0002   0.0000  0.0000  0.0052
Age      0.9960 0.1190   0.0104 0.0000        0.0000 0.1752 0.0000 0.0000 0.0139 0.6110 0.0000   0.0000  0.0000  0.0000
Gen      0.0744 0.0003   0.4513 0.0002 0.0000        0.9713 0.0017 0.2108 0.0969 0.7845 0.0000   0.0026  0.3968  0.0000
Eth      0.1352 0.7873   0.0000 0.1679 0.1752 0.9713        0.0652 0.0049 0.4858 0.4721 0.5217   0.0000  0.0000  0.1015
EduExp   0.5463 0.3924   0.1398 0.0000 0.0000 0.0017 0.0652        0.0000 0.2509 0.4212 0.0103   0.0000  0.0003  0.0011
IndExp   0.0560 0.0000   0.1379 0.0000 0.0000 0.2108 0.0049 0.0000        0.0535 0.8216 0.2884   0.2105  0.0000  0.0000
OrgExp   0.6590 0.2942   0.0000 0.8182 0.0139 0.0969 0.4858 0.2509 0.0535        0.4996 0.3230   0.1379  0.0024  0.0918
OccExp   0.5853 0.0443   0.7111 0.0012 0.6110 0.7845 0.4721 0.4212 0.8216 0.4996        0.0204   0.0007  0.0000  0.0000
EduDivF2 0.7129 0.0008   0.2204 0.0002 0.0000 0.0000 0.5217 0.0103 0.2884 0.3230 0.0204          0.4700  0.0009  0.0000
IndDivF  0.0321 0.0002   0.9145 0.0000 0.0000 0.0026 0.0000 0.0000 0.2105 0.1379 0.0007 0.4700          0.0000  0.0008
OrgDivF  0.0007 0.0000   0.3199 0.0000 0.0000 0.3968 0.0000 0.0003 0.0000 0.0024 0.0000 0.0009   0.0000          0.0000
OccDivF2 0.2775 0.0305   0.1569 0.0052 0.0000 0.0000 0.1015 0.0011 0.0000 0.0918 0.0000 0.0000   0.0008  0.0000
>
```

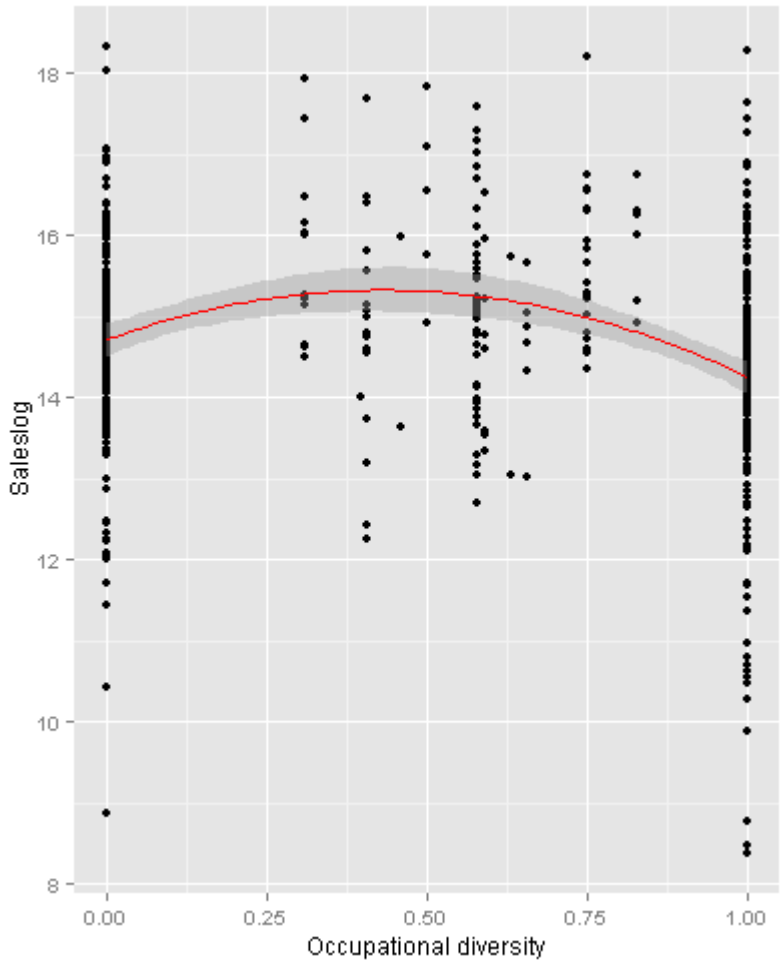
Histogram of Educational diversity



Histogram of Occupational diversity




Saleslog
18
16
14
12
10
8
0.00
0.25
0.50
0.75
1.00
Educational diversity


Saleslog
18
16
14
12
10
8
0.00
0.25
0.50
0.75
1.00
Occupational diversity

Linear Survival Model

```
Call:
glm(formula = Sur ~ Year + NVT + Age + Gen + Eth + EduExp + IndExp +
    OrgExp + OccExp + EduDivF2 + IndDivF + OrgDivF + OccDivF2,
    family = "binomial", data = Sample5, na.action = na.omit)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8925  -1.2213   0.8284   1.0307   1.5126

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.126632   0.759999  -2.798  0.00514 **
Year2003     0.542920   0.340763   1.593  0.11110
Year2004     0.332628   0.310250   1.072  0.28366
Year2005     0.383633   0.309815   1.238  0.21562
Year2006     0.641438   0.302787   2.118  0.03414 *
Year2007     0.804154   0.298620   2.693  0.00708 **
Year2008     0.362751   0.328011   1.106  0.26877
Year2009     0.113219   0.366462   0.309  0.75736
NVT          0.287275   0.091884   3.127  0.00177 **
Age          0.007629   0.009486   0.804  0.42125
Gen          0.482430   0.297994   1.619  0.10546
Eth          0.490262   0.464387   1.056  0.29110
EduExp      -0.081805   0.160925  -0.508  0.61121
IndExp      -0.003462   0.009482  -0.365  0.71506
OrgExp       0.064401   0.187986   0.343  0.73191
OccExp       0.037241   0.185052   0.201  0.84051
EduDivF2     0.055005   0.190633   0.289  0.77293
IndDivF     -0.112799   0.240646  -0.469  0.63926
OrgDivF      0.230173   0.145706   1.580  0.11417
OccDivF2    -0.228188   0.178523  -1.278  0.20118
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1202.7  on 885  degrees of freedom
Residual deviance: 1163.3  on 866  degrees of freedom
  (5 observations deleted due to missingness)
AIC: 1203.3

Number of Fisher Scoring iterations: 4
```

Linear Sales Model

```
Call:
lm(formula = Saleslog ~ Year + NVT + Age + Gen + Eth + EduExp +
    IndExp + OrgExp + OccExp + EduDivF2 + IndDivF + OrgDivF +
    OccDivF2, data = Sample5Sur, na.action = na.omit)

Residuals:
    Min      1Q  Median      3Q     Max
-5.3961 -0.5783  0.0760  0.7730  4.0686

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 15.055859   0.718990  20.940  < 2e-16 ***
Year2003     0.047887   0.298550   0.160 0.872636
Year2004     0.501812   0.282167   1.778 0.075990 .
Year2005     0.279669   0.283358   0.987 0.324166
Year2006     0.562801   0.274107   2.053 0.040611 *
Year2007     0.238760   0.264946   0.901 0.367969
Year2008     0.436119   0.299762   1.455 0.146378
Year2009    -0.150418   0.357862  -0.420 0.674443
NVT          0.190225   0.074877   2.540 0.011394 *
Age         -0.033071   0.008557  -3.865 0.000127 ***
Gen          0.830154   0.269170   3.084 0.002163 **
Eth         -0.796615   0.457633  -1.741 0.082394 .
EduExp      -0.091714   0.134503  -0.682 0.495663
IndExp       0.008348   0.007572   1.103 0.270805
OrgExp       0.068549   0.157722   0.435 0.664039
OccExp       0.179679   0.151661   1.185 0.236727
EduDivF2    -0.525360   0.170098  -3.089 0.002132 **
IndDivF     -0.396176   0.212228  -1.867 0.062567 .
OrgDivF      0.608304   0.132212   4.601 5.43e-06 ***
OccDivF2    -0.185777   0.156956  -1.184 0.237167
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.289 on 464 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.2294,    Adjusted R-squared:  0.1978
F-statistic:  7.27 on 19 and 464 DF,  p-value: < 2.2e-16
```

Squared Survival Model

```
Call:
glm(formula = Sur ~ Year + NVT + Age + Gen + Eth + EduExp + IndExp +
    OrgExp + OccExp + EduDivF2 + I(EduDivF2^2) + IndDivF +
I(IndDivF^2) +
    OrgDivF + I(OrgDivF^2) + OccDivF2 + I(OccDivF2^2), family =
"binomial",
    data = Sample5, na.action = na.omit)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8047  -1.2375   0.8197   1.0255   1.5873

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.732542   0.940352  -1.842  0.06541 .
Year2003       0.505392   0.342209   1.477  0.13972
Year2004       0.309426   0.311187   0.994  0.32006
Year2005       0.390054   0.311061   1.254  0.20986
Year2006       0.631031   0.304089   2.075  0.03797 *
Year2007       0.802323   0.299458   2.679  0.00738 **
Year2008       0.375953   0.329038   1.143  0.25321
Year2009       0.092158   0.368231   0.250  0.80238
NVT            0.265991   0.127105   2.093  0.03638 *
Age            0.005965   0.009599   0.621  0.53436
Gen            0.503618   0.299027   1.684  0.09215 .
Eth            0.490060   0.472584   1.037  0.29974
EduExp        -0.104121   0.161920  -0.643  0.52020
IndExp        -0.001066   0.009689  -0.110  0.91243
OrgExp         0.054060   0.188917   0.286  0.77476
OccExp         0.048896   0.185217   0.264  0.79178
EduDivF2       0.750293   1.025354   0.732  0.46433
I(EduDivF2^2) -0.679493   0.992373  -0.685  0.49352
IndDivF       -1.278414   0.895227  -1.428  0.15328
I(IndDivF^2)   0.380563   0.286692   1.327  0.18437
OrgDivF        1.064419   0.398888   2.668  0.00762 **
I(OrgDivF^2)  -0.273969   0.119716  -2.288  0.02211 *
OccDivF2      -0.518475   1.001383  -0.518  0.60463
I(OccDivF2^2)  0.279100   0.996170   0.280  0.77934
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 1202.7  on 885  degrees of freedom
Residual deviance: 1157.2  on 862  degrees of freedom
  (5 observations deleted due to missingness)
AIC: 1205.2
Number of Fisher Scoring iterations: 4
```

Squared Sales Model

```
Call:
lm(formula = Saleslog ~ Year + NVT + Age + Gen + Eth + EduExp +
    IndExp + OrgExp + OccExp + EduDivF2 + I(EduDivF2^2) + IndDivF +
    I(IndDivF^2) + OrgDivF + I(OrgDivF^2) + OccDivF2 +
I(OccDivF2^2),
    data = Sample5Sur, na.action = na.omit)

Residuals:
    Min      1Q  Median      3Q     Max
-5.3413 -0.5770  0.0893  0.7743  4.1057

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   15.167670   0.832650  18.216  < 2e-16 ***
Year2003       0.052164   0.300087   0.174 0.862075
Year2004       0.511007   0.283175   1.805 0.071797 .
Year2005       0.270232   0.285267   0.947 0.343986
Year2006       0.566591   0.275859   2.054 0.040548 *
Year2007       0.241314   0.265820   0.908 0.364454
Year2008       0.441765   0.300960   1.468 0.142827
Year2009      -0.146721   0.358178  -0.410 0.682268
NVT            0.132416   0.102165   1.296 0.195590
Age           -0.033470   0.008602  -3.891 0.000114 ***
Gen            0.828983   0.269691   3.074 0.002239 **
Eth           -0.755400   0.464339  -1.627 0.104458
EduExp        -0.091579   0.136558  -0.671 0.502796
IndExp         0.009368   0.007832   1.196 0.232286
OrgExp         0.085544   0.158090   0.541 0.588693
OccExp         0.174752   0.152962   1.142 0.253861
EduDivF2      -1.078312   0.837921  -1.287 0.198780
I(EduDivF2^2)  0.552355   0.804662   0.686 0.492779
IndDivF       -0.301400   0.769051  -0.392 0.695305
I(IndDivF^2)  -0.029872   0.243808  -0.123 0.902539
OrgDivF        0.487548   0.382454   1.275 0.203027
I(OrgDivF^2)   0.034107   0.109019   0.313 0.754533
OccDivF2       1.290690   0.784213   1.646 0.100479
I(OccDivF2^2) -1.504022   0.783829  -1.919 0.055626 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.29 on 460 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.2358,    Adjusted R-squared:  0.1976
F-statistic: 6.171 on 23 and 460 DF,  p-value: 2.828e-16
```

Relative categorical Educational & Occupational Entropy diversities

Discussion: For the categorical Edu and Occ measures, we do observe a pattern. In case of Occ the inverted U-shaped pattern is confirmed due to the significance of some moderate diversity categories. The entire variable shows a marginal significant relationship to sales. Moreover, the categorization of Occ has led to a higher R2 for sales. In case of Edu, we only see significance in the low diversity category, but the whole variable is significant nonetheless. Also, the categorization of Edu leads to a lower R2 than the continuous alternative. Hence, I conclude that Edu is negatively linearly related to sales.

Linear Survival Model

```
Call:
glm(formula = Sur ~ Year + NVT + Age + Gen + Eth + EduExp + IndExp +
    OrgExp + OccExp + EduDivF2C + IndDivF + OrgDivF + OccDivF2C,
    family = "binomial", data = Sample5, na.action = na.omit)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9795  -1.2147   0.8211   1.0300   1.5268

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.356241   0.804437  -2.929  0.00340 **
Year2003     0.551096   0.341727   1.613  0.10681
Year2004     0.340135   0.313358   1.085  0.27772
Year2005     0.380837   0.311802   1.221  0.22193
Year2006     0.652996   0.305061   2.141  0.03231 *
Year2007     0.824181   0.300223   2.745  0.00605 **
Year2008     0.363252   0.329881   1.101  0.27083
Year2009     0.138857   0.367992   0.377  0.70592
NVT          0.323726   0.147998   2.187  0.02872 *
Age          0.007324   0.009524   0.769  0.44193
Gen          0.465513   0.298615   1.559  0.11902
Eth          0.517876   0.467191   1.108  0.26765
EduExp      -0.096916   0.161973  -0.598  0.54961
IndExp      -0.003383   0.009570  -0.354  0.72368
OrgExp       0.069555   0.188466   0.369  0.71208
OccExp       0.051428   0.185470   0.277  0.78156
EduDivF2C1  -0.102598   0.193293  -0.531  0.59556
EduDivF2C2   0.415521   0.490667   0.847  0.39708
EduDivF2C3  -0.090275   0.260943  -0.346  0.72937
EduDivF2C4   0.005102   0.379938   0.013  0.98929
IndDivF     -0.120638   0.241018  -0.501  0.61670
OrgDivF      0.225187   0.146149   1.541  0.12336
OccDivF2C1   0.221034   0.180553   1.224  0.22088
OccDivF2C2  -0.071606   0.432157  -0.166  0.86840
OccDivF2C3   0.200907   0.281900   0.713  0.47604
OccDivF2C4  -0.392702   0.413604  -0.949  0.34239
---
    Null deviance: 1202.7  on 885  degrees of freedom
Residual deviance: 1160.7  on 860  degrees of freedom
  (5 observations deleted due to missingness)
AIC: 1212.7
Number of Fisher Scoring iterations: 4
```

Squared Survival Model

```
Call:
glm(formula = Sur ~ Year + NVT + Age + Gen + Eth + EduExp + IndExp +
    OrgExp + OccExp + EduDivF2C + IndDivF + I(IndDivF^2) + OrgDivF +
    I(OrgDivF^2) + OccDivF2C, family = "binomial", data = Sample5,
    na.action = na.omit)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.938  -1.229   0.806   1.029   1.570

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.9461214  0.9763422  -1.993  0.04623 *
Year2003      0.5086754  0.3431829   1.482  0.13828
Year2004      0.3140519  0.3142588   0.999  0.31763
Year2005      0.3881452  0.3128095   1.241  0.21467
Year2006      0.6400221  0.3059427   2.092  0.03644 *
Year2007      0.8154388  0.3009077   2.710  0.00673 **
Year2008      0.3725744  0.3306863   1.127  0.25988
Year2009      0.1172910  0.3694701   0.317  0.75090
NVT           0.3055656  0.1503146   2.033  0.04207 *
Age           0.0054751  0.0096070   0.570  0.56874
Gen           0.4869824  0.2994919   1.626  0.10394
Eth           0.5157207  0.4744724   1.087  0.27707
EduExp       -0.1131989  0.1628066  -0.695  0.48687
IndExp       -0.0006349  0.0097273  -0.065  0.94796
OrgExp        0.0605332  0.1891997   0.320  0.74901
OccExp        0.0605096  0.1856746   0.326  0.74451
EduDivF2C1   -0.1116113  0.1948504  -0.573  0.56678
EduDivF2C2    0.4721523  0.4925205   0.959  0.33774
EduDivF2C3   -0.0518564  0.2619467  -0.198  0.84307
EduDivF2C4    0.0492845  0.3809019   0.129  0.89705
IndDivF      -1.2691514  0.8942547  -1.419  0.15583
I(IndDivF^2)  0.3769916  0.2863721   1.316  0.18803
OrgDivF       1.0205549  0.4018844   2.539  0.01110 *
I(OrgDivF^2) -0.2614035  0.1209879  -2.161  0.03073 *
OccDivF2C1    0.2365817  0.1816044   1.303  0.19267
OccDivF2C2   -0.0596963  0.4328312  -0.138  0.89030
OccDivF2C3    0.1801688  0.2820456   0.639  0.52296
OccDivF2C4   -0.3089993  0.4169304  -0.741  0.45862
---
Null deviance: 1202.7  on 885  degrees of freedom
Residual deviance: 1155.4  on 858  degrees of freedom
  (5 observations deleted due to missingness)
AIC: 1211.4

Number of Fisher Scoring iterations: 4
```

Linear Sales Model

```
Call:
lm(formula = Saleslog ~ Year + NVT + Age + Gen + Eth + EduExp +
    IndExp + OrgExp + OccExp + EduDivF2C + IndDivF + OrgDivF +
    OccDivF2C, data = Sample5Sur, na.action = na.omit)

Residuals:
    Min      1Q  Median      3Q     Max
-5.3219 -0.5557  0.1071  0.7336  4.1092

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.549406   0.750511  19.386  < 2e-16 ***
Year2003     0.020817   0.300419   0.069  0.94479
Year2004     0.450885   0.287906   1.566  0.11802
Year2005     0.239061   0.289427   0.826  0.40924
Year2006     0.518872   0.278311   1.864  0.06291 .
Year2007     0.216748   0.267758   0.809  0.41865
Year2008     0.405297   0.302714   1.339  0.18127
Year2009    -0.187797   0.360108  -0.522  0.60227
NVT          0.105444   0.117572   0.897  0.37027
Age         -0.034086   0.008579  -3.973 8.24e-05 ***
Gen          0.820985   0.269412   3.047  0.00244 **
Eth         -0.770767   0.459180  -1.679  0.09392 .
EduExp      -0.076864   0.137005  -0.561  0.57505
IndExp       0.010396   0.007659   1.357  0.17533
OrgExp       0.081396   0.158104   0.515  0.60692
OccExp       0.170904   0.152261   1.122  0.26226
EduDivF2C1   0.490141   0.174409   2.810  0.00516 **
EduDivF2C2   0.394738   0.349432   1.130  0.25921
EduDivF2C3   0.055414   0.214618   0.258  0.79637
EduDivF2C4   0.013151   0.287647   0.046  0.96356
IndDivF     -0.391847   0.212858  -1.841  0.06629 .
OrgDivF      0.589345   0.133121   4.427 1.20e-05 ***
OccDivF2C1   0.251265   0.160518   1.565  0.11819
OccDivF2C2   0.298177   0.317887   0.938  0.34874
OccDivF2C3   0.564818   0.229445   2.462  0.01420 *
OccDivF2C4   0.585493   0.318311   1.839  0.06651 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 1.288 on 458 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.2412,    Adjusted R-squared:  0.1998
F-statistic: 5.823 on 25 and 458 DF,  p-value: 4.546e-16
```

Significance of categorical variables

Single term deletions

```
Model:
Saleslog ~ Year + NVT + Age + Gen + Eth + EduExp + IndExp + OrgExp +
    OccExp + EduDivF2C + IndDivF + OrgDivF + OccDivF2C
          Df Sum of Sq    RSS    AIC  Pr(>Chi)
<none>                 759.73 270.22
Year       7    18.385 778.11 267.80  0.115505
NVT        1     1.334 761.06 269.07  0.356766
Age        1    26.186 785.91 284.62 5.125e-05 ***
Gen        1    15.404 775.13 277.94  0.001828 **
Eth        1     4.674 764.40 271.19  0.084904 .
EduExp     1     0.522 760.25 268.56  0.564183
IndExp     1     3.056 762.78 270.17  0.163327
OrgExp     1     0.440 760.17 268.50  0.596695
OccExp     1     2.090 761.82 269.55  0.248883
EduDivF2C  4    15.980 775.71 272.30  0.039188 *
IndDivF    1     5.621 765.35 271.79  0.058901 .
OrgDivF    1    32.512 792.24 288.50 6.685e-06 ***
OccDivF2C  4    13.089 772.81 270.49  0.082251 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

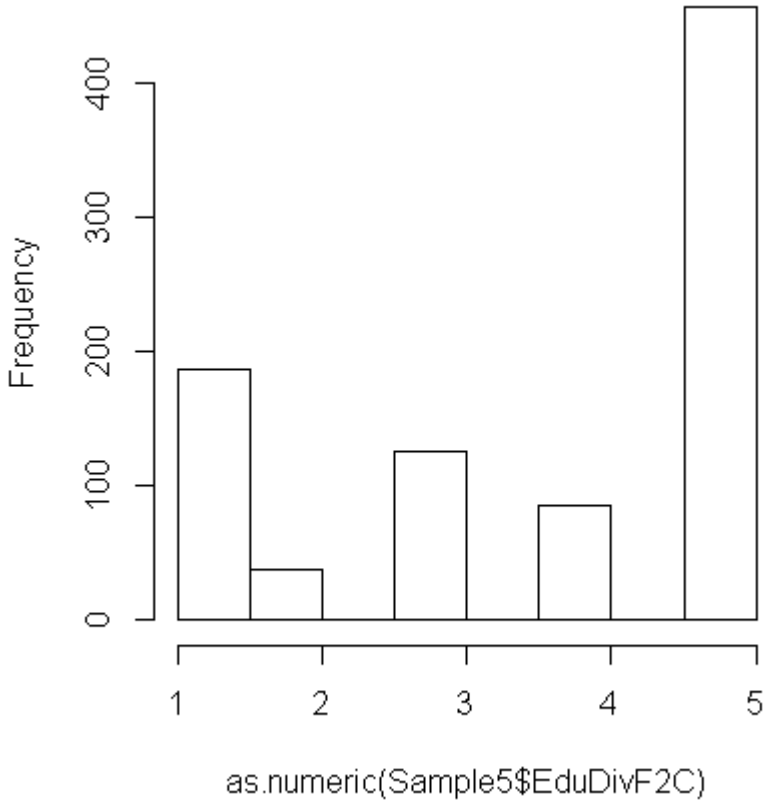
Squared Sales Model

```
Call:
lm(formula = Saleslog ~ Year + NVT + Age + Gen + Eth + EduExp +
    IndExp + OrgExp + OccExp + EduDivF2C + IndDivF + I(IndDivF^2) +
    OrgDivF + I(OrgDivF^2) + OccDivF2C, data = Sample5Sur, na.action = na.omit)

Residuals:
    Min      1Q  Median      3Q     Max
-5.3182 -0.5552  0.1089  0.7271  4.1116

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  14.535220   0.870095  16.705  < 2e-16 ***
Year2003      0.021550   0.301605   0.071  0.94307
Year2004      0.451279   0.288575   1.564  0.11855
Year2005      0.238245   0.290282   0.821  0.41223
Year2006      0.519594   0.279092   1.862  0.06328 .
Year2007      0.217488   0.268453   0.810  0.41827
Year2008      0.403678   0.303856   1.329  0.18467
Year2009     -0.187217   0.360934  -0.519  0.60422
NVT           0.106057   0.120056   0.883  0.37748
Age          -0.034024   0.008625  -3.945 9.24e-05 ***
Gen           0.820095   0.270433   3.033  0.00256 **
Eth          -0.766925   0.466318  -1.645  0.10073
EduExp       -0.075672   0.137678  -0.550  0.58284
IndExp        0.010220   0.007907   1.293  0.19681
OrgExp        0.081057   0.158511   0.511  0.60934
OccExp        0.169641   0.153034   1.109  0.26822
EduDivF2C1    0.491452   0.176957   2.777  0.00571 **
EduDivF2C2    0.393344   0.350355   1.123  0.26215
EduDivF2C3    0.053638   0.215475   0.249  0.80353
EduDivF2C4    0.012460   0.288526   0.043  0.96557
IndDivF      -0.334950   0.769581  -0.435  0.66360
I(IndDivF^2) -0.017845   0.244067  -0.073  0.94175
OrgDivF       0.540781   0.385586   1.402  0.16145
I(OrgDivF^2)  0.014889   0.110043   0.135  0.89243
OccDivF2C1    0.250319   0.161053   1.554  0.12081
OccDivF2C2    0.298402   0.318667   0.936  0.34956
OccDivF2C3    0.565073   0.229968   2.457  0.01437 *
OccDivF2C4    0.581189   0.321108   1.810  0.07096 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.291 on 456 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.2412,    Adjusted R-squared:  0.1963
F-statistic: 5.369 on 27 and 456 DF,  p-value: 2.671e-15
```

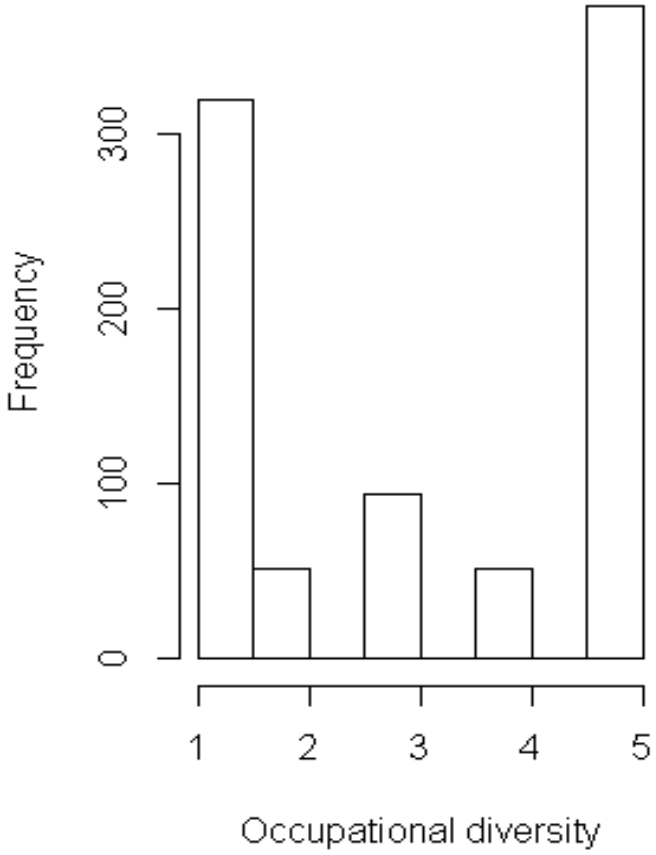
Significance of categorical variables

```
Single term deletions

Model:
Saleslog ~ Year + NVT + Age + Gen + Eth + EduExp + IndExp + OrgExp +
    OccExp + EduDivF2C + IndDivF + I(IndDivF^2) + OrgDivF + I(OrgDivF^2) +
    OccDivF2C
            Df Sum of Sq    RSS    AIC  Pr(>Chi)
<none>                   759.69 274.20
Year         7   18.3551 778.05 271.76  0.116177
NVT          1    1.3001 760.99 273.03  0.362964
Age          1   25.9255 785.62 288.44 5.576e-05 ***
Gen          1   15.3208 775.01 281.87  0.001879 **
Eth          1    4.5062 764.20 275.06  0.090670 .
EduExp       1    0.5033 760.20 272.52  0.571283
IndExp       1    2.7836 762.48 273.97  0.183362
OrgExp       1    0.4356 760.13 272.48  0.598363
OccExp       1    2.0472 761.74 273.50  0.253754
EduDivF2C    4   15.6646 775.36 276.08  0.042527 *
IndDivF      1    0.3156 760.01 272.40  0.653898
I(IndDivF^2) 1    0.0089 759.70 272.21  0.939955
OrgDivF      1    3.2770 762.97 274.29  0.148920
I(OrgDivF^2) 1    0.0305 759.72 272.22  0.889138
OccDivF2C    4   12.9717 772.67 274.40  0.084707 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Histogram of Educational diversity



Histogram of Occupational diversity

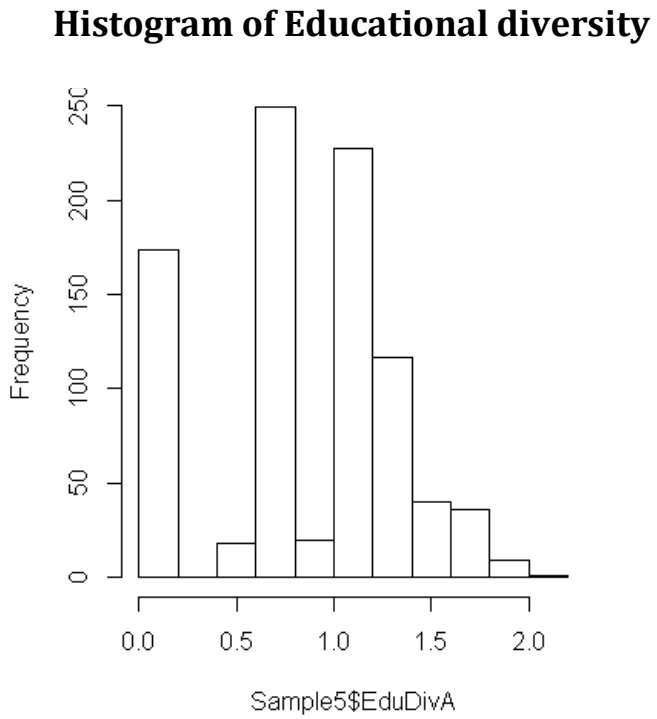


Control variables NVT

Discussion: due to the interaction between NVT and the diversity dimensions, all the models are also estimated without the NVT measures. In every single model the explained variance is lower in case of exclusion of the NVT size variable. Moreover, the effect on the other variables is rather limited. In some cases the significance of Occupational and Educational diversity is slightly lower; in others this is slightly higher. Overall there is no reason to exclude NVT from the model, because the variable does not appear to interfere with the results (as was indicated by the VIF values as well).

Separate Educational Operationalization

Discussion: Because single individuals can have more educational experiences, in this case not only the last experience but all experiences are taken into account. However, because each case can only be assigned to one category, I see each educational status as a different design. Because an individual can only have one occurrence for an educational category the balance requirements are still met. However, keep in mind that all teams with zero or one educational status are considered completely homogeneous. Because the diversity in this case is not restricted by the size of the NVT, I do not have to correct for a theoretical maximum value (this is supported by the spread of diversity for NVTs with 2 members). The correlation with NVT remains of .4, but this does not lead to multicollinearity problems. As with the measure that only considers the last education, this measure is negatively related to the Sales of the new venture and the effect appears slightly more significant (it remains significant in squared model). However, the explained variance for this method is lower than for the relative educational measure with only the last status being taken into account (with which it has a .5 correlation).



Correlation matrix of different Educational diversity measures and NVT size

```
          NVT    EduDivF EduDivF2 EduDivA EduDivA2
NVT       1.00    0.43    -0.12    0.35     0.01
EduDivF   0.43    1.00     0.80    0.48     0.50
EduDivF2 -0.12    0.80     1.00    0.30     0.55
EduDivA   0.35    0.48     0.30    1.00     0.76
EduDivA2  0.01    0.50     0.55    0.76     1.00

n= 891


P
         NVT    EduDivF EduDivF2 EduDivA EduDivA2
NVT             0.0000  0.0002   0.0000  0.7133
EduDivF  0.0000         0.0000   0.0000  0.0000
EduDivF2 0.0002 0.0000           0.0000  0.0000
EduDivA  0.0000 0.0000  0.0000           0.0000
EduDivA2 0.7133 0.0000  0.0000   0.0000
>
```

Linear Survival Model

```
Call:
lm(formula = Sur ~ Year + NVT + Age + Gen + Eth + EduExp + IndExp +
    OrgExp + OccExp + EduDivA + IndDivF + OrgDivF + OccDivF2,
    data = Sample5, na.action = na.omit)

Residuals:
    Min      1Q  Median      3Q     Max
-0.8703 -0.5268  0.2938  0.4129  0.6723

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0130964  0.1766110   0.074  0.94090
Year2003     0.1309971  0.0806733   1.624  0.10478
Year2004     0.0804865  0.0738760   1.089  0.27624
Year2005     0.0921789  0.0737014   1.251  0.21138
Year2006     0.1532178  0.0715530   2.141  0.03253 *
Year2007     0.1877879  0.0699903   2.683  0.00743 **
Year2008     0.0884557  0.0782014   1.131  0.25831
Year2009     0.0267835  0.0876548   0.306  0.76002
NVT          0.0635948  0.0212969   2.986  0.00291 **
Age          0.0018466  0.0022448   0.823  0.41096
Gen          0.1112058  0.0703339   1.581  0.11422
Eth          0.1180149  0.1107513   1.066  0.28691
EduExp      -0.0180920  0.0401294  -0.451  0.65222
IndExp      -0.0009281  0.0021557  -0.431  0.66693
OrgExp       0.0140000  0.0442522   0.316  0.75180
OccExp       0.0096707  0.0433530   0.223  0.82353
EduDivA      0.0036981  0.0387862   0.095  0.92406
IndDivF     -0.0284528  0.0571755  -0.498  0.61886
OrgDivF      0.0554399  0.0341758   1.622  0.10512
OccDivF2    -0.0515531  0.0416163  -1.239  0.21577
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 0.4875 on 866 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.04325,   Adjusted R-squared:  0.02225
F-statistic:  2.06 on 19 and 866 DF,  p-value: 0.004889
```

Squared Survival Model

```
Call:
lm(formula = Sur ~ Year + NVT + Age + Gen + Eth + EduExp + IndExp +
    OrgExp + OccExp + EduDivA + I(EduDivA^2) + IndDivF +
I(IndDivF^2) +
    OrgDivF + I(OrgDivF^2) + OccDivF + I(OccDivF^2), data = Sample5,
    na.action = na.omit)

Residuals:
    Min      1Q  Median      3Q     Max
-0.8199 -0.5325  0.2821  0.4125  0.7026

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.0669395  0.2191592   0.305  0.76011
Year2003      0.1217010  0.0807420   1.507  0.13210
Year2004      0.0838667  0.0741378   1.131  0.25827
Year2005      0.1005285  0.0737681   1.363  0.17331
Year2006      0.1566284  0.0716134   2.187  0.02900 *
Year2007      0.1923410  0.0700282   2.747  0.00615 **
Year2008      0.1008904  0.0784823   1.286  0.19896
Year2009      0.0234126  0.0880823   0.266  0.79045
NVT           0.0731922  0.0238112   3.074  0.00218 **
Age           0.0015243  0.0022546   0.676  0.49918
Gen           0.1177343  0.0703932   1.673  0.09478 .
Eth           0.1139179  0.1115833   1.021  0.30758
EduExp       -0.0225428  0.0400827  -0.562  0.57398
IndExp       -0.0003361  0.0021985  -0.153  0.87853
OrgExp        0.0075441  0.0442554   0.170  0.86468
OccExp        0.0124199  0.0440219   0.282  0.77791
EduDivA      -0.1198981  0.0971317  -1.234  0.21740
I(EduDivA^2)  0.0801801  0.0599173   1.338  0.18119
IndDivF      -0.2617834  0.2043057  -1.281  0.20042
I(IndDivF^2)  0.0759393  0.0654129   1.161  0.24599
OrgDivF       0.2547886  0.0923728   2.758  0.00593 **
I(OrgDivF^2) -0.0646335  0.0276947  -2.334  0.01984 *
OccDivF      -0.0685841  0.1150075  -0.596  0.55110
I(OccDivF^2)  0.0087690  0.1078241   0.081  0.93520
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4866 on 862 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.05144,   Adjusted R-squared:  0.02613
F-statistic: 2.032 on 23 and 862 DF,  p-value: 0.00293

>
```

Linear Sales Model

```
Call:
lm(formula = Saleslog ~ Year + NVT + Age + Gen + Eth + EduExp +
    IndExp + OrgExp + OccExp + EduDivA + IndDivF + OrgDivF +
    OccDivF2C, data = Sample5Sur, na.action = na.omit)

Residuals:
    Min      1Q  Median      3Q     Max
-5.3860 -0.5844  0.0898  0.6999  3.8706

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.753742   0.742619  19.867  < 2e-16 ***
Year2003    -0.026245   0.299394  -0.088   0.9302
Year2004     0.409231   0.283386   1.444   0.1494
Year2005     0.167592   0.284530   0.589   0.5561
Year2006     0.443439   0.276012   1.607   0.1088
Year2007     0.162643   0.265733   0.612   0.5408
Year2008     0.339505   0.301334   1.127   0.2605
Year2009    -0.239823   0.359443  -0.667   0.5050
NVT          0.135307   0.099032   1.366   0.1725
Age         -0.034431   0.008568  -4.018 6.84e-05 ***
Gen          0.859589   0.269161   3.194   0.0015 **
Eth         -0.856813   0.460388  -1.861   0.0634 .
EduExp      -0.024106   0.145691  -0.165   0.8687
IndExp       0.011363   0.007604   1.494   0.1358
OrgExp       0.084449   0.158233   0.534   0.5938
OccExp       0.143206   0.152247   0.941   0.3474
EduDivA     -0.326537   0.140565  -2.323   0.0206 *
IndDivF     -0.276884   0.214267  -1.292   0.1969
OrgDivF      0.569990   0.132233   4.311 1.99e-05 ***
OccDivF2C1   0.357682   0.153845   2.325   0.0205 *
OccDivF2C2   0.347336   0.307782   1.129   0.2597
OccDivF2C3   0.543528   0.219786   2.473   0.0138 *
OccDivF2C4   0.684000   0.308463   2.217   0.0271 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.29 on 461 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.2342,    Adjusted R-squared:  0.1976
F-statistic: 6.408 on 22 and 461 DF,  p-value: < 2.2e-16
```

Squared Sales Model

```
Call:
lm(formula = Saleslog ~ Year + NVT + Age + Gen + Eth + EduExp +
    IndExp + OrgExp + OccExp + EduDivA + I(EduDivA^2) + IndDivF +
    I(IndDivF^2) + OrgDivF + I(OrgDivF^2) + OccDivF + I(OccDivF^2),
    data = Sample5Sur, na.action = na.omit)

Residuals:
    Min      1Q  Median      3Q     Max
-5.4125 -0.6125  0.0828  0.7626  3.6898

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  14.862328   0.841019  17.672  < 2e-16 ***
Year2003      0.027610   0.303708   0.091 0.927605
Year2004      0.489903   0.286879   1.708 0.088367 .
Year2005      0.246683   0.287529   0.858 0.391372
Year2006      0.521438   0.278922   1.869 0.062191 .
Year2007      0.202682   0.267802   0.757 0.449536
Year2008      0.400327   0.305275   1.311 0.190389
Year2009     -0.174385   0.367464  -0.475 0.635324
NVT           0.288141   0.084671   3.403 0.000725 ***
Age          -0.034066   0.008662  -3.933  9.7e-05 ***
Gen           0.879513   0.272126   3.232 0.001317 **
Eth          -0.812185   0.469990  -1.728 0.084643 .
EduExp       -0.026003   0.145550  -0.179 0.858288
IndExp        0.010085   0.007874   1.281 0.200880
OrgExp        0.062738   0.159663   0.393 0.694548
OccExp        0.130060   0.156106   0.833 0.405192
EduDivA      -0.610563   0.358925  -1.701 0.089602 .
I(EduDivA^2)  0.170984   0.216994   0.788 0.431121
IndDivF      -0.513797   0.768856  -0.668 0.504302
I(IndDivF^2)  0.074198   0.242490   0.306 0.759755
OrgDivF       0.528137   0.388661   1.359 0.174856
I(OrgDivF^2)  0.017762   0.110511   0.161 0.872377
OccDivF      -0.177991   0.413598  -0.430 0.667144
I(OccDivF^2) -0.091254   0.378134  -0.241 0.809409
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.3 on 460 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.2236,    Adjusted R-squared:  0.1848
F-statistic:  5.76 on 23 and 460 DF,  p-value: 6.26e-15
```

Two step Heckman Approach

Discussion: This method is calculated to correct for the sampling bias that might results from only including the surviving ventures in the Sales models. The results of this approach are largely similar to the seperate approach used before, indicating that there is no large selection problem. Small changes that occur are slightly lower significane of Occ and Org dimensions in the squared models. This might however also be the result of the use of a probit estimation (standard in Two step Heckman) instead of a Logit estimation. Furthermore, due to the fact that all surviving cases are automatically selected for the second step, I am not able to control for takeovers. Takeover new ventures lack the sales data and would distort the results. This approach is less precise in this respect. Moreover, there are no way of controlling the significance of entire categorical variables. Concludingly, the Two step Heckman approach shows largely similar results but has some disadvantages with respect to the traditional method. Because no indication for severe sampling bias exists, the traditional approach is preferred.

Linear model

```
--------------------------------------------
Tobit 2 model (sample selection model)
Maximum Likelihood estimation
Newton-Raphson maximisation, 18 iterations
Return code 1: gradient close to zero
Log-Likelihood: -1385.858
886 observations (402 censored and 484 observed)
48 free parameters (df = 838)
Probit selection equation:
             Estimate Std. error t value Pr(> t)
(Intercept) -1.621485   0.494767  -3.277 0.00105 **
Year2003     0.411627   0.212107   1.941 0.05230 .
Year2004     0.189318   0.193656   0.978 0.32827
Year2005     0.184877   0.192329   0.961 0.33642
Year2006     0.338748   0.187590   1.806 0.07095 .
Year2007     0.449021   0.184591   2.433 0.01499 *
Year2008     0.125951   0.204566   0.616 0.53809
Year2009    -0.098527   0.229932  -0.429 0.66828
NVT          0.180067   0.074161   2.428 0.01518 *
Age          0.005299   0.005851   0.906 0.36512
Gen          0.223443   0.184075   1.214 0.22480
Eth          0.337183   0.291378   1.157 0.24719
EduExp      -0.043743   0.099195  -0.441 0.65923
IndExp       0.003146   0.005764   0.546 0.58515
OrgExp       0.127823   0.116159   1.100 0.27115
OccExp       0.128522   0.113671   1.131 0.25820
EduDivF2     0.051426   0.118249   0.435 0.66364
IndDivF     -0.011302   0.148056  -0.076 0.93915
OrgDivF      0.081050   0.089785   0.903 0.36668
OccDivF2C1   0.169850   0.111439   1.524 0.12747
OccDivF2C2  -0.093102   0.242505  -0.384 0.70104
OccDivF2C3   0.118744   0.164581   0.721 0.47061
OccDivF2C4  -0.153122   0.238848  -0.641 0.52147
Outcome equation:
             Estimate Std. error t value  Pr(> t)
(Intercept) 15.076734   1.048249  14.383  < 2e-16 ***
Year2003     0.013221   0.309010   0.043  0.96587
Year2004     0.456469   0.280402   1.628  0.10354
Year2005     0.222666   0.280886   0.793  0.42794
Year2006     0.509161   0.282075   1.805  0.07107 .
Year2007     0.209798   0.281274   0.746  0.45574
Year2008     0.402831   0.294559   1.368  0.17145
Year2009    -0.208920   0.351069  -0.595  0.55178
NVT          0.078537   0.104248   0.753  0.45123
Age         -0.033852   0.008442  -4.010 6.07e-05 ***
Gen          0.830454   0.268245   3.096  0.00196 **
Eth         -0.779299   0.455712  -1.710  0.08725 .
EduExp      -0.089332   0.132559  -0.674  0.50037
IndExp       0.010460   0.007438   1.406  0.15963
OrgExp       0.075286   0.156921   0.480  0.63139
OccExp       0.179585   0.151163   1.188  0.23483
EduDivF2    -0.498012   0.166577  -2.990  0.00279 **
IndDivF     -0.386825   0.206700  -1.871  0.06129 .
OrgDivF      0.586734   0.130608   4.492 7.05e-06 ***
OccDivF2C1   0.245339   0.160739   1.526  0.12693
OccDivF2C2   0.319460   0.300182   1.064  0.28723
OccDivF2C3   0.512193   0.216726   2.363  0.01811 *
OccDivF2C4   0.632301   0.302694   2.089  0.03672 *
Error terms:
      Estimate Std. error t value Pr(> t)
sigma  1.25447    0.04059  30.909  <2e-16 ***
rho    0.01840    0.33397   0.055   0.956
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
--------------------------------------------
>
```

Squared model

```
--------------------------------------------
Tobit 2 model (sample selection model)
Maximum Likelihood estimation
Newton-Raphson maximisation, 12 iterations
Return code 1: gradient close to zero
Log-Likelihood: -1373.15
886 observations (402 censored and 484 observed)
54 free parameters (df = 832)
Probit selection equation:
                Estimate Std. error t value Pr(> t)
(Intercept)   -1.2485465  0.5661442  -2.205  0.0274 *
Year2003       0.3487293  0.2081048   1.676  0.0938 .
Year2004       0.1385215  0.1904003   0.728  0.4669
Year2005       0.2682209  0.1901698   1.410  0.1584
Year2006       0.3346042  0.1852813   1.806  0.0709 .
Year2007       0.4258987  0.1806809   2.357  0.0184 *
Year2008       0.1731100  0.2011704   0.861  0.3895
Year2009      -0.0232224  0.2257825  -0.103  0.9181
NVT            0.1407778  0.0838434   1.679  0.0931 .
Age            0.0050086  0.0058495   0.856  0.3919
Gen            0.1999854  0.1794324   1.115  0.2650
Eth            0.3521899  0.2889240   1.219  0.2229
EduExp        -0.0668334  0.0966582  -0.691  0.4893
IndExp        -0.0008152  0.0056113  -0.145  0.8845
OrgExp         0.1299521  0.1128604   1.151  0.2496
OccExp         0.1209820  0.1119280   1.081  0.2797
EduDivF2       0.4170056  0.6127728   0.681  0.4962
I(EduDivF2^2) -0.3100167  0.5922097  -0.523  0.6006
IndDivF       -0.7573253  0.5194855  -1.458  0.1449
I(IndDivF^2)   0.2126330  0.1663365   1.278  0.2011
OrgDivF        0.5466332  0.2430579   2.249  0.0245 *
I(OrgDivF^2)  -0.1355058  0.0725457  -1.868  0.0618 .
OccDivF2C1     0.1632399  0.1094920   1.491  0.1360
OccDivF2C2    -0.0521280  0.2371142  -0.220  0.8260
OccDivF2C3     0.0779082  0.1638371   0.476  0.6344
OccDivF2C4    -0.0862129  0.2328688  -0.370  0.7112
Outcome equation:
               Estimate Std. error t value  Pr(> t)
(Intercept)   17.578335   0.990384  17.749  < 2e-16 ***
Year2003      -0.340041   0.344625  -0.987 0.323790
Year2004       0.267717   0.322377   0.830 0.406287
Year2005       0.117102   0.322148   0.364 0.716227
Year2006       0.231092   0.313826   0.736 0.461507
Year2007      -0.180073   0.304242  -0.592 0.553935
Year2008       0.332536   0.340449   0.977 0.328690
Year2009       0.021492   0.397482   0.054 0.956879
NVT           -0.013890   0.132056  -0.105 0.916231
Age           -0.036453   0.009772  -3.730 0.000191 ***
Gen            0.595097   0.304895   1.952 0.050961 .
Eth           -1.109581   0.511709  -2.168 0.030130 *
EduExp        -0.054052   0.157729  -0.343 0.731836
IndExp         0.004004   0.009098   0.440 0.659897
OrgExp        -0.009457   0.183089  -0.052 0.958806
OccExp         0.042727   0.178213   0.240 0.810524
EduDivF2      -1.317371   0.983163  -1.340 0.180267
I(EduDivF2^2)  0.796706   0.947090   0.841 0.400228
IndDivF       -0.061798   0.871943  -0.071 0.943498
I(IndDivF^2)  -0.124904   0.277438  -0.450 0.652563
OrgDivF        0.179198   0.425631   0.421 0.673742
I(OrgDivF^2)   0.119873   0.123487   0.971 0.331678
OccDivF2C1     0.078503   0.182070   0.431 0.666345
OccDivF2C2     0.384482   0.364030   1.056 0.290886
OccDivF2C3     0.440442   0.261742   1.683 0.092427 .
OccDivF2C4     0.679483   0.364957   1.862 0.062629 .
Error terms:
      Estimate Std. error t value Pr(> t)
sigma  1.69294    0.08668   19.53  <2e-16 ***
rho   -0.87938    0.03089  -28.46  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
--------------------------------------------
```

This page intentionally left blank