Two's Company: Directional Diversity and Performance of Entrepreneurial Pairs

Alex Coad Bram Timmermans

August 29, 2012

"PRELIMINARY AND INCOMPLETE VERSION FOR WORKSHOPS DEADLINE – DO NOT CITE, DO NOT QUOTE, DO NOT DISTRIBUTE"

Abstract

We study the effects of diversity on the survival and growth of new ventures using the Danish Linked Employer-Employee database. To get cleaner measures of diversity, we focus on entrepreneurial dyads, and also investigate the asymmetric effects of 'directed' diversity by distinguishing between the 'primary' and the 'secondary' founder. We complement existing work by showing that diversity depends on the asymmetric hierarchical structure within the team, and that a unidimensional diversity indicator (which is usually applied) fails to capture a number of performance effects of diversity. Ventures with a STEM-educated primary founder and a Business-educated secondary founder have high employment growth, while the opposite combination (Business first, STEM second) has low employment growth. Pairs of younger individuals have lower survival chances but higher employment growth. Performance of mixed-race and mixed-gender ventures depends upon the identity of the primary founder. Family firms have lower employment growth.

JEL codes: L25, L26, J21

Keywords: Diversity, Entrepreneurial Teams, Survival, Growth

1 Introduction

From the late 1980s forward, we have observed a gradual shift from treating entrepreneurship as an act of one individual towards entrepreneurship as a collective activity (Cooney, 2005; Harper, 2008).[1] Consequently, researchers within the field of entrepreneurship started to investigate the composition of these entrepreneurial teams (e.g. Ruef et al., 2003). In close relation to identifying this composition, here is also an interest in investigating whether the particular composition of these entrepreneurial teams affects the performance of these teams, which varies from member entry and exit to growth and survival, and if so what the nature of this relationship is. Studies have focused on various compositional measures of teams; for example, industry experience (Delmar and Shane, 2006) entrepreneurial experience(Ucbasaran et al., 2003; Delmar and Shane, 2006), and prior affiliations of team members (Beckman et al., 2007). However, inspired by the upper echelon theories on diversity in top management teams (Hambrick and Mason, 1984), there has been an increased focus on diversity in entrepreneurial teams arguing that the heterogeneity of these teams affects how they work together, which ultimately will affect their performance.

Not surprisingly, and in accordance with studies on top management teams, the impact of diversity is inconclusive. Studies have not specified whether, how and why team diversity affects positively or negatively the performance of start-ups. This can attributed to: (i) the assumption that the same level of diversity might have a different impact on various performance indicators, i.e. diversity might be good for firm growth but bad for firm survival, (ii) that the approach in investigating diversity is too ambitious by constructing an overall (unidirectional) diversity measure, and (iii) studies investigate entrepreneurial teams of different sizes, thereby introducing an undesirable level of complexity making it even more difficult to estimate the impact of diversity on performance. In this paper, we will focus on a subset of entrepreneurial teams – partnerships. Despite that we focus on such a subset, this is the most common form of start-up (even more common than single-person start-ups; furthermore, this narrow focus allows us to explore the effects of diversity in great detail. In addition, instead of treating diversity as a unidirectional compositional measure, we will take into account the direction of this diversity.

To investigate the issue, we use the Danish Integrated Database for Labour Market Research (IDA) to identify these entrepreneurial pairs. This database provides

[1] Decades earlier, researchers already started to promote the notion of entrepreneurial teams but the work of Kamm et al. (1990) and Gartner et al. (1994) called for a more systemic approach in studying the performance of entrepreneurial teams (Ruef, 2010).

detailed information on the demographic characteristics of individuals, e.g. age, gender, education, and the dynamics of organization, i.e. birth, growth and exit of firms, which allows us to analyse the relation that exist between the demographic characteristic of the entrepreneurial team and new venture performance. The direction of the diversity is determined by the position of the individuals in the firm where a higher ranked individual, based on ownership and occupation code, is considered to be the *primus motor* of the start-up. We select a sample of 4,219 start-ups in the Danish private sector in the period 1999-2003 and follow these start-ups up to five year after founding.

We contribute to the literature in a number of ways. We investigate the effect of diversity on performance using an especially rich dataset that contains details on a number of variables including educational background and family ties. While much previous work has focused on small samples, we provide representative large-sample evidence using administrative data. We investigate the performance of new businesses in terms of both survival and employment growth. While previous work has grouped together ventures of different ages, we observe new ventures from their first year of business (as indicated by their date of official registration). In response to calls for diversity research to focus more on dynamic effects (Horwitz and Horwitz, 2007), we exploit our longitudinal data to consider lagged effects of diversity (that is, the effects of diversity of start-up and pre-start characteristics on five-year performance). We focus on entrepreneurial pairs only, in order to get a cleaner measure of diversity (Harper, 2008). We also distinguish ourselves from the literature that uses unidimensional diversity indicators to have richer measures of diversity. Furthermore, we contribute to the literature by moving on from assuming that power relations are symmetric between team members – we distinguish between the primary and secondary founder, and investigate which characteristics matter for each of the two founders.

The preliminary analysis indicate that when focusing on entrepreneurial pairs. There is indeed a difference in performance depending on the direction of this diversity. With regards to education, the best performing firms are not composed of similar individuals. Ventures with a STEM-educated primary founder and a Business-educated secondary founder enjoy relatively high employment growth, while interestingly enough the opposite combination (Business first, STEM second) has conspicuously low employment growth. Pairs of younger individuals have lower survival chances but higher employment growth. Performance of mixed-race and mixed-gender ventures depends upon the identity of the primary founder. Family firms have equal survival chances but lower employment growth – consistent with suggestions that they persist for an unnecessarily long period of time.

The remainder of the paper is structured as follows. We review the related literature in Section 2. Our methodology is described in Section 3. We present our data in Section 4. Section 5 contains our analysis, where we begin with non-parametric representations of diversity and performance before moving on to parametric regressions. Section 6 contains a synthetic discussion of our findings and revisits our hypotheses. Section 7 concludes.

2 Overview of the literature

Issues on team diversity are not a new phenomenon; on the contrary, a survey of the literature indicates that there exists a long tradition in linking the diverse composition of teams with their performance (see, e.g., Williams and O'Reilly (1998) and Horwitz (2005) for a literature review). However, a closer inspection of these studies reveals that the interest is traditionally based upon teams in larger organizational settings, e.g. top management and product development teams (Murray, 1989; Bantel and Jackson, 1989; Ancona and Caldwell, 1992; Pelled, 1996; Dahlin et al., 2005). More recently, studies that investigate the diverse composition of entrepreneurial teams have emerged and an increase in the number of studies is visible. This steady increase runs parallel with the increased focus of entrepreneurial teams in general (Cooney, 2005; Harper, 2008).

Studies that investigate the composition of entrepreneurial teams (e.g. Baron et al., 1999; Ruef et al., 2003; Steffens et al., 2011) show that entrepreneurial teams[2] are mainly characterized by homophily, at least regarding gender, ethnicity and occupation (more visible characteristics), while we can observe more heterogeneity in terms of functionality and status. The homophily in these teams can be explained by the be social selection mechanism behind recruitment that often are driven by interpersonal attraction (Forbes et al., 2006); not only because they rely on social networks (Aldrich and Langton, 1998; Aldrich and Ruef, 2006), which are homogeneous (McPherson et al., 2001), but also based on the other recruitment channels. The underlying rationale is that interpersonal attraction based on the demographic attributes will cause less (personal) trouble in start-ups (Beckman et al., 2007); consequently, the limited resources will be used to deal with the liability issues that start-ups face.

[2]The same also holds for other organizational units that rely on voluntary participation (McPherson et al., 2001).

2.1 Competing Views on Diversity and Performance

Despite the underlying reasoning for why homogeneity is beneficial for entrepreneurial teams there is not much consistency related to the effects of diversity in the diversity and entrepreneurial team literature. This inconsistency is nicely illustrated by the selection of studies on diversity in entrepreneurial teams and the impact on various performance indicators in Table 1.

The underlying premise on why diversity affects performance is similar to the arguments put forward by studies on top management teams. Overall, one can make a distinction between two competing approaches with regard to the expected effect of diversity on team performance (Williams and O'Reilly, 1998; Horwitz, 2005; Horwitz and Horwitz, 2007). The first approach has its roots in social psychology, represented by the social categorization and social identity theories (Tajfel, 1981; Turner, 1987), and the similarity-attraction paradigm (Byrne, 1971). These approaches associate diversity with conflict and consequently a negative impact on performance. The approach that claims the opposite is referred to as cognitive resource diversity theory (Cox and Blake, 1991). This approach assumes a positive effect of diversity on the performance of teams.

In the occurrence of social categorization, individuals place themselves and others in various categories rather than treating them as separate individuals; a process that most often occurs in groups that are heterogeneous (Williams and O'Reilly, 1998; Joshi and Jackson, 2003). In this case, it is not the categorization that is problematic but rather the interaction between individuals as a result of this categorization. This behavior is often referred to as social identity and similarity-attraction. With social identify, individuals created in-group and out-group membership. Eventually, individuals will create a more favorable bias toward those that are part of the in-group and out-group members are regarded as less attractive, trustworthy, honest, and cooperative, which will lead to conflict (Joshi and Jackson, 2003). The similarity attraction paradigm argues for the interpersonal attraction that arises from the similarities that exist between the different group members. This attraction is a result of shared experiences and values, which eases the communication and interaction between members and enhances cohesiveness; in addition to the stronger connection within the group there might, just as in social identity theory, occur a dislike of members in other groups (Horwitz, 2005). However, one has to consider that high levels of diversity and conflict risky constellations will be avoided due to the recruitment process; consequently, this results in the earlier mentioned rather homogeneous composition of teams in general and entrepreneurial teams in particular.

Table 1: Studies on Entrepreneurial Teams and Diversity

Study	Issue	No. Teams	Effect of diversity
Chowdhury (2005)	This study investigates the influence of diversity in (a) age, (b) gender, and (c) functional background on entrepreneurial team effectiveness	79	(a)non-significant (b) non-significant (c) non-significant
Ensley et al. (1998)	This study investigates the effect of diversity in the entrepreneurial teams' skill composition on (a) sales growth, (b) profitably, and (c) revenues in new ventures	88	(a)negative (b) non-significant (c) negative
Beckman (2006)	This study investigates the influence of diversity in previous affiliation on the (a) exploration and (b) exploitation strategy of firms and (c) products speed to market and (d) firm growth.	170	(a) positive (b) non-significant (c) non-significant (d) only positive in combination with shared prior affiliation
Beckman et al. (2007)	This study investigates the impact of functional diversity, turnover and background affiliation of founding and early top management team composition on (a) the ability to attract venture capital (b) ability to successfully complete an initial public offering.	161	(a) positive (b) positive
Ucbasaran et al. (2003)	This study examines the impact of functional diversity on the (a) entry and (b) exit of team members.	92	(a) non-significant (b) non-significant
Foo et al. (2005)	The study examines how team diversity, making a distinction between (a) task and (b) non-task related diversity, affects external evaluation of the teams business ideas	154	(a) positive and (b) negative
Chandler et al. (2004)	This study investigates impact of diversity in (a) gender (b) religion (c) political affiliation (d) education, (e) industry experience, and (f) functional diversity of the initial team on on the (1) addition and (2) departure of team members.	124	(a1) non-significant, (b1) non-significant, (c1) non-significant, (d1) positive, (e1) positive, (f1) positive, (a2) non-significant, (b2) positive, (c2) non-significant (d2) non-significant, (e2) positive, (f2) non-significant
Watson et al. (2003)	This study tests how differences between partners on human capital, organizational demography, and interpersonal processes affects the partners' perceptions of firm (a) profit and (b) growth.	175	(a) non-significant (b) non-significant
Amason et al. (2006)	This study investigate the impact of diversity in (a) age, (b) level of education, (c) specialization of education, and (d) functional background on sales growth profitability and market performance. In this study the diversity measures are interacted with the novelty of the products and services offered.	174	Overall non-significant but interacting novelty with high novelty there is a negative effect of diversity.
Steffens et al. (2011)	This study investigates how more homogeneity (less diversity) in (a) start-up experience, (b) age and (c) gender affects (1) short term performance and (2) long run performance.	202	(a1) non-significant, (b1) non-significant, (c1) non-significant (a2) positive, (b2) positive, (c2) non-significant.
Eisenhardt and Schoonhoven (1990)	This research investigates how heterogeneity in of industry experience of high tech founding teams affects organizational growth.	92	positive.

In contrast to the above-mentioned approaches stands cognitive resource diversity theory. Scholars that are in favor of this approach argue for the positive impact of diversity on performance as a result of the unique set of skills, abilities and knowledge that are brought into the team (Cox and Blake, 1991; Hambrick et al., 1996; Williams and O'Reilly, 1998; Horwitz, 2005). This line of argument is similar to other approaches within management theory, in particular the resource based view of the firm, which argues that a heterogeneous resource composition, including human resources, determines a firm's competitive advantage (Barney, 1991).

One domain within management that has paid much attention to the positive effect of diversity on performance is the upper echelon studies on top management teams, which have their origin in Hambrick and Mason (1984). In this stream of literature, it is widely accepted that it is important that these teams collectively possess the skills that are necessary to run a successful business (Beckman et al., 2007). The majority of studies on entrepreneurial teams share this perspective as their superior performance compared to solo entrepreneurs is believed to be driven by the access to various forms of human capital and the presence of different perspectives (Kamm et al., 1990; Eisenhardt and Schoonhoven, 1990; Watson et al., 1995). Nevertheless, most studies on these entrepreneurial teams focus on human capital theory and look at overall team characteristics (e.g. average level of education or length of experiences) to explain performance (Beckman et al., 2007). This approach is very helpful in explaining the performance of individual entrepreneurs (Davidsson and Honig, 2003) but fails to capture the impact of the diversity in skills that are present in a collective. By adopting an organizational demography approach, we can consider both the average characteristics of the human resources in ventures and the differences between the human resources (Beckman et al., 2007).

The above-mentioned theoretical approaches provide sound but contradictory arguments on the potential effect of team diversity on team performance. It is therefore not surprising that empirical studies have found both positive, negative and non-significant effects of diversity in teams. However, another, rather unexplored perspective on why there is so much ambiguity of diversity in entrepreneurial teams might be because of the rather ambitious way in which studies have treated diversity. In the remainder of this paper we will address this issue empirically by investigating how a diverse composition in a particular subset of entrepreneurial teams – partnerships– affects new venture performance.

3 Method

In the majority of the above-mentioned studies on team diversity, diversity is defined as the distribution of differences among team members with respect to a common attribute. Consequently, diversity is often regarded as a unit-level compositional construct (Harrison and Klein, 2007). Overall, diversity on these attributes can be measured on three dimensions, i.e. variety, balance, disparity (Stirling, 2001; Harrison and Klein, 2007). Variety takes into account the number of categories within a certain attribute where more categories result in higher diversity. With balance the shares of the specific category are measured and a more equal balance between categories results in a higher degree of diversity. Disparity refers to the distance between the outer boundaries of the various categories within one characteristic. Harrison and Klein (2007) distinguished between separation and disparity where the first relates to horizontal differences, i.e. diversity based on opinions or expertise, and the latter on vertical differences, i.e. diversity based on hierarchy or power. To study this diversity, we will follow the methods proposed in the existing work on team diversity. The majority of these studies have used the techniques of organizational demography. This means that the level of diversity is measured based on observable demographic characteristics, where demography is defined as: "the composition, in terms of basic attributes such as age, sex, educational level, length of service, race and so forth of the social unit under study" (Pfeffer, 1983, p. 303).

Such an empirical strategy leads to several challenges when investigating the impact of diversity on the performance of the team. First, researchers create one overall, undirectional, measure of diversity for each attribute; this approach does not take into account that diversity might in reality have a directional character. Second, there is the challenge on how to find a concise representation of the high dimensionality, i.e. large teams have more nodes leading to a higher level of complexity.

3.1 Focus on pairs only

To keep the dimensionality manageable, we focus on entrepreneurial pairs. Focusing on entrepreneurial dyads is a meaningful way of simplifying the analysis of entrepreneurial teams (Harper, 2008). With pairs, there is only one possible relationship in which diversity can be measured – that is, the relationship of A to B. With triads, one may look at the diversity between A and B, or A and C, or B and C; and the analysis of diversity becomes even more complex with four or more founders.

Another main reason why we focus on pairs is that, contrary to other studies that investigate entrepreneurial team performance, we consider that entrepreneurial teams of different sizes are qualitatively different. In pairs, for example, there is always the tension of a head-on conflict, and disputes are resolved essentially through the mechanism of 'my word against yours.' In keeping with insights from geometry (that is, the stability of triangular structures), an entrepreneurial team of three founders will have more stability as the dynamics of majority rule is more flexible, with each individual taking turns as the swing voter and arbiter, and being able to move from side to side to form new majority coalitions with one of the two others. With teams of four individuals, there may be a tendency to split into rival groups (of pairs) within the team, for individuals to seek strong pair-bonds within the team, or for minority views to acquiesce relatively easily. In short, there may be nonlinearities between number of team members and the nature of diversity within the team, because integers can be seen as being qualitatively different (Schimmel, 1994). Teams of different sizes have fundamentally different opportunities for specialization, that do not scale up with team size in a linear way. To keep our observations as comparable as possible, we focus only on the most numerous team-size, which is the team of two individuals.

3.2 Quantifying diversity

Table 2: Indicators of diversity used in the literature

Indicator	Formula	Types of variables	Examples
Coefficient of variation	$c_v = \frac{\sigma}{\mu}$	Continuous	Pelled et al. (1999, p11); Foo et al. (2005, p393)
Herfindahl-Hirschman index	$H = 1 - \sum_{l=1}^{l}(P_i)^2$	Categorical	Pelled et al. (1999, p11); Foo et al. (2005, p393); Beckman et al. (2007, p157)
Shannon index	$H = -\sum_{l=1}^{l} P_i(lnP_i)$	Categorical	Pelled et al. (1999, p16); Ucbasaran et al. (2003, p116); Beckman et al. (2007, p156)

Notes: The Shannon index is referred to as Teachman's index in Pelled et al. (1999, p16) and Ucbasaran et al. (2003, p116). The Herfindahl-Hirschman Index is referred to as Blau's heterogeneity index in Pelled et al. (1999, p11).

Table 2 summarizes the most common indicators of diversity used in the literature.We will argue that these measures of diversity have a number of drawbacks. First of all, the numerical value of such an index may have no intuitive interpretation. Second, we may be interested in asymmetric roles (due to power structures in a hierarchy) for individuals i and j, instead of assuming the two to be interchangeable. Third, the benefits of diversity may vary across the distribution of x (for example, being ten years younger may be more important if your partner is 30 than if your partner is 60). This will be difficult to quantify without making the

results difficult to interpret. Therefore, instead of trying to quantify diversity, we instead aim to present information on diversity in the most accessible way possible.

In our view, the standard scalar indicators of diversity suffer from problems related to extreme reductionist simplification. For example, a team of two men and one women is treated as having an identical gender composition as a team of two women and one man, or a team of four women and two men (because firm size is seldom interacted with the diversity indices). The maximum possible amount of diversity also depends on the group size (e.g. the maximum score for gender diversity in a team of three is not the same as the maximum score for a team of four). With regards to information on educational background, standard diversity measures provide information on the number of different backgrounds but they remain mute on *which* backgrounds are represented. To deal with these problems, we develop a less parametric approach to investigating diversity and performance.

3.3 Empirical strategy

We begin with some non-parametric illustrative statistics of the performance of pairs to give the reader an intuitive grasp of the diversity of teams and their performance outcomes. In recognition of the fact that the human brain is not well adapted to considering graphs containing three or more dimensions (which would be problematic if we had more than two team members), we plot the two founders on two axes and report the outcome in the resulting two-dimensional plane, using contour plots and cross-tabulations.

We then complement our 'raw' non-parametric results with parametric regressions, that have the advantage of allowing us to include control variables. In our parametric regressions, we prefer not to collapse information on diversity into a single summary diversity index, because this might not have a ready or 'intuitive' interpretation. Instead, we include a dummy variable for each category of combinations of partners. This gives us a different problem – that of having to include a large number of dummy variables for each pair-wise combination of characteristics. To deal with this latter issue, we adopt a 'stepwise' regression approach, whereby we repeat our regressions in iterative progression, at each step removing the least significant variable, and continuing this way until all of the remaining explanatory variables are above a minimum threshold level of significance.

3.4 Hypotheses development

The previous literature has generally formulated hypotheses in terms of how diversity in one particular dimension (e.g. age, gender, race, prior professional affiliations) affects the performance of the firm. In our paper, we do not focus on specific variables, partly because the previous literature has already developed some plausible hypotheses, and partly for reasons of space.

We focus instead on developing some 'meta-hypotheses' to loosely guide our empirical investigations, that will be used to evaluate the validity of our novel empirical approach. To begin with, we deliberately distinguish between the 'primary' and 'secondary' founding entrepreneur in our analysis, and hypothesize that the effect of diversity on performance is not invariant to which individual has which characteristics. For example, it may be that entrepreneurial pairs need one brash, energetic young individual to take the leading role, with an older and wiser individual acting as a 'guiding hand.' It may also be that the primary founder needs to have sound technical knowledge of the product, while benefitting from commercial advice from a supporting partner.

Hypothesis 1 *Structures of power and authority within teams are not symmetric, and the 'direction' of diversity moderates the effect of diversity on performance*

We also take a non-standard approach to measuring diversity, because we suspect that the standard practice of reducing diversity to a single summary scalar index leads to the loss of considerable information. Consider the variable age: we suspect that age has a non-linear effect on performance (from the liability of youth to the 'golden age' to senescence), and also that 10 years difference in age matters more when the two founders are on average 25 years old than when they are both on average 60 years old. Therefore we posit:

Hypothesis 2 *Diversity cannot easily be reduced to a single summary scalar index because of: a) nonlinearities, and b) the role of diversity depends on the 'average' level of the individuals*

Another feature of our paper is that we have two performance indicators: survival and growth. While these two indicators are both associated with firm performance, they shed light on different facets of performance. We prefer growth as an indicator of success, because some firms may survive and persist even if they experience poor performance (the so-called 'living dead').

Hypothesis 3 *Diversity has different effects for survival and growth*

4 Data

To investigate whether the direction of the employment diversity affects the performance of the new venture we make use of the information gathered from Danish government registers. This database, which is maintained by Statistics Denmark, is known under the name Danish Integrated Database for Labor Market Research (from now on referred to by its Danish acronym IDA). IDA is suitable for the analysis as its longitudinal characteristic allows us to follow individuals, establishments and firms over time. As a result, firm dynamics (birth, death and growth rate of firms) and the employment history of the active labor force can be identified. The database holds information on various demographic characteristics (e.g. gender, age, country of origin, type and level of education, which university the individuals attended, occupation and work experience). Because these individuals can be matched to a firm at any given year, it is possible to measure the level of diversity in the start-up.[3]

4.1 Start-ups and Entrepreneurial Pairs

To conduct the various analyses, we created a sample of all start-ups in the period 1999 to 2003 where we exclude all start-ups in the primary and public sector.[4] Table H.2 in the Appendix shows a more detailed industry breakdown of ventures in our sample. The motivation for selecting the time-period is two-fold. First, we want to be able to use the growth in sales as one of the firm growth measures; due to the break in the data between 1998 and 1999 it is problematic to include start-ups founded prior to 1999. Second, we want to follow the start-up for a number of years after founding to identify whether they survive and to establish their growth rates. The current dataset has data up to 2008, which allows us to follow each start-up for up to at least five years after founding.

To select our sample of start-ups it is important to identify the founding year. To do so, we use information on the firm's founding date from the company register in combination with the plant and firm identification number.We identify a start-up as a one-plant firm with no prior firm and plant identification number, which is in

[3]See Timmermans (2010) for a more detailed description of the database.

[4]Start-ups that are not within the 15 and 75 two-digit level NACE code are excluded. Within these two two-digit codes there is one classification, 40 to 45 (energy), that is a mix of both public and private firms, which also will be omitted.

line with Dahl and Reichstein (2007). Furthermore, to select genuinely new firms, we exclude all start-ups that are the result of a separation or merger of previously existing plants. Based on the above-mentioned selection criteria we identify 12,861 start-ups in the period 1999-2003.

To identify the employee diversity we need to identify the persons that are involved in the start-up in the year of founding. These persons are identified by merging two datasets: i.e. (i) the entrepreneurship database, which provides detailed information on who is the owner of the start-up; and (ii) the employee dataset that provides information on a person's primary and secondary workplace. We add all these individuals to identify the size of the start-up in the year of founding. We are aware that a firm changes their human resource composition but the motivation for choosing the human resources in the first year is: (i) the observation that most firms start small and hardly change in size during their lifetime (Aldrich and Ruef, 2006); (ii) the initial resource profile can be used to predict start-up performance, including failure (Cooper et al., 1994); and (iii) early hiring decisions have lasting consequences for new organizations (Baron et al., 1999). The size distribution of new ventures is presented in Table 3. Since we investigate prior joint work relations, we remove all the start-ups with only one person, i.e. 3,171 firms. Furthermore, as explained in Section 3 we will focus only on two-person start-ups, which is, as shown in Table 3, the most common start-up form. The final sample of start-ups will be 4,219, which is 32.80 percent of all start-ups and 43.54 percent of all multi-person start-ups. More descriptive statistics on the overall sample, including the performance indicators and demographic variables presented later in this paper, are presented in Table H.1 in the Appendix.

Table 3: Size distribution of start-ups in year 1

Number of owners and/or employees	Frequency	Percent of Total	Percent of "Teams"
1	3,171	24.66	.
2	4,219	32.80	43.54
3	1,856	14.43	19.15
4	953	7.41	9.83
5	606	4.71	6.25
6-10	1,232	9.58	12.71
11-25	824	6.41	8.50
26+	206	1.60	2.13
all	12,861	100.00	100.00

4.2 Survival and Growth

As mentioned above, we investigate the impact of this directional diversity on firm performance. The first performance measure is firm survival – due to the unique identification number associated to firms and plants we can follow the status of these organizational units in all years up to a change in this identification number.[5] A change in this identification number is always connected to a variable that indicates the status; when this status is identified as being a closure we consider it to be a non-survivor. In reality, firms might re-enter into the same or in different industries; however, for analytical purposes we will not consider this as a new entry and these firms do not re-appear in the sample. In addition to the closure of a business a firm might continue in another form, e.g. as a result of a merger or acquisition. We will treat these firms as survivors but these observations will be censored due to the structural change of these firms. This is the case for 149 firms in the sample. In total, 1,674 firms survive up to the fifth year.

We also investigate the impact of diversity on firm growth. To measure growth we use the growth in employees, both in total number of employees and in full time equivalent, and track the employment growth of the firm after 5 years. It is straightforward for us to measure the employment growth of our firms, because they all start with two individuals – we need only consider the number of employees in year 5.

4.3 Directional Diversity

As presented earlier in this paper, we will treat diversity as a directional measure. The first step to investigate the directionality of the diversity is to assign a *primus motor* for each two person start-up. To do so, we conduct several steps to find this individual, which is a combination of occupation code where the higher placed individual is considered to be higher ranked, whether the person has the start-up as the primary business, the number of days connected to the business, the highest income from the business and the self-employment history. Afterwards, we identify the diversity of this pair in terms of age, gender, and education (both on level, degree, and discipline), which are the most common diversity indicators.

[5] Timmermans (2010) discusses in more detail when firms and plants change identify. In short, IDA follows a person oriented approach towards change. Consequently, an establishments identification number remains the same from one year to the other whenever one of the following criteria is fulfilled: (i) a plant has the same owner and is active in the same industry; (ii) a plant has the same owner and the same labor force; or (iii) a plant has the same labor force and is located on the same address or is active in the same industry.

5 Analysis

5.1 Non-parametric analysis

In the following subsections we will present the various non-parametric analysis that provides us with a first indication on the effect of directional diversity on new venture performance. We will start with discussing the impact of age, followed by gender and education, will be discussed both on the level of education and the type of education, and ethnicity.

5.1.1 Age

We start with presenting some non-parametric analysis on the age of the founders. As presented in Table H.1 in the Appendix, the average age of the individuals in the sample is around 35 years of age.

Figure 1 shows the survival rates of entrepreneurial pairs, conditional on the age of individuals i and j. On the whole, there appears to be a rather uniform pattern – the survival of startups seems to be constant across the distribution of ages of i and j. That said, we observe a slightly longer survival of firms depending on the age of the entrepreneurial team, i.e. where both i and j are older.

Figure 2 shows the employment growth outcomes associated with different partnership combinations according to age. A first observation is that the best performing ventures, in terms of employment growth, are those where the primary founder has an age of around 20, while the secondary partner has an age of around 30. This suggests that both partners should be relatively young, to cope energetically with the workload of starting a new venture, although the secondary founder should be noticeably older than the first. Hence, some diversity in age can be valuable. Other regions associated with high employment creation are also visible, such as when age_i=45 and age_j=40. A second observation is that job creation generally seems to decrease with age of both the primary and the secondary partner, although the relationship is not smooth or linear.

5.1.2 Gender

In the second non-parametric analysis we focus on the gender composition and direction of the entrepreneurial pair. As expected, men are disproportionately represented as founders and women are most often identified as individual j compared to men.

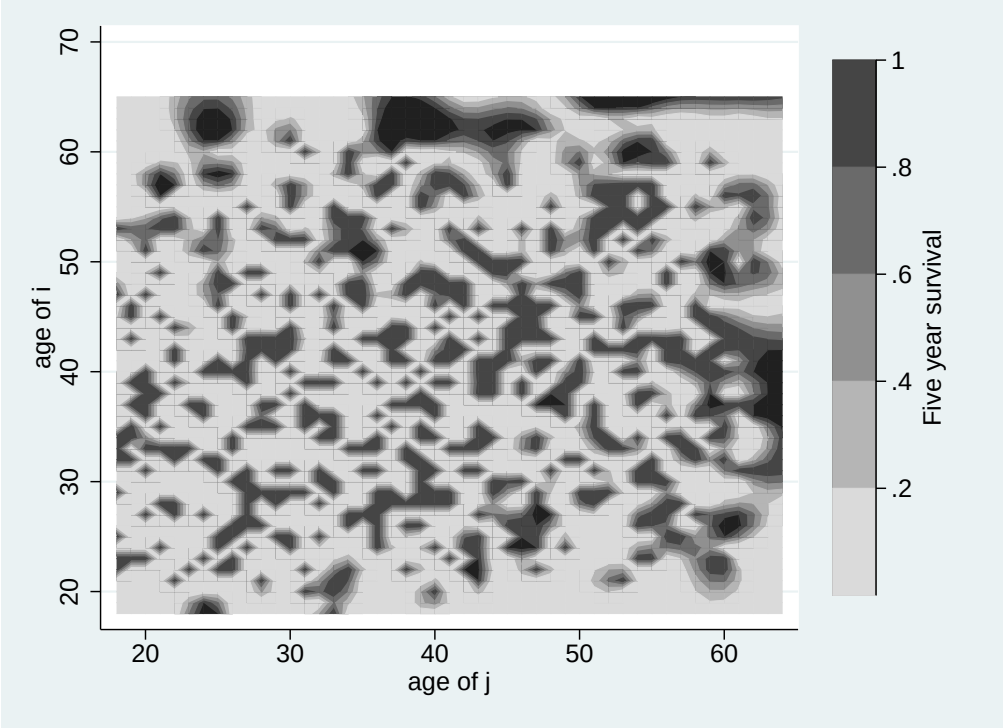

Figure 1: Contour plot of the survival of entrepreneurial pairs. z-axis: average survival after 5 years. Contour plot produced using thin-plate-spline interpolation.

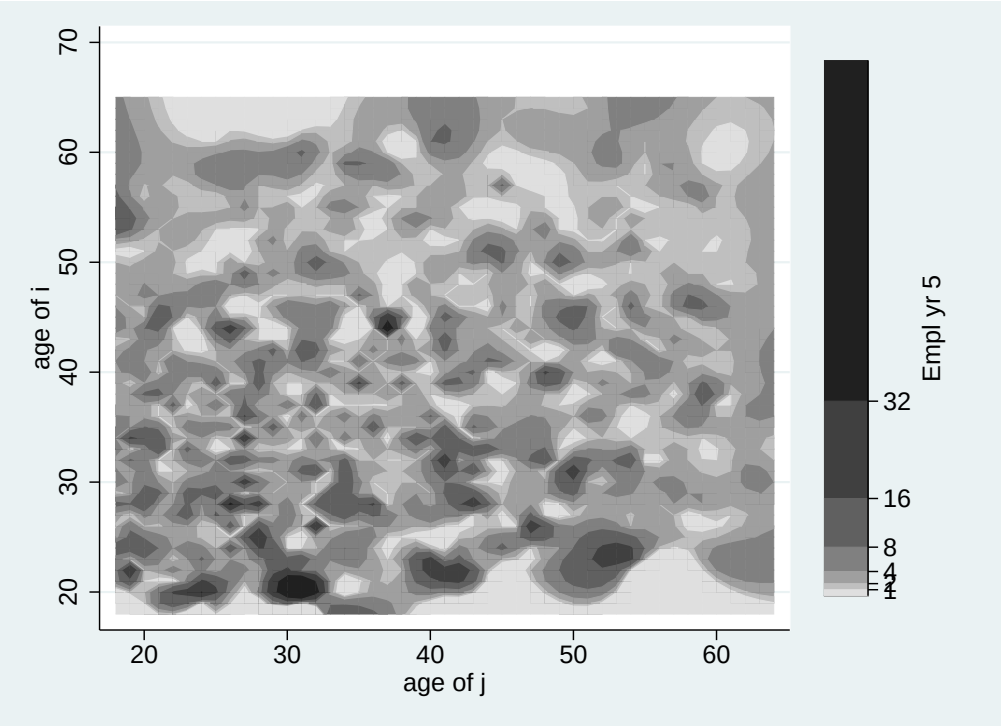

Figure 2: Contour plot of the outcomes associated with entrepreneurial pairs. z-axis: employment after 5 years, measured in terms of number of employees at the date of annual compulsory registration (in November of each year). Contour plot produced using thin-plate-spline interpolation.

		j	
		female	male
i	female	0.336	0.311
	male	0.362	0.386

Figure 3: Survival rates after 5 years, by gender, for individuals i (primary founder) and j. Cells with above-median values are highlighted.

		j	
		female	male
i	female	3.700	4.282
	male	3.921	4.782

Figure 4: Means of number of employees after 5 years, by gender, for individuals i (primary founder) and j. Cells with above-median values are highlighted.

Figure 3 looks at survival while Figure 4 looks at growth. These two tables provide slightly contrasting results, suggesting that the gender balance has different effects for survival and growth.

Figure 3 shows that the highest survival rates are for male-male pairs; the lowest is for pairs of females. For mixed-sex pairs, we observe higher survival chances if the principal entrepreneur is male.

Figure 4 shows that the highest employment growth is for pairs of males, and the lowest for pairs of females, similar to the pattern observed for survival in Figure 3. However, the second highest job creation rate is for mixed-sex pairs led by females, of the type i=female, j=male. This contrasts with the results for survival, where better mixed-sex outcomes were observed when the male was more dominant.

5.1.3 Education

In our analysis on the impact of directional in diversity on education we would like to take into account both the level of education and the type of education. The structure of the education variable in IDA allows us to identify the level of education (based on the first two-digits of an eight-digit code) and the discipline being taught (digit three and four of the eight-digit education code). We begin by considering education level before moving on to education type.

Education level In preparing the sample for the non-parametric analysis on education, we first remove those firms where the individuals are that have two-digit education-level-code 17, which stands for independent school or free school. The motivation for this step is the low number of individuals in the sample that have this education ($n = 2$). For this analysis we also drop individuals for who we do not know the education they obtained; however, these individuals are included in the regressions presented later on.

Table 4: Survival rates after 5 years, by education level, for individuals i (primary founder) and j. Empty cells correspond to cases where we have no observations for this particular combination of education levels. Cells with above-median values are highlighted.

		j								
		10	20	25	35	40	50	60	65	70
	10	0.222	0.247	0.296	0.342	0.326	0.274	0.500	0.269	
	20	0.293	0.429	0.800	0.356	0.688	0.423	0.250	0.250	0.000
	25	0.333	0.250	0.000	0.537	0.667	0.400	0.500	0.333	
	35	0.336	0.414	0.485	0.456	0.547	0.467	0.429	0.378	0.000
i	40	0.349	0.455	0.333	0.471	0.500	0.273	0.600	0.222	0.000
	50	0.304	0.400	0.500	0.267	0.600	0.558	0.500	0.214	1.000
	60	0.000	0.000		0.700	0.400	0.667	0.333	0.462	0.000
	65	0.083	0.200	1.000	0.400	0.222	0.545	0.400	0.610	0.750
	70								0.000	0.000

Taking the above-mentioned removals into account we have the following education level variables, the codes are indicated in brackets: Public/Primary School (10); Higher Preparatory Courses (15); General secondary education (Gymnasium and Higher Preparatory Examination(20); Higher Commercial Examination Programme and Higher Technical Examination Programme (25); Vocational basic (30); Vocational training and main course (35); Training of skilled workers / skilled workers (39); Vocationally oriented short-cycle higher education programmes (40); Professional Bachelor (50); Academic Bachelor (60); Master (65); PhD Programmes (70).

In Table 4 we present the survival rates of the start-ups based on the combination of education levels of the founders and the directionality of these education levels. We observe that those with the least education (i.e. public/primary school) generally seem to have the lowest survival rates. For the other education levels, however, there is no clear evidence to suggest that higher education enhances survival. We also observe that some diversity in education level seems to be beneficial, because the highest survival rates are not usually observed along the diagonal.

Table 5 shows the number of employees (november headcount) after 5 years by education level. No clear patterns emerge from this table, although it would appear that post-entry growth is low when $Educ_j = 10$ (that is, Public/Primary

Table 5: Number of employees after 5 years, by education level, for individuals i (primary founder) and j. Empty cells correspond to cases where we have no observations for this particular combination of education levels. Cells with above-median values are highlighted.

		j								
		10	20	25	35	40	50	60	65	70
	10	4.000	4.550	4.500	4.176	4.857	3.850	3.000	5.571	
	20	4.647	7.067	4.375	4.419	3.091	6.909	3.500	2.667	
	25	4.556	13.333		6.000	7.500	4.500	7.000	3.000	
	35	4.120	3.759	9.813	3.991	5.462	3.371	4.667	3.357	
i	40	3.267	4.000	8.000	6.417	6.647	7.000	2.000	9.500	
	50	3.714	7.667	2.333	3.667	4.556	3.000	2.000	2.000	9.000
	60				3.857	16.500	1.500	8.667	4.167	
	65	2.000	8.000	6.333	4.200	2.500	9.833	1.500	6.240	33.333
	70									

school education for the secondary founder) and relatively high when $Educ_j = 40$ (vocationally oriented short-cycle higher education programmes for the secondary founder).

Education type From the level of education we try to identify whether the type of education matters for survival and growth. The education types are divided in four categories. One type are all the programmes in vocational training and below ($\leq$) Voc Tr) and three where the educational programmes above this level have been divided in: degrees within science technology, engineering and mathematics (STEM), business related degrees (Business); and other degrees (Other). These different types of education where Figure 5 shows the mean survival rates for each category, while Figure 6 shows the mean number of employees (november headcount) after 5 years for a given combination of i and j.

Figure 5 contains a number of interesting results, among which some evidence that the direction of diversity matters – $EducType_i$=STEM, $EducType_j$ = Business has a high survival rate (0.471) while $EducType_i$ = Business, $EducType_j$ = STEM has a remarkably low survival rate (0.125). This pattern is also visible in Figure 6, which pertains to growth. The highest employment growth (9.375) is associated with $EducType_i$=STEM, $EducType_j$ = Business; while the lowest employment growth (1.000) is associated with $EducType_i$=Business, $EducType_j$ = STEM.

		j			
		$\leq$ Voc Tr	STEM	Business	Other
i	$\leq$ Voc Tr	0.348	0.394	0.466	0.265
	STEM	0.5	0.535	0.471	0.643
	Business	0.450	0.125	0.514	0.450
	Other	0.304	0.333	0.467	0.388

Figure 5: Survival rates after 5 years, by education type, for individuals i (primary founder) and j. Cells with above-median values are highlighted.

		j			
		$\leq$ Voc Tr	STEM	Business	Other
i	$\leq$ Voc Tr	4.195	4.095	5.379	4
	STEM	5.142	7.385	9.375	3.722
	Business	4.731	1	6.211	3.222
	Other	3.624	5.667	4	5.667

Figure 6: Means of number of employees after 5 years, by education type, for individuals i (primary founder) and j. Cells with above-median values are highlighted.

(More generally, we observe that the two lowest outcomes (1.000, 3.222) occur when the primary entrepreneur (i.e. individual i) has a business education.) This clearly shows that STEM and Business education backgrounds complement each other in complex ways.

5.1.4 Ethnicity

When looking at the issue of ethnicity, we focus only on the distinction Danes vs Non-Danes. The motivation for this distinction is that about 90% of the individuals in our sample are Danes. Our results for five-year survival (shown in Figure 7) show that the highest survival rates are for ventures founded by two Danes. With regards to mixed pairs, it matters little whether the non-Dane is first (i) or second (j).

The results for 5-year employment growth are in Figure 8. In contrast to the results for survival, we observe that pairs of Danes perform well, but that the best outcome is for ventures of mixed pairs where the primary entrepreneur i is a Dane. Pairs of non-Danes perform worst in terms of employment growth, which was also the case for survival.

5.2 Regressions

The non parametric analysis we presented above give an indication on the role of direction diversity on new venture performance. To study this effect in more detail we will now move into other estimation techiques to control for other factors that might explain new venture performance. To do so, we estimate the following regression equation:

		j	
		Dane	non-Dane
i	Dane	0.392	0.221
	non-Dane	0.225	0.151

Figure 7: Survival rate after 5 years, by ethnicity, for individuals i (primary founder) and j. Cells with above-median values are highlighted.

		j	
		Dane	non-Dane
i	Dane	4.604	4.840
	non-Dane	3.805	2.600

Figure 8: Means of number of employees after 5 years, by ethnicity, for individuals i (primary founder) and j. Cells with above-median values are highlighted.

$$\begin{aligned} Y_k &= \beta_0 + \beta_1 \cdot \sum_{i=0,1} \sum_{j=0,1} Sex_{ij} + \beta_2 \cdot \sum_{i=1...3} \sum_{j=1...3} AgeGroup_{ij} \\ &+ \beta_3 \cdot \sum_{i=0,11} \sum_{j=0,11} Education_{ij} + \beta_4 \cdot \sum_{i=0,1} \sum_{j=0,1} Danish_{ij} \\ &+ \beta_6 \cdot VentureCharacteristics_k + \varepsilon_k \end{aligned} \tag{1}$$

The unit of observation is the performance of venture k, and the explanatory variables include a constant term β_0, characteristics of the two founders i and j, as well as some venture-specific controls.

We recoded our education variables to take into account the interdependence of education level and education type. Those with the lowest educational qualifications have not had the opportunity to specialize, and therefore the types of education refer only to those above a minimum level of education. To take this into account, we recoded our education variables $Education_{ij}$ such that i and j can take the following values: 1 for all up to (and including) highschool; 2 for vocational training; 3_1, 3_2 and 3_3 for vocational oriented short-cycle education programes that specializes in either STEM, Business or other (respectively); 4_1, 4_2 and 4_3 for undergraduate (both academic and professional bachelor degrees), that specializes in either STEM, Business or other, respectively. and 5_1, 5_2 and 5_3 graduate and PhD education that specializes in either STEM, Business or other, respectively.

In addition to recoding the education variables we also recoded the age categories where we divided age in three separate dummies, i.e.: less or equal to 30; age between 31-45 years, and 45 years and up.

Our matrix of variables in $VentureCharacteristics_k$ includes a set of control vari-

ables. The first control variables we include are industry controls. In some industries, such as manufacturing, we have only a few firms present. To deal with this, we regroup some sectors together, following the Eurostat industry classification scheme for manufacturing sectors.[6] We also have few firms in two-digit NACE sectors 65 and 67 (banking, insurance, etc) and so we merge these sectors together with 66 (life insurance, pensions, etc) to generate a new industry group which corresponds to the Eurostat definition of "Knowledge-intensive financial services".[7] Second, the entrepreneurial pairs might be based on family relationships. As this relationship can influence the performance of the firms in different ways we included four dummies making a distinction whether the entrepreneurial pairs are spouses, siblings, father and son/daughter, or mother and son/daughter. Family firms account for 20% of our sample. Third, similar industry experience is an important factor that explains new venture performance, in particular survival (Dahl and Reichstein, 2007). To control for this factor we created two variable that iwhether the two entrepreneurs have common industry experience in the previous 5 years;[8] Fourth, we introduce cohort dummies, which correspond to the year (1999-2003) in which the firm was founded. Finally, to control for the regional dimension, we created a set of five region dummies that correspond to the five Danish administrative regions.[9]

Altogether, we will have a large set of variables in our initial regressions. To maintain overview of the analysis we conduct a stepwise regression technique. In this technique we take Equation (1) as a starting point before stepwise removal of the least significant explanatory variable. We stop when the least significant explanatory variable is significant at the 20% level.

5.2.1 Survival

Table 6 contains the regression results for survival after 5 years. Teams that include old old founders tend to have a high survival rate. The two best performing age combinations are when the secondary founder is in the older age category. Mixed-gender teams headed by a female have the lowest survival prospects.

[6]The Eurostat manufacturing industry classification scheme has the following four categories: high tech, medium-high tech, medium tech and low tech. More details can be found at http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Glossary:High-tech_classification_of_manufacturing_industries.

[7]See http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/Annexes/hrst_st_esms_an9.pdf.

[8]That is, whether one or both have worked in the same 4-digit NACE industry class.

[9]These regions are: Capital Region of Denmark, Region Zealand, North Denmark Region, Region of Southern Denmark, Central Denmark Region.

Education seems to help survival, because all of the significant education dummies are positive (with respect to the base case dummy which is two founders with minimal education). In terms of ethnicity we observe that teams of two danes have the best expected survival. Furthermore, we also observe that prior industry experience helps, especially if both founders can build on this experience. Finally, if the teams consist of individuals that are in a registered partnership (spouse) the new venture has higher survival prospects.

5.2.2 Employment growth

Table 7 contains the regression results for number of employees (November headcount) after 5 years. For age, we observe results that contrast to our findings for survival – the base case (youngest age category for both founders) has the highest expected employment growth, judging by the fact that all of our significant dummies are negative. In fact, teams of two old founders have the strongest negative coefficient, which indicates lowest employment growth. Firms with older aged pairs firms demonstrate stronger entrepreneurial persistence, but firms composed of young founders demonstrate, conditional on survival, higher growth. This mirrors findings that younger firms grow faster (i.e. years since start-up, at the business level rather than at the level of individuals).

Interestingly, conditional on survival, mixed gender teams headed by a female have the highest employment growth (even though we saw in Table 6 that they have the lowest survival chances). Education dummies show a mixed bag of results. With respect to the base case (minimal education for both founders) some dummies are positive while others are negative. The largest positive coefficient is for the combination $i = 3\_2, j = 4\_2$, although the reverse ordering $i = 4\_2, j = 3\_2$ has a negative coefficient on average. In fact, of the three dummies representing teams led by $i = 4\_2$, these all have negative coefficients. Thus, we can only conclude that the ordering has a large impact on the effects of diversity on performance. Regarding ethnicity, the highest-growth ventures are those led by a Dane (but it matters little whether the second founder is a Dane or a foreigner). Family firms consistently have negative coefficients – whether we consider pairs with siblings, with a father, with a mother, or with a spouse. Finally, it is curious to see that previous industry experience displays a negative coefficient on employment growth.

5.2.3 Robustness analysis

To deal with the risk that our stepwise procedure will remove potentially valuable variables, we compared our stepwise estimates with those of the full model

Table 6: Stepwise logit regression of equation (1), where the binary dependent variable is survival after 5 years.

	Coefficient	z-stat
Age group dummies		
agegrp_dummy_12	0.241	1.74
agegrp_dummy_21	0.359	3.47
agegrp_dummy_22	0.356	3.61
agegrp_dummy_23	0.450	3.25
agegrp_dummy_32	0.286	1.98
agegrp_dummy_33	0.378	2.65
Gender dummies		
man_dummy_01	-0.238	-2.15
Education		
education_dummy_2__2	0.334	3.85
education_dummy_2__4_3	-0.662	-1.75
education_dummy_3_1__2	0.788	3.12
education_dummy_3_1__4_2	1.400	1.44
education_dummy_3_1__4_3	1.515	1.34
education_dummy_3_2__1	0.762	1.72
education_dummy_3_2__2	0.752	2.14
education_dummy_3_2__3_2	1.222	1.91
education_dummy_3_3__2	0.821	1.41
education_dummy_4_1__1	0.569	1.53
education_dummy_4_1__2	0.743	2.03
education_dummy_4_1__4_1	1.322	2.13
education_dummy_5_1__5_1	1.411	2.98
education_dummy_5_3__1	-0.996	-1.63
education_dummy_5_3__5_3	1.097	1.42
Nationality dummies		
dane_dummy_11	0.600	5.55
(Shared) industry experience		
share_indexp_2	0.332	4.18
share_indexp_3	0.709	6.99
Family firms		
spouse	0.370	2.91
Constant	-1.354	-8.88
Industry dummies	yes	
Region dummies	yes	
Year dummies	yes	
Observations	4.176	
Pseudo-R^2	0.080	

Table 7: Stepwise OLS regression of equation (1), where the dependent variable is number of employees after 5 years. Robust standard errors obtained from the Huber/White/Sandwich estimator.

	Coefficient	t-stat
Age group dummies		
agegrp_dummy_21	-1.858	-4.46
agegrp_dummy_22	-1.851	-4.17
agegrp_dummy_23	-2.713	-5.96
agegrp_dummy_31	-2.424	-4.40
agegrp_dummy_32	-2.519	-5.36
agegrp_dummy_33	-3.384	-7.68
Gender dummies		
man_dummy_01	1.195	2.11
man_dummy_10	0.999	1.54
man_dummy_11	0.763	1.86
Education dummies		
education_dummy_1_3_2	-1.771	-1.87
education_dummy_1_5_1	-1.692	-3.68
education_dummy_2_3_1	1.436	1.36
education_dummy_2_4_1	-0.760	-1.47
education_dummy_2_4_2	-2.831	-7.76
education_dummy_3_1_2	1.511	2.15
education_dummy_3_1_3_1	4.626	3.35
education_dummy_3_1_3_2	6.984	14.01
education_dummy_3_1_4_1	-2.000	-3.06
education_dummy_3_1_4_2	-2.163	-3.79
education_dummy_3_1_5_3	-3.346	-4.65
education_dummy_3_2_3_1	-3.729	-5.99
education_dummy_3_2_4_2	27.835	65.32
education_dummy_3_2_5_2	-1.003	-2.03
education_dummy_3_3_3_2	-4.046	-5.38
education_dummy_3_3_3_3	-2.803	-3.86
education_dummy_3_3_4_1	3.867	7.79
education_dummy_3_3_4_3	-3.649	-5.21
education_dummy_4_1_1	3.756	1.77
education_dummy_4_1_3_1	-1.527	-1.46
education_dummy_4_1_3_2	6.826	11.90
education_dummy_4_1_5_1	19.293	10.56
education_dummy_4_2_1	-2.106	-3.93
education_dummy_4_2_3_2	-3.126	-5.29
education_dummy_4_2_4_2	-2.032	-3.06
education_dummy_4_2_5_2	-5.606	-7.28
education_dummy_4_2_5_3	-3.861	-7.76
education_dummy_4_3_3_2	10.402	13.53
education_dummy_4_3_4_2	-3.075	-2.33
education_dummy_4_3_5_1	1.435	2.37
education_dummy_4_3_5_2	-0.971	-2.26
education_dummy_4_3_5_3	-3.513	-5.95
education_dummy_5_1_2	-1.827	-3.03
education_dummy_5_1_4_1	-4.403	-3.50
education_dummy_5_1_4_2	-5.174	-9.02
education_dummy_5_1_5_1	2.334	1.53
education_dummy_5_1_5_2	16.612	1.83
education_dummy_5_1_5_3	-2.299	-2.03
education_dummy_5_2_4_3	-0.878	-1.84
education_dummy_5_2_5_3	-2.693	-2.25
education_dummy_5_3_3_3	7.599	13.07
education_dummy_5_3_5_2	-2.181	-3.48
Nationality dummies		
dane_dummy_01	1.360	2.05
dane_dummy_10	2.056	3.77
dane_dummy_11	2.223	6.09
Family firms		
siblings	-1.252	-2.34
dad	-1.380	-2.84
mom	-1.902	-5.22
spouse	-1.639	-3.32
(Shared) industry experience		
share_indexp_2	-0.608	-2.65
Constant	5.046	6.51
Industry dummies	yes	
Region dummies	yes	
Year dummies	yes	
Number of obs	1652	
R^2	0.165	

(all variables included) to see if there were any major discrepancies; however on discrepancies were found.

We also changed the time period over which performance was measured – instead of focusing on the five years after entry, we also investigated three-year and four-year periods for our two performance indicators (survival and employment growth). These results are presented in Tables H.4 - H.9 in the Appendix.

We also took an alternative employment growth indicator. In our baseline analysis, we measured employment taking the number of employees (November headcount). To check the robustness of our measure, however, we also considered employment (growth) measured using full time equivalent These results are presented in Table H.3 in the Appendix. All in all, the results are very similar.

To check that our identification of 'primary' and 'secondary' entrepreneur is valid, we repeated the analysis on a subsample of businesses where we could be more confident that our attribution of founders as 'primary' and 'secondary' was meaningful. This analysis is presented in the Appendix in Tables H.10, H.11, and H.12. The results from this analysis are not very significant (because of a lower number of observations) but generally are in accord with our main findings.

6 Discussion

In this section we will seek to 'digest' our findings by referring to our three hypotheses.

Hypothesis 1 stated that the effects of diversity on the outcomes of new businesses were heavily moderated by the 'position' or 'status' within the hierarchy. We find considerable support for this hypothesis because our results were far from 'symmetric' in a number of cases. This suggests that beneficial characteristics of the primary founder are not necessarily those that would best befit the secondary founder. With regards to age, growth tends to be higher if the primary founder is younger than the second. With regards to type of education, we obtained a mixed set of results, although businesses with a commercially-minded individual playing a secondary role performed better in terms of survival and growth than when a commercially-minded individual was the primary founder. More generally, our results for education type were far from symmetric. With some of our other variables, however, symmetry in characteristics space was associated with better outcomes (such as two Danes as founders; or two men with respect to firm survival; or two founders with low education having the worst survival chances).

Hypothesis 2 stated that the effects of diversity were non-linear and complex and could not easily be represented using a linear unidimensional indicator. We observed that the 'optimal' position in characteristics space was not monotonically increasing – for example, low education was associated with low survival, but there was little dependence of survival on education above a certain threshold. We also observed that the 'optimal' position in characteristics space depended on the characteristics of the partner – a powerful illustration of this idea is that, controlling for other factors, a configuration of education types $i = 3\_2, j = 4\_2$ had the highest expected employment growth, while the inverse configuration ($i = 4\_2, j = 3\_2$) yielded a negative coefficient (with respect to the base case of minimal education), and furthermore, changing the characteristics of the 'second fiddle' turned the coefficient from strongly positive to negative (in the case of $i = 3\_2, j = 3\_1$). Finally, another problem is that diversity will probably interact with firm size (this was not examined here because all businesses in our sample have the same start-up size: 2 individuals).

Taken together, our support for hypotheses 1 and 2 provide justification for our new methodology, which has identified effects of diversity on performance that could not have been uncovered using the standard econometric approach.

Hypothesis 3 predicted that diversity has different effects on survival and employment growth (even though these two could be considered as indicators of firm performance). In the case of family firms, we observe that they generally have an average performance for survival (because we observe no significant coefficients apart from a positive coefficient for spouses), although family firms are associated with slower employment growth. Regarding our employment growth regressions, we observe the largest negative effects for firms founded with mothers, then spouses, then siblings, then fathers. This hints that family firms are under pressure to keep the family 'tradition' alive (perhaps even in the face of prolonged poor performance), although they do not seek employees either through a mistrust of 'outsiders' or an aversion to the perceived risks or growth. Similarly, firms composed of older founders have better survival rates, but lower employment growth. Pairs of young founders have the highest employment growth. This could be because pairs of older founders do not want to take risks or over-exert themselves, but would prefer to 'coast along' before retirement. Younger pairs seem to be more willing to 'experiment' in their businesses, having higher exit hazards but often experiencing faster employment growth.

7 Conclusion

This study on more than 4000 entrepreneurial pairs and the amount of detail that is provided by the Danish register data on these pairs, provide interesting insights in how diversity affects performance. In particular it places question marks on the way diversity is treated in the various studies that exist on the topic. First, we provided proof that the diversity is heavily affected by the hierarchy that exist in the firm. Second, diversity is clearly not a linear and unidimensional indicator. This calls for an overall re-evaluation of the existing measures that exist. Third, diversity has a different impact on different performance measures.

The findings reported in this paper will also be of interest to the following:

- Entrepreneurs who are interested in choosing a partner for their new business idea.
- Angel investors who are interested in the outcomes associated with entrepreneurial partnerships
- University entrepreneurship promotion schemes will be interested in our results concerning the diversity of education subjects. For example, our finding that pairs of educational profiles consisting of STEM and business do best, may have implications for encouraging collaboration between engineering departments and business schools.
- Policy makers interested in offering assistance to potentially high-impact new ventures will be interested in information that helps identify which ventures will grow faster.
- Policy makers seeking to have a more efficient entrepreneurship policy should observe that family firms do worse in our case, and should perhaps rethink the specific benefits these firms get. For example, it is not clear why, in the UK, family firms get implicit subsidies (tax relief from inheritance tax) even though they are observed to be noticeably unproductive (Bloom and Van Reenen, 2010).

Finally, we would like to provide some suggestions for further work. First, we consider that there is still plenty of opportunity for finding richer indicators of diversity. It seems slightly ironic to us that it is frequently acknowledged that diversity is a 'double-edged sword' and often yields mixed results, and yet researchers generally compress the numerous dimensions of diversity into a single

indicator and then calculate the 'average effect' through standard regressions. We would like to see more 'diversity' in research into the role of diversity in teams. For example, future work could try to decompose the two edges of the 'sword' to investigate which factors affect conflict more than creativity (that is, distinguishing between the 'gross' and the 'net' costs and benefits of diversity). Second, it would be interesting to see if the degree of diversity in an entrepreneurial team affects the likelihood that the founder will stay with the firm in later years. True, there exist literature on team member exit (Ucbasaran et al., 2003; Chandler et al., 2004) but the time span of these studies are limited.

References

Aldrich, H. E. and Langton, N. (1998). Human resource management practices and organizational life cycles. In Reynolds, P., Bygrave, W. D., Davidsson, P., Gartner, W. B., and Carter, N. M., editors, *Frontiers of Entrepreneurship Research*, pages 349–357. Babson Park, MA: Babson College, Center for Entrepreneurial Studies.

Aldrich, H. E. and Ruef, M. (2006). *Organizations Evolving*. SAGE publications ltd, second edition.

Amason, A. C., Shrader, R. C., and Tompson, G. H. (2006). Newness and novelty: Relating top management team composition to new venture performance. *Journal of Business Venturing*, 21:125–148.

Ancona, D. G. and Caldwell, D. F. (1992). Demography and design: Predictors of new product team performance. *Organization Science*, 3(3):321–341.

Bantel, K. and Jackson, S. (1989). Top management and innovations in banking: Does the composition of the top team make a difference? *Strategic Management Journal*, 10:107–124.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1):99–120.

Baron, J., Hannan, M., and Burton, M. (1999). Building the iron cage: Determinants of managerial intensity in the early years of organizations. *American Sociological Review*, 64(4):527–547.

Beckman, C. M. (2006). The influence of founding team company affilitations on firm behavior. *The Academy of Management Journal*, 49(4):741–758.

Beckman, C. M., Burton, M., and O'Reilly, C. (2007). Early teams: The impact of teams demography on vc financing and going public. *Journal of Business Venturing*, 22(147-173).

Bloom, N. and Van Reenen, J. (2010). Why do management practices differ across firms and countries? *Journal of Economic Perspectives*, 24(1):203–224.

Byrne, D. (1971). *The Attraction Paradigm*. Academic Press: New York.

Chandler, G. N., Honig, B., and Wiklund, J. (2004). Antecedents, moderators, and performance consequences of membership change in new venture teams. *Journal of Business Venturing*, 20:705–725.

Chowdhury, S. (2005). Demographic diversity for building an effective entrepreneurial team: is it important? *Journal of Business Venturing*, 20:727–746.

Cooney, T. M. (2005). Editorial: What is an entrepreneurial team. *International Small Business Journal*, 23(3):226–235.

Cooper, A. C., Gimeno-Gascon, F. J., and Woo, C. Y. (1994). Initial human and financial capital as predictors of new venture performance. *Journal of Business Venturing*, 9:371–395.

Cox, T. H. and Blake, S. (1991). Managing cultural diversity: Implications for organizational competitiveness. *The Executive*, 5(3):45–56.

Dahl, M. S. and Reichstein, T. (2007). Are you experienced? prior experience an the survival of new organizations. *Industry and Innovation*, 14(5):497–511.

Dahlin, K., Weingart, L., and Hinds, P. (2005). Team diversity and information use. *The Academy of Management Journal*, 48(6):1107–1123.

Davidsson, P. and Honig, B. (2003). The role of social and human capital among nascent entrepreneurs. *Journal of Business Venturing*, 18:301–331.

Delmar, F. and 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.

Eisenhardt, K. and Schoonhoven, C. (1990). Organizational growth: Linking founding team, strategy, environment, and growth among U.S. semiconductor ventures, 1978-1988. *Administrative Science Quarterly*, 35(3):504–529.

Ensley, M. D., Carland, J. W., and Carland, J. C. (1998). The effect of entrepreneurial team skill heterogeneity and functional diversity on new venture performance. *Journal of Business and Entrepreneurship*, 10(1):1–15.

Foo, M. D., Wong, P. K., and Ong, A. (2005). Do others think you have a viable business idea? team diversity and judges' evaluation of ideas in a business plan competition. *Journal of Business Venturing*, 20:385–402.

Forbes, D., Borchert, P., and Zellmer-Bruhn, M. (2006). Entrepreneurial team formation: An exploration of new member addition. *Entrepreneurship Theory and Practice*, pages 225–246.

Gartner, W. B., Shaver, K. G., Gatewood, E., and Katz, J. A. (1994). Finding the entrepreneur in entrepreneurship. *Entrepreneurship Theory and Practice*, Spring 1994:5–9.

Hambrick, D., Cho, T., and Chen, M. (1996). The influence of top management team heterogeneity on firms' competitive moves. *Administrative Science Quarterly*, 41:659–684.

Hambrick, D. and Mason, P. (1984). Upper echelon: The organization as a reflection of its top managers. *The Academy of Management Review*, 9:193–206.

Harper, D. (2008). Towards a theory of entrepreneurial teams. *Journal of Business Venturing*, 23(6):613–626.

Harrison, D. and Klein, K. (2007). What's the difference? diversity constructs as separation, variety, or disparity in organizations. *The Academy of Management Review*, 32(4):1199–1228.

Horwitz, S. (2005). The compositional impact of team diversity on performance: Theoretical considerations. *Human Resource Development Review*, 4(2):219–245.

Horwitz, S. and Horwitz, I. (2007). The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of Management*, 6:987–1015.

Joshi, A. and Jackson, S. (2003). Managing workforce diversity to enhance cooperation in organizations. In West, M., Tjosvold, D., and Smith, K., editors, *International Handbook of Organizational Teamwork and Cooperative Working*. John Willey & Sons, Ltd.

Kamm, J., Shuman, J. C., Seeger, J., and Nurick, A. J. (1990). Entrepreneurial teams in new venture creation: A research agenda. *Entrepreneurship Theory and Practice*, Summer 1990:12.

McPherson, M., Smith-Lovin, L., and Cook, J. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444.

Murray, A. (1989). Top management heterogeneity and firm performance. *Strategic Management Journal*, 10:125–141.

Pelled, L. H. (1996). Demographic diversity, conflict, and work group outcomes: An intervening process theory. *Organization Science*, 7(6):615–631.

Pelled, L. H., Eisenhardt, K., and Xin, K. (1999). Exploring the black box: An analysis of work group diversity, conflict, and performance. *Administrative Science Quarterly*, 44(1):1–28.

Pfeffer, J. (1983). Organizational demography. In Staw, B. and Cummings, L., editors, *Research in Organizational Behavior*, volume 5, pages 299–357. Greenwhich, CT: JAI press.

Ruef, M. (2010). Entrepreneurial groups. In Landström, H. and Lohrke, F., editors, *Historical Foundations of Entrepreneurship Research*. Edward Elgar Publishing, Aldershot, UK.

Ruef, M., Aldrich, H., and Carter, N. (2003). The structure of founding teams: Homophily, strong ties, and isolation among U.S. entrepreneurs. *American Sociological Review*, 68(2):195–222.

Schimmel, A. (1994). *The Mystery of Numbers*. Oxford University Press.

Steffens, P., Terjesen, S., and Davidsson, P. (2011). Birds of a feather get lost together: new venture team composition and performance. 10.1007/s11187-011-9358-z.

Stirling, A. (2001). A general framework for analyzing divirsity in science, technology and society. *SPRU Working Papers 156*.

Tajfel, H. (1981). *Human Groups and Social Categories*. Cambridge University Press. Cambridge, U.K.

Timmermans, B. (2010). The danish integrated database for labor market research: Towards demystification for the english speaking audience. *Druid Working Paper Series*, 10-16.

Turner, J. (1987). *Rediscovering the Social Group: A Self-Categorization Theory*. Basil Blackwell: Oxford.

Ucbasaran, D., Lockett, A., Wright, M., and Westhead, P. (2003). Entrepreneurial founder teams: Factors associated with member entry and exit. *Entrepreneurship Theory and Practice*, Winter 2003:1042–2587.

Watson, W., Ponthieu, L., and Critelli, J. (1995). Team interpersonal process effectiveness in venture partnerschips and its connection to perceived success. *Journal of Business Venturing*, 10:393–411.

Watson, W. E., Stewart, W. H., and BarNir, A. (2003). The effects of human capital, organizational demography, and interpersonal processes on venture partner perceptions of firm profit and growth. *Journal of Business Venturing*, 18:145–164.

Williams, K. and O'Reilly, C. (1998). Demograhpy and diversity in organizations: A review of 40 years of research. In Staw, B. and Cummings, L., editors, *Research in Organizational Behavior*, pages 77–140. JAI Press, Greenwich.

8 Appendix

Table H.1: Summary statistics.

Variable	Obs	Mean	Std. Dev.	Min	Max
primary1	4219	0.988	0.109	0	1
primary2	4219	0.797	0.402	0	1
plantyear	4219	2001.008	1.367	1999	2003
age1	4219	36.895	10.303	15	91
age2	4219	34.349	11.859	15	75
educ_level1	4219	2.018	1.310	0	7
educ_level2	4219	1.755	1.162	0	7
educ_type1	4219	0.379	0.872	0	3
educ_type2	4219	0.277	0.779	0	3
man1	4219	0.705	0.456	0	1
man2	4219	0.629	0.483	0	1
survive1_5	4219	0.363	0.481	0	1
antnov_5	1652	4.388	5.198	0	88
aarsvrk_5	1652	3.108	3.867	0	60
brother	4219	0.032	0.177	0	1
father	4219	0.049	0.216	0	1
mother	4219	0.029	0.168	0	1
spouse	4219	0.091	0.288	0	1
family firm	4219	0.201	0.401	0	1
share_indexp	4219	0.300	0.370	0	1
share_hieduc	4219	0.137	0.270	0	1
Copenhagen region	4219	0.451	0.498	0	1

8.1 Alternative employment indicator

Table H.3 provides an alternative employment growth indicator, i.e. FTE growth over a five-year period.

8.2 Outcomes over 3 or 4 years instead of 5

Tables H.4, H.5 and H.6 show the results when a three-year period is considered.

Tables H.7, H.8 and H.9 show the results when a four-year period is considered.

8.3 Robustness sample

Table H.10 shows the results for survival for the robustness sample Table H.11 shows the results for antnov_5 employment indicator, and Table H.12 shows the results for the aarsvrk5 employment indicator.

Table H.2: Frequency of ventures disaggregated by 2-digit NACE sectors.

NACE02	Frequency	Percent
15	22	0.52
17	4	0.09
18	9	0.21
19	1	0.02
20	6	0.14
22	26	0.62
24	3	0.07
25	4	0.09
26	9	0.21
27	1	0.02
28	59	1.40
29	20	0.47
30	3	0.07
31	12	0.28
32	3	0.07
33	7	0.17
35	6	0.14
36	11	0.26
37	1	0.02
45	611	14.48
50	124	2.94
51	302	7.16
52	811	19.22
55	862	20.43
60	270	6.40
61	7	0.17
63	32	0.76
64	28	0.66
65	8	0.19
66	1	0.02
67	8	0.19
70	59	1.40
71	25	0.59
72	267	6.33
73	11	0.26
74	586	13.89
	4219	100.00

Table H.3: Alternative employment (growth) indicator: FTE after 5 years.

FTE_5	Coef.	t
constant	4.272	6.83
agegrp_dummy_13	-0.797	-1.32
agegrp_dummy_21	-1.085	-3.55
agegrp_dummy_22	-0.990	-3.00
agegrp_dummy_23	-1.649	-4.65
agegrp_dummy_31	-1.523	-3.69
agegrp_dummy_32	-1.514	-4.40
agegrp_dummy_33	-2.117	-6.70
man_dummy_01	0.825	2.00
man_dummy_10	0.783	1.75
man_dummy_11	0.616	2.19
education_dummy_1__3_2	-1.340	-1.81
education_dummy_1__5_1	-1.130	-2.20
education_dummy_2__3_2	1.774	1.32
education_dummy_2__4_2	-1.723	-4.64
education_dummy_2__4_3	0.711	1.61
education_dummy_3_1__2	1.393	2.16
education_dummy_3_1__3_1	5.236	3.83
education_dummy_3_1__3_2	5.611	15.21
education_dummy_3_1__3_3	0.727	1.68
education_dummy_3_1__4_1	-2.019	-3.69
education_dummy_3_1__5_3	-4.149	-6.36
education_dummy_3_2__3_1	-2.202	-4.55
education_dummy_3_2__3_2	1.362	1.39
education_dummy_3_2__4_2	22.106	58.94
education_dummy_3_2__5_2	-0.969	-2.73
education_dummy_3_3__3_2	-1.790	-3.14
education_dummy_3_3__3_3	-4.157	-6.67
education_dummy_3_3__4_1	3.209	8.07
education_dummy_3_3__4_3	-2.632	-4.73
education_dummy_4_1__1	3.549	1.70
education_dummy_4_1__3_1	-1.241	-1.46
education_dummy_4_1__3_2	-2.232	-4.22
education_dummy_4_1__5_1	10.245	9.13
education_dummy_4_2__1	-1.322	-1.79
education_dummy_4_2__3_2	-2.525	-5.11
education_dummy_4_2__4_2	-1.140	-1.98
education_dummy_4_2__5_2	-4.440	-6.50
education_dummy_4_2__5_3	-2.779	-7.61
education_dummy_4_3__3_2	8.786	17.18
education_dummy_4_3__4_1	0.801	1.73
education_dummy_4_3__4_2	-2.333	-2.21
education_dummy_4_3__4_3	-0.798	-1.34
education_dummy_4_3__5_1	3.142	7.12
education_dummy_4_3__5_3	-2.162	-3.01
education_dummy_5_1__1	-1.371	-1.77
education_dummy_5_1__2	-1.265	-1.84
education_dummy_5_1__4_1	-2.752	-2.18
education_dummy_5_1__4_2	-4.331	-7.71
education_dummy_5_1__5_1	1.923	1.55
education_dummy_5_1__5_2	11.405	2.11
education_dummy_5_1__5_3	-1.935	-2.03
education_dummy_5_2__1	-1.386	-2.97
education_dummy_5_2__4_3	-1.806	-4.80
education_dummy_5_2__5_2	-1.634	-1.86
education_dummy_5_2__5_3	-2.905	-3.25
education_dummy_5_3__1	-1.126	-1.56
education_dummy_5_3__3_3	5.998	11.71
education_dummy_5_3__5_2	-3.426	-5.58
share_indexp_2	-0.281	-1.69
dane_dummy_10	0.687	1.68
dane_dummy_11	0.934	3.54
siblin	-1.144	-3.11
dad	-0.928	-2.11
mom	-1.084	-3.48
spouse	-1.438	-4.06
Industry dummies	yes	
Region dummies	yes	
Year dummies	yes	
Number of obs	1652	
R^2	0.205	

Table H.4: Robustness analysis: survival after 3 years.

survive1_3	Coef.	z
constant	-1.300	-6.74
agegrp_dummy_12	0.238	1.80
agegrp_dummy_21	0.410	4.17
agegrp_dummy_22	0.367	3.90
agegrp_dummy_23	0.389	2.91
agegrp_dummy_32	0.400	2.89
agegrp_dummy_33	0.473	3.42
man_dummy_01	-0.215	-1.98
man_dummy_10	-0.150	-1.64
education_dummy_1__2	0.209	1.85
education_dummy_1__3_3	0.610	1.50
education_dummy_2__1	0.226	2.44
education_dummy_2__2	0.433	4.64
education_dummy_3_1__2	0.729	2.80
education_dummy_3_1__3_1	0.917	1.43
education_dummy_3_2__1	0.740	1.63
education_dummy_3_2__2	0.682	1.83
education_dummy_3_2__3_2	1.222	1.98
education_dummy_3_3__2	1.177	2.02
education_dummy_4_1__4_1	0.838	1.39
education_dummy_4_3__2	0.440	1.56
education_dummy_4_3__4_3	0.925	1.99
education_dummy_5_1__3_1	-1.355	-1.36
education_dummy_5_1__5_1	1.217	2.50
education_dummy_5_3__5_3	1.735	1.64
dane_dummy_01	0.329	1.51
dane_dummy_10	0.284	1.40
dane_dummy_11	0.758	4.66
share_indexp_2	0.360	4.73
share_indexp_3	0.633	6.39
sibling	0.278	1.53
spouse	0.282	2.13
Industry dummies	yes	
Region dummies	yes	
Year dummies	yes	
Obs	4175.	
Pseudo-R2	0.068	

Table H.5: Employees (november headcount) after 3 years.

Employees (november headcount) _3	Coef.	t
constant	4.376	9.97
agegrp_dummy_21	-0.725	-2.80
agegrp_dummy_22	-0.965	-3.61
agegrp_dummy_23	-1.177	-3.77
agegrp_dummy_31	-1.279	-3.64
agegrp_dummy_32	-1.359	-4.15
agegrp_dummy_33	-1.850	-6.87
man_dummy_01	0.622	2.24
education_dummy_1__4_2	-1.533	-1.37
education_dummy_1__5_1	-0.880	-2.28
education_dummy_2__4_2	-1.326	-4.68
education_dummy_2__5_1	-1.700	-2.52
education_dummy_3_1__2	1.024	1.97
education_dummy_3_1__3_2	6.640	18.65
education_dummy_3_1__3_3	1.594	4.39
education_dummy_3_1__4_1	-2.556	-5.82
education_dummy_3_1__5_3	-2.050	-3.29
education_dummy_3_2__3_1	-2.729	-7.01
education_dummy_3_2__4_2	6.217	20.29
education_dummy_3_2__5_2	1.114	3.55
education_dummy_3_2__5_3	-1.506	-3.69
education_dummy_3_3__3_3	-3.781	-10.18
education_dummy_3_3__4_1	1.193	3.40
education_dummy_3_3__4_3	-2.487	-7.89
education_dummy_3_3__5_2	-3.398	-5.38
education_dummy_4_1__1	4.507	1.83
education_dummy_4_1__3_1	-0.938	-3.47
education_dummy_4_1__3_2	-2.032	-5.99
education_dummy_4_1__5_1	11.648	9.48
education_dummy_4_2__4_1	-3.193	-8.28
education_dummy_4_2__4_2	-3.563	-5.07
education_dummy_4_2__5_2	-3.287	-8.60
education_dummy_4_2__5_3	-3.611	-10.11
education_dummy_4_3__3_2	3.202	6.57
education_dummy_4_3__3_3	2.414	5.75
education_dummy_4_3__4_1	-0.978	-2.34
education_dummy_4_3__4_2	-2.203	-3.00
education_dummy_5_1__3_1	-1.476	-4.54
education_dummy_5_1__4_1	-3.206	-5.73
education_dummy_5_1__4_2	-8.745	-1.92
education_dummy_5_1__5_1	1.888	1.47
education_dummy_5_1__5_2	8.634	2.00
education_dummy_5_1__5_3	-1.571	-1.83
education_dummy_5_2__3_3	0.567	1.68
education_dummy_5_2__4_2	-0.865	-1.83
education_dummy_5_2__4_3	-1.206	-1.90
education_dummy_5_2__5_2	-0.765	-2.63
education_dummy_5_2__5_3	-2.030	-3.30
education_dummy_5_3__2	-0.839	-1.77
education_dummy_5_3__3_3	5.217	17.03
education_dummy_5_3__5_2	-2.792	-8.27
dane_dummy_10	1.011	2.68
dane_dummy_11	1.381	5.43
share_indexp_2	-0.250	-1.49
sibling	-1.103	-2.94
dad	-1.291	-4.26
mom	-1.353	-4.36
spouse	-1.078	-5.02
Industry dummies	yes	
Region dummies	yes	
Year dummies	yes	
Number of obs	2100	
R-squared	0.131	

Table H.6: FTE after 3 years.

FTE_3	Coef.	t
constant	4.243	6.60
agegrp_dummy_21	-0.275	-1.57
agegrp_dummy_22	-0.346	-1.87
agegrp_dummy_23	-0.627	-2.79
agegrp_dummy_31	-0.693	-2.70
agegrp_dummy_32	-0.585	-2.77
agegrp_dummy_33	-0.865	-4.58
man_dummy_01	0.533	2.38
man_dummy_10	0.302	1.42
man_dummy_11	0.426	2.22
education_dummy_1__2	0.304	1.35
education_dummy_1__5_2	0.975	1.57
education_dummy_2__4_2	-1.006	-1.99
education_dummy_3_1__2	0.784	1.73
education_dummy_3_1__3_1	1.615	2.30
education_dummy_3_1__3_2	1.913	8.64
education_dummy_3_1__4_1	-1.635	-3.84
education_dummy_3_1__5_3	-1.653	-3.25
education_dummy_3_2__3_1	-1.308	-4.81
education_dummy_3_2__4_2	3.016	12.81
education_dummy_3_2__5_2	1.614	6.40
education_dummy_3_2__5_3	-2.925	-9.15
education_dummy_3_3__3_3	-3.264	-9.00
education_dummy_3_3__4_1	2.546	9.31
education_dummy_3_3__4_3	-1.462	-5.30
education_dummy_3_3__5_2	-2.441	-3.28
education_dummy_4_1__1	3.651	1.51
education_dummy_4_1__5_1	7.067	9.89
education_dummy_4_2__4_1	-1.337	-4.32
education_dummy_4_2__4_2	-2.590	-6.07
education_dummy_4_2__5_2	-2.713	-6.46
education_dummy_4_2__5_3	-1.411	-5.40
education_dummy_4_3__3_2	2.743	3.47
education_dummy_4_3__3_3	1.357	4.01
education_dummy_4_3__4_1	0.705	2.29
education_dummy_4_3__4_2	-1.502	-2.70
education_dummy_4_3__5_2	0.336	1.28
education_dummy_5_1__3_1	-1.254	-2.31
education_dummy_5_1__4_1	-2.021	-3.31
education_dummy_5_1__4_2	-7.093	-1.81
education_dummy_5_1__5_1	1.128	1.44
education_dummy_5_1__5_2	5.781	2.46
education_dummy_5_2__1	-0.642	-2.14
education_dummy_5_2__4_2	-0.849	-1.68
education_dummy_5_2__4_3	-0.996	-3.97
education_dummy_5_2__5_2	-0.956	-1.37
education_dummy_5_2__5_3	-1.479	-2.88
education_dummy_5_3__1	-0.848	-1.83
education_dummy_5_3__2	-0.882	-1.88
education_dummy_5_3__3_3	3.691	13.49
education_dummy_5_3__5_2	-1.868	-5.95
dane_dummy_10	0.522	1.82
dane_dummy_11	0.758	4.12
share_indexp_3	0.203	1.48
sibling	-1.005	-4.19
dad	-0.879	-3.80
mom	-1.036	-5.02
spouse	-0.981	-5.57
Industry dummies	yes	
Region dummies	yes	
Year dummies	yes	
Number of obs	2100	
R-squared	0.184	

Table H.7: Survival after 4 years.

survive1_4	Coef.	z
constant	-14.814	-19.17
agegrp_dummy_12	0.216	1.59
agegrp_dummy_21	0.397	3.99
agegrp_dummy_22	0.390	4.07
agegrp_dummy_23	0.449	3.34
agegrp_dummy_32	0.368	2.64
agegrp_dummy_33	0.467	3.36
man_dummy_01	-0.409	-3.02
man_dummy_10	-0.177	-1.49
man_dummy_11	-0.162	-1.51
education_dummy_2__1	0.157	1.72
education_dummy_2__2	0.320	3.56
education_dummy_3_1__2	0.674	2.66
education_dummy_3_2__1	0.896	1.95
education_dummy_3_2__2	0.532	1.52
education_dummy_3_2__3_2	0.915	1.43
education_dummy_4_1__2	0.501	1.38
education_dummy_4_1__4_1	1.062	1.73
education_dummy_4_3__2	0.418	1.48
education_dummy_5_1__5_1	1.156	2.48
education_dummy_5_2__1	0.858	1.49
education_dummy_5_3__1	-1.246	-2.03
dane_dummy_11	0.472	4.54
share_indexp_2	0.366	4.74
share_indexp_3	0.666	6.66
spouse	0.367	2.73
Industry dummies	yes	
Region dummies	yes	
Year dummies	yes	
Obs	4196	
Pseudo-R2	0.072	

Table H.8: Employees (November headcount) after 4 years.

Employees (November headcount_4	Coef.	t
constant	4.474	10.20
agegrp_dummy_21	-1.278	-4.24
agegrp_dummy_22	-1.197	-3.66
agegrp_dummy_23	-1.520	-3.96
agegrp_dummy_31	-1.550	-3.14
agegrp_dummy_32	-1.799	-5.01
agegrp_dummy_33	-2.417	-7.54
man_dummy_01	0.541	1.46
education_dummy_1__2	0.580	1.30
education_dummy_1__3_2	-1.276	-1.46
education_dummy_1__5_1	-1.280	-2.37
education_dummy_1__5_3	0.523	1.36
education_dummy_2__4_2	-2.248	-6.53
education_dummy_3_1__2	1.928	2.78
education_dummy_3_1__3_1	3.511	3.28
education_dummy_3_1__3_2	4.925	13.20
education_dummy_3_1__3_3	0.533	1.41
education_dummy_3_1__4_1	-2.490	-2.64
education_dummy_3_1__4_2	-1.514	-2.80
education_dummy_3_1__5_3	-2.988	-5.61
education_dummy_3_2__3_1	-3.570	-8.98
education_dummy_3_2__4_2	10.122	28.44
education_dummy_3_2__5_2	-3.199	-10.27
education_dummy_3_2__5_3	-1.907	-5.03
education_dummy_3_3__2	1.542	1.61
education_dummy_3_3__3_2	-1.193	-2.32
education_dummy_3_3__3_3	-2.629	-7.51
education_dummy_3_3__4_1	2.093	5.52
education_dummy_3_3__4_3	-2.489	-7.40
education_dummy_3_3__5_2	-2.922	-3.10
education_dummy_4_1__1	3.350	1.68
education_dummy_4_1__3_1	-2.995	-3.48
education_dummy_4_1__3_2	1.969	6.00
education_dummy_4_1__5_1	14.616	2.10
education_dummy_4_2__3_2	-2.613	-10.50
education_dummy_4_2__4_2	-4.309	-11.10
education_dummy_4_2__5_2	-4.572	-8.85
education_dummy_4_2__5_3	-3.477	-9.08
education_dummy_4_3__3_3	1.580	3.12
education_dummy_4_3__4_1	-1.161	-2.39
education_dummy_4_3__4_2	-2.694	-3.45
education_dummy_4_3__5_1	3.516	8.09
education_dummy_5_1__2	-1.388	-2.77
education_dummy_5_1__3_1	-1.258	-2.26
education_dummy_5_1__4_1	-3.793	-5.28
education_dummy_5_1__4_2	-3.910	-12.94
education_dummy_5_1__5_1	2.783	1.81
education_dummy_5_1__5_2	10.789	1.77
education_dummy_5_1__5_3	-1.702	-2.05
education_dummy_5_2__3_3	0.679	1.82
education_dummy_5_2__4_3	-2.147	-5.11
education_dummy_5_2__5_3	-1.966	-2.15
education_dummy_5_3__2	-0.956	-1.56
education_dummy_5_3__3_3	6.122	17.20
education_dummy_5_3__5_1	-0.710	-1.84
education_dummy_5_3__5_2	-2.112	-5.70
dane_dummy_01	0.887	1.92
dane_dummy_10	1.312	2.82
dane_dummy_11	1.835	5.95
share_indexp_2	-0.402	-2.15
sibling	-0.712	-1.66
dad	-1.298	-3.20
mom	-1.646	-4.44
spouse	-1.337	-5.18
Industry dummies	yes	
Region dummies	yes	
Year dummies	yes	
Number of obs	1872	
R-squared	0.131	

Table H.9: FTE after 4 years.

FTE_4	Coef.	t
constant	3.404	6.91
agegrp_dummy_21	-0.732	-3.38
agegrp_dummy_22	-0.532	-2.20
agegrp_dummy_23	-1.047	-3.79
agegrp_dummy_31	-0.816	-2.43
agegrp_dummy_32	-1.074	-4.45
agegrp_dummy_33	-1.378	-6.02
man_dummy_01	0.714	2.15
man_dummy_10	0.545	1.95
man_dummy_11	0.586	2.57
education_dummy_1__2	0.487	1.55
education_dummy_1__3_2	-1.047	-1.63
education_dummy_1__5_1	-1.214	-2.23
education_dummy_2__4_2	-1.309	-3.79
education_dummy_3_1__2	1.549	2.60
education_dummy_3_1__3_1	3.321	4.95
education_dummy_3_1__3_2	2.359	8.57
education_dummy_3_1__4_1	-1.626	-2.32
education_dummy_3_1__5_3	-2.313	-4.73
education_dummy_3_2__3_1	-1.673	-4.40
education_dummy_3_2__3_2	1.606	2.10
education_dummy_3_2__4_2	7.531	25.07
education_dummy_3_2__4_3	-1.301	-1.42
education_dummy_3_2__5_2	-1.546	-5.15
education_dummy_3_2__5_3	-3.683	-9.22
education_dummy_3_3__3_2	-1.079	-2.46
education_dummy_3_3__3_3	-3.365	-7.24
education_dummy_3_3__4_1	4.049	12.11
education_dummy_3_3__4_3	-1.861	-5.57
education_dummy_3_3__5_2	-2.079	-4.52
education_dummy_4_1__1	3.319	1.59
education_dummy_4_1__3_1	-2.375	-4.31
education_dummy_4_1__3_2	1.466	3.86
education_dummy_4_1__5_1	8.201	2.39
education_dummy_4_2__4_2	-3.117	-6.78
education_dummy_4_2__5_2	-3.662	-6.79
education_dummy_4_2__5_3	-1.857	-6.02
education_dummy_4_3__4_1	0.713	1.98
education_dummy_4_3__4_2	-1.946	-2.21
education_dummy_4_3__5_1	3.199	9.97
education_dummy_4_3__5_2	0.712	2.64
education_dummy_4_3__5_3	-0.730	-1.43
education_dummy_5_1__2	-1.451	-3.10
education_dummy_5_1__3_1	-1.284	-2.74
education_dummy_5_1__4_1	-3.324	-4.89
education_dummy_5_1__4_2	-3.985	-8.08
education_dummy_5_1__5_1	2.118	1.78
education_dummy_5_1__5_2	7.578	2.38
education_dummy_5_1__5_3	-1.339	-1.50
education_dummy_5_2__1	-0.967	-2.47
education_dummy_5_2__4_3	-2.251	-3.21
education_dummy_5_2__5_2	-1.521	-1.75
education_dummy_5_2__5_3	-2.308	-3.15
education_dummy_5_3__2	-0.611	-1.33
education_dummy_5_3__3_3	4.539	12.17
education_dummy_5_3__5_2	-2.796	-5.86
dane_dummy_10	0.615	1.89
dane_dummy_11	0.917	4.88
share_indexp_2	-0.223	-1.62
sibling	-0.900	-2.76
dad	-1.249	-4.17
mom	-1.248	-4.83
spous	-1.237	-4.81
Industry dummies	yes	
Region dummies	yes	
Year dummies	yes	
Number of obs	1872	
R-squared	0.193	

Table H.10: Robustness sample – survival after 5 years.

survive1_5	Coef.	z
constant	-13.440	-9.67
agegrp_dummy_12	0.184	0.57
agegrp_dummy_13	-0.145	-0.34
agegrp_dummy_21	0.331	1.38
agegrp_dummy_22	0.635	2.69
agegrp_dummy_23	0.793	2.61
agegrp_dummy_31	0.344	0.88
agegrp_dummy_32	0.132	0.38
agegrp_dummy_33	0.482	1.50
man_dummy_01	-0.345	-1.14
man_dummy_10	-0.390	-1.62
man_dummy_11	-0.227	-0.98
education_dummy_1__2	0.275	1.02
education_dummy_1__3_1	0.149	0.12
education_dummy_1__3_3	1.685	2.25
education_dummy_1__4_3	0.054	0.07
education_dummy_2__1	0.282	1.26
education_dummy_2__2	0.656	2.96
education_dummy_2__3_3	1.239	1.42
education_dummy_2__4_1	0.044	0.04
education_dummy_2__4_3	-0.215	-0.22
education_dummy_2__5_3	1.578	1.53
education_dummy_3_1__1	-0.156	-0.16
education_dummy_3_1__2	1.655	2.57
education_dummy_3_2__1	1.417	1.45
education_dummy_3_2__2	0.679	0.84
education_dummy_3_2__3_2	0.503	0.52
education_dummy_3_3__2	0.817	0.59
education_dummy_4_1__1	0.903	0.78
education_dummy_4_1__2	0.118	0.13
education_dummy_4_2__1	0.128	0.10
education_dummy_4_2__2	0.854	0.78
education_dummy_4_3__1	0.334	0.69
education_dummy_4_3__2	0.754	1.27
education_dummy_4_3__4_3	0.990	1.17
education_dummy_5_3__1	0.006	0.01
education_dummy_5_3__2	0.842	1.23
dane_dummy_01	0.211	0.38
dane_dummy_10	0.852	1.70
dane_dummy_11	1.007	2.35
share_indexp_2	0.471	2.83
share_indexp_3	0.804	3.81
sibling	-0.005	-0.01
dad	-0.015	-0.04
mom	0.085	0.21
spouse	-0.332	-0.97
Industry dummies	yes	
Region dummies	yes	
Year dummies	yes	
Obs	1111	
Pseudo-R2	0.113	

Table H.11: Robustness sample – antnov_5 employment indicator.

antnov_5	Coef.	t
constant	3.973	1.32
agegrp_dummy_12	-0.045	-0.04
agegrp_dummy_13	-0.725	-0.78
agegrp_dummy_21	-0.223	-0.32
agegrp_dummy_22	-0.383	-0.46
agegrp_dummy_23	-0.543	-0.58
agegrp_dummy_31	-0.305	-0.34
agegrp_dummy_32	-1.073	-1.29
agegrp_dummy_33	-1.547	-1.83
man_dummy_01	-0.034	-0.04
man_dummy_10	1.118	1.35
man_dummy_11	0.537	0.89
education_dummy_1__2	-0.949	-1.37
education_dummy_1__3_1	2.540	1.70
education_dummy_1__3_3	-2.695	-2.27
education_dummy_1__4_3	-2.170	-1.90
education_dummy_2__1	-0.285	-0.36
education_dummy_2__2	-0.548	-0.85
education_dummy_2__3_3	1.195	0.50
education_dummy_2__4_1	-1.372	-1.33
education_dummy_2__4_3	0.951	0.67
education_dummy_2__5_3	-3.137	-2.21
education_dummy_3_1__1	-0.073	-0.05
education_dummy_3_1__2	1.804	1.03
education_dummy_3_2__1	-1.498	-0.61
education_dummy_3_2__2	0.311	0.16
education_dummy_3_2__3_2	-0.309	-0.20
education_dummy_3_3__2	-0.944	-0.28
education_dummy_4_1__1	7.030	5.48
education_dummy_4_1__2	-0.691	-0.56
education_dummy_4_2__1	-0.587	-0.55
education_dummy_4_2__2	-2.304	-2.17
education_dummy_4_3__1	-0.402	-0.40
education_dummy_4_3__2	-1.488	-1.42
education_dummy_4_3__4_3	-0.823	-0.47
education_dummy_5_1__5_3	-1.498	-0.97
education_dummy_5_3__1	-1.530	-1.01
education_dummy_5_3__2	-1.834	-1.36
dane_dummy_01	0.570	0.46
dane_dummy_10	1.593	1.29
dane_dummy_11	1.295	1.26
share_indexp_2	-0.809	-1.86
share_indexp_3	-0.769	-1.37
sibling	-0.006	0.00
dad	-1.605	-2.10
mom	-2.052	-3.34
spouse	-1.333	-1.66
Industry dummies	yes	
Region dummies	yes	
Year dummies	yes	
Number of obs	370	
R-squared	0.180	

Table H.12: Robustness sample – FTE_5 employment indicator.

FTE_5	Coef.	t
constant	2.490	1.24
agegrp_dummy_12	-0.179	-0.25
agegrp_dummy_13	-0.316	-0.45
agegrp_dummy_21	-0.339	-0.69
agegrp_dummy_22	-0.213	-0.37
agegrp_dummy_23	-0.024	-0.03
agegrp_dummy_31	-0.233	-0.38
agegrp_dummy_32	-0.619	-0.96
agegrp_dummy_33	-0.549	-0.92
man_dummy_01	-0.365	-0.96
man_dummy_10	0.628	1.33
man_dummy_11	0.747	1.86
education_dummy_1__2	-0.576	-1.26
education_dummy_1__3_1	1.760	1.73
education_dummy_1__3_3	-0.598	-0.93
education_dummy_1__4_3	-1.289	-1.78
education_dummy_2__1	-0.130	-0.25
education_dummy_2__2	-0.166	-0.41
education_dummy_2__3_3	1.591	0.90
education_dummy_2__4_1	-1.725	-2.59
education_dummy_2__4_3	0.948	0.88
education_dummy_2__5_3	-2.095	-2.64
education_dummy_3_1__1	-0.440	-0.42
education_dummy_3_1__2	1.741	0.99
education_dummy_3_2__1	-0.547	-0.56
education_dummy_3_2__2	-0.214	-0.30
education_dummy_3_2__3_2	1.587	1.26
education_dummy_3_3__2	-0.898	-0.66
education_dummy_4_1__1	6.724	7.46
education_dummy_4_1__2	0.530	0.32
education_dummy_4_2__1	-0.172	-0.26
education_dummy_4_2__2	-1.620	-2.24
education_dummy_4_3__1	-1.009	-1.38
education_dummy_4_3__2	-0.454	-0.69
education_dummy_4_3__4_3	-0.520	-0.59
education_dummy_5_1__5_3	0.056	0.06
education_dummy_5_3__1	-2.335	-3.00
education_dummy_5_3__2	-0.615	-0.70
dane_dummy_01	-0.100	-0.13
dane_dummy_10	0.408	0.66
dane_dummy_11	0.721	1.27
share_indexp_2	-0.138	-0.49
share_indexp_3	-0.205	-0.53
sibling	-0.182	-0.21
dad	-1.385	-2.66
mom	-0.552	-1.37
spouse	-0.339	-0.73
Industry dummies	yes	
Region dummies	yes	
Year dummies	yes	
Number of obs	370	
R-squared	0.260	