The Origins of Entrepreneurship: Evidence from Sibling Correlations

Matthew J. Lindquist* Joeri Sol† C. Mirjam van Praag‡
Theodor Vladasel§

(**Preliminary and incomplete, please do not cite**)

February 25, 2016

Abstract

Vast amounts of money are currently being spent on policies aimed at promoting entrepreneurship. The success of such policies, however, rests in part on the assumption that individuals are not 'born entrepreneurs'. In this paper, we assess the importance of family background and neighborhood effects as determinants of entrepreneurship. We start by estimating sibling correlations in entrepreneurship. We find that between 20 and 50 percent of the variance in different entrepreneurial outcomes is explained by factors that siblings share. The average is 28 percent, hinting at relatively equal chances for becoming an entrepreneur. Allowing for differential treatment within families by gender and birth order does little to further increase our estimates of the importance of family-wide factors. We then go on to show that neighborhood effects, sibling peer effects, and parental income and education explain very little of these correlations. Parental entrepreneurship – occupational immobility – does play a large role, as do shared genes.

JEL Classification: D13, J62, L26.

Keywords: Entrepreneurship; Family Background; Intergenerational Persistence; Neighborhood Effects; Sibling Correlations.

*Stockholm University, matthew.lindquist@sofi.su.se
†University of Amsterdam, J.Sol@uva.nl
‡Copenhagen Business School and University of Amsterdam, mvp.ino@cbs.dk
§Corresponding author. Copenhagen Business School, tvl.ino@cbs.dk

1 Introduction

Entrepreneurship has been hailed as a driver of growth and as an avenue for upward social mobility (Holtz-Eakin, 2000; Holtz-Eakin et al., 2000; van Praag and Versloot, 2007; Shane, 2009). Policies aimed at encouraging successful entrepreneurship have therefore been adopted in many countries. For example, vast amounts of money are spent in the attempt to facilitate access to finance for entrepreneurs with great ideas, but limited personal capital (Lerner, 2009; Lelarge et al., 2010). Entrepreneurial education has also permeated academic curricula, from primary school, through tertiary education, and to active labor market programs, with the goal of supporting the transition of high ability individuals into entrepreneurship (Oosterbeek et al., 2010; Fairlie et al., 2015). The success of such policies, however, rests in part on the assumption that individuals are not 'born entrepreneurs'. If entrepreneurship is largely determined at a young age by factors outside of an individual's control, such as parental entrepreneurship experience, then adult-stage policies may miss their target.

In this paper, we assess the importance of family background and neighborhood effects as determinants of entrepreneurship. Do individuals start life with equal chances of engaging in entrepreneurship? If not, then what family- and/or community-wide factors do children and young adults face that may later limit their opportunities of becoming entrepreneurs?

Previous literature suggests that entrepreneurial preferences may be formed at a young age. In fact, entrepreneurship education has, so far, only been shown to influence the 'entrepreneurial intentions' of students in their early teens (Huber et al., 2014). Moreover, strong intergenerational associations in entrepreneurship have attracted considerable attention. Part of this relationship has been shown to be genetic, but parental role-modeling also appears to have a strong impact on subsequent labor market decisions (Sørensen, 2007; Nicolaou et al., 2008; Lindquist et al., 2015). More generally, family influences have been shown to play a critical role in the formation of human capital and in determining occupational choice (Becker, 1988; Björklund and Jäntti, 2012).

In contrast to the parent-child intergenerational transmission literature, we quantify the importance of family- and community-wide factors (experienced when young) as determinants of entrepreneurship by estimating sibling correlations in various measures of entrepreneurship and entrepreneurial success. Sibling correlations provide us with much broader measures of the importance of family background and neighborhood effects than do, for example, simple, single-trait intergenerational associations (Solon, 1999; Björklund and Jäntti, 2012). Sibling correlations can also be viewed as an inverse measure of equality of opportunity. In our case, if sibling correlations are low, then individuals are relatively more free to choose their occupations (all else equal), which implies a large degree of equal opportunity.

To compute sibling correlations in entrepreneurship, we use detailed data drawn from Sweden's Multigenerational Register. Our data set includes information on more than 700,000 siblings. We have extensive data on individual and family-wide socio-economic variables, including information on parental education and income, parental entrepreneurship, family structure and parish of residence when individuals' were young. As outcomes, we create a wide set of variables at the extensive and intensive margins, which allow us to obtain a complete picture of entrepreneurial processes: it is not enough that anyone can *become* an entrepreneur, it is also important for individuals to *survive and thrive* as entrepreneurs.

Our outcome variables are defined as follows. First, we create a dichotomous variable, *Entrepreneur*, that is equal to one if a person has ever received the majority of their annual labor income from a non-incorporated business that they owned in part or in full, and zero otherwise. We also create a dichotomous variable, *Incorporated*, that is equal to one if a person has ever received the majority of their annual labor income from a private (non-listed) incorporated (limited liability) business that they owned in part or in full, and zero otherwise. To study entrepreneurial survival, we create dichotomous variables that are equal to one for those who are entrepreneurs for four or more years or who are incorporated for five or more years, where four and five years are the median number of years for each category in our data (after excluding zeros). We also look at more continuous measures, such as the number of years spent as an entrepreneur or incorporated, and incomes generated from these ventures.

Our main results show that close to 25 percent of the variance in individuals' decisions to become entrepreneurs is explained by family background and community influences; for incorporation this is close to 35 percent. At the intensive margin, roughly 25 percent of the variation in *Entrepreneur* ≥ 4 *years* and more than 40 percent of the variation in *Incorporated* ≥ 5 *years*, is explained by family background and community influences. Brother correlations are always larger than sister correlations, with the largest correlation for males being close to 50 percent for *Incorporated* ≥ 5 *years*. Mixed sibling correlations are consistently smaller than same-sex correlations. We also find that the sibling correlations for widely spaced siblings are just as large as those for closely spaced siblings.[1]

As discussed in Björklund and Jäntti (2012), sibling correlations should actually be viewed as lower bounds on the importance of family background, since families may also generate im-

[1] In the bulk of the sibling correlation literature, closely spaced siblings tend to have more similar outcomes than widely spaced siblings, even when they have the same biological parents (see, e.g., Eriksson et al., 2016).

portant differences between siblings if parents treat children differently based on, for example, birth order and/or gender (see also Conley, 2004).[2] As differential treatment by gender or birth order is likely to have an impact on entrepreneurship (Black et al., 2005; Lindquist et al., 2015), and since these family-generated differences are not captured by our sibling correlations, we apply the two-step estimation method proposed by Björklund and Jäntti (2012) to assess the potential downward 'bias' of our measures of the importance of family background. On average, our revised estimates increase by 4 percent after controlling for within-family differential treatment. Thus, our sibling correlations appear to be rather tight lower bounds on the importance of family background.

We then go on to investigate the role played by specific factors in generating sibling similarities in entrepreneurship. We examine the roles played by (i) neighborhood effects, (ii) parental characteristics, (iii) genes, and (iv) sibling peer effects.

To investigate the role of neighborhood factors we estimate neighborhood correlations in entrepreneurship (see, e.g., Solon et al., 2000). Since we do find some room for neighborhoods to play a role, we do not directly contradict the findings of Giannetti and Simonov (2009) and Guiso et al. (2015), but we find that the scope for such effects in our setting is quite limited: less than 10 percent of sibling correlations or 2.5 percent of total entrepreneurial variance can be explained by common neighborhood effects shared by siblings.

To investigate the role of parental characteristics in generating sibling similarities we run an accounting exercise similar to those performed in Mazumder (2008) and Björklund et al. (2010). We re-estimate our sibling correlations controlling for observable parental characteristics. Our results show that parental entrepreneurship and incorporation are quite important, with parental education and income a distant second; together, these factors account for 20 percent of sibling correlations, and around 5 percent of total variance. Parental immigration status, which is typically associated with higher participation in entrepreneurship, does not contribute more than 1.5 percent of sibling correlations, while family structure accounts for less than 1 percent.

Interestingly, parental entrepreneurship is a prime explanatory force in individual entrepreneurship, but not incorporation, and parental incorporation explains best individual incorporation, but not entrepreneurship. This suggests that individual entrepreneurship (or, rather, self-employment) and incorporation are different aspects of entrepreneurship, in line with Henrekson and Sanandaji (2014), and that there are different transmission mechanisms

[2] Consider also the fact that each person receives all of her genes from her parents, but shares only part of her genome with her sibling, unless, of course, they are identical twins.

depending on the type of entrepreneurial engagement of the parents.

To investigate the role of genes, we perform the genetic decomposition developed by Nicoletti and Rabe (2013). Our results imply a relatively important role for genes, similar to Nicolaou et al. (2008). Nonetheless, it is difficult to measure the exact contribution of genes to sibling correlations, since we (and researchers, generally) cannot be sure that the genes siblings actually share are the ones related to the entrepreneurship outcomes.

Lastly, we perform a correlated random effects exercise (Altonji et al., 2010) to compute an upper bound on the potential contribution of sibling peer effects to our sibling correlations. Our estimated peer effects are generally non-significant (with the exception of male pairs) and very small in magnitude – less than 10 percent of the sibling correlation in absolute terms could potentially be explained by sibling peer effects. This result is corroborated by the stable sibling interaction patterns across the distribution of spacings that we observe.

We contribute to the literature on the determinants of entrepreneurship in several ways. We apply a set of methods that have not been used to analyze entrepreneurial outcomes before. Our results generate a novel measure of equality of opportunity in entrepreneurial entry and success. Our approach allows us to investigate a quite comprehensive set of potential mechanisms in order to ascertain the channels through which families influence their children's entrepreneurial choices and successes. In doing so, we put many previous results in the literature – regarding the role of neighborhoods, parental income, parental entrepreneurship, and genes – into perspective, since we are able to investigate the relative importance of various mechanisms (previously studied one at a time) within a unified empirical framework.

We conclude that parental entrepreneurship and genes are the two most important factors generating sibling similarities in entrepreneurship. Despite these findings, our main message to policy makers is that there appears to be scope for policies aimed at encouraging entrepreneurship later in life. Entrepreneurship in our setting is clearly not fully determined at birth.

The remainder of this paper is structured as follows. Section 2 describes the data used in our empirical analyses and presents descriptive statistics. In section 3, we lay out our empirical strategy and report baseline sibling correlations. In Section 4, we assess the potential role of differential treatment of siblings within the family and we present the results of this exercise. Section 5 looks at the mechanisms that drive sibling similarities: (i) neighborhood effects, (ii) parental characteristics, (iii) genes, and (iv) sibling peer effects. Each subsection presents an appropriate extension of our empirical model and results concerning each specific mechanism. Section 6 concludes.

2 Data

To create the dataset used in this paper, we start with a 25 percent random sample from Sweden's Multigenerational Register, which includes all persons born from 1932 onwards who have lived in Sweden at any time since 1961. We then match on all their brothers and sisters, where siblings are defined as those having the same biological or adoptive mother.[3] This matching is made possible by the fact that all family ties (biological and adoptive) are recorded in Sweden's Multigenerational Register. We later check that results are similar when defining the family on the basis of the father. Given the years for which entrepreneurship data is available (1985-2012), we restrict our sample to those born between 1960 and 1970: thus, we follow the oldest cohort from age 25 to 52, and the youngest cohort from age 15 to 42.[4] Those who died or left Sweden before 1985 are dropped from the sample. These cohort restrictions imply that siblings are born at most 10 years apart and that some individuals have siblings that are not included in our sample.

Consistent with the Swedish tax authorities, we define individuals as entrepreneurs when they derive the majority of their taxable labor from a business they own in full or in part. For the years 1985 to 2012, we have information on (sole and shared) business ownership for unincorporated business. For the years 1993 to 2012, we also know if a person received the majority of his or her taxable labor income from an incorporated enterprise owned in part or in full by him- or herself (and possibly employing personnel). An incorporated business in our data is a privately owned (i.e. non-listed) limited liability stock company. This type of business is better able to capture a type of entrepreneurship more likely to be associated with job creation, innovation and growth than simple self-employment (Levine and Rubinstein, 2013). In any given year, no individual is classified as both self-employed and incorporated.

We use information on the type of business owned in order to characterize people as different types of entrepreneurs.[5,6] Our first measure of entrepreneurship, *Entrepreneur*, is a dichotomous

[3] We drop all observations with missing mothers from the analysis. In total, xx observations have missing mothers; xx percent of these observations are immigrants, many of whom emigrated to Sweden without their mothers. Among the dropped observations, xx percent have been entrepreneurs, as compared to 14.6 percent of the observations kept in the sample.

[4] We impose these sample restrictions such that we are able to observe the occupational choices of the parents of the old cohort for a number of years.

[5] We do not have information before 1993 on those working in their own incorporated enterprise. This implies we are underestimating the true extent of entrepreneurship for the years 1985-1992. For 1993-2001, roughly 2 percent of the sample is in this position. This might be approximately true for 1985-1992, as well.

[6] Note that farmers are included in Statistics Sweden's definition of business owners, since farms are typically run as companies (either incorporated or unincorporated). In 2004, Statistics Sweden changed their routines for collecting information on business ownership, as well as its definition. Since then, it includes business owners who report zero profits or even losses.

variable equal to one if the individual is ever categorized as self-employed in an unincorporated business that they own in full or in part, and zero otherwise. Our second measure of entrepreneurship, *Incorporated*, is equal to one if the individual has ever been incorporated, and zero otherwise.[7] We add two stricter definitions of entrepreneurship by using the median of the number of years individuals spend in entrepreneurship (or incorporation) as a threshold in labeling an individual as an entrepreneur (or incorporated). In our sample, $Entrepreneur \geq 4y$ takes a value of one for individuals who have been entrepreneurs for at least of 4 years, and $Incorporated \geq 5y$ equals one for individuals incorporated for a minimum of 5 years. We define a *High income entrepreneur* as an individual whose entrepreneurial income is above the median.

We also create several continuous measures of Entrepreneurship. We start with the number of years individuals have been self-employed or incorporated: *Years entrepreneur* and *Years incorporated*. These variables represent established measures of entrepreneurial duration used in the literature. We also study income from entrepreneurial activities. This variable is created from annual tax register data for the years 1985-2012. For each person in each year, we have a measure of pre-tax total factor income, which includes earnings, taxable benefits (e.g. unemployment insurance, parental insurance, sick pay, etc.) and net capital gains (e.g. dividends, interest received or paid, etc.). In each year a person is self-employed, we label this income as 'entrepreneurial income'. We then take the average across all years of entrepreneurial income and then take the log of this average. This is our measure of *Entrepreneurial income*.

We have also created a set of family-wide background variables to use in our accounting exercise. We define parental entrepreneurship and incorporation in the same way as we do for their children. We also have information on parental education, immigrant status and income. The latter is defined as the log of the average of a parent's pre-tax total factor income for all available years from 1968 to 2012, calculated separately for mothers and fathers. Parental education is measured in seven different levels spanning the minimum, nine-year compulsory level to graduate school. These indicate the highest degree completed in Sweden, and as such, it is missing for older immigrants who have not attended school in Sweden.[8]

We include several measures of family type and family structure. We have created a variable for the mother's age at the birth of her first child; this is her age at her true first birth, since we have information on *all* children, including those born before 1960 that we do not include

[7] Individuals who are *Incorporated* are not a subsample of those who are *Entrepreneurs*. In any given year, these two variables are mutually exclusive. There are individuals who have been incorporated, but not self-employed, and viceversa, as well as individuals who have experienced both types of entrepreneurship at some point in their careers.

[8] In some cases, their education is still included if it has been recorded by the immigration authorities.

in our sample of siblings. Similarly, we can also create correct variables for the number of children in a family and a variable indicating whether a sibling is the first born child or not. We create a dummy variable if the father is unknown and a variable for family structure when a sibling is 15 years old, which is allowed to vary across siblings from the same family. This variable is based on information we have about actual cohabitation; it contains six categories: missing, both parents present, single mother, single father, mother with new husband, father with new wife. We have also constructed two other family structure variables, namely the mother's partner count (i.e. the number of individuals she has conceived children with) and whether the household includes both biological and adoptive children – a motive usually found in the bequests literature. Lastly, we use information on the parishes siblings live in at age 15 to define neighborhoods.[9] We have 2,650 parishes in our data: the largest includes 5,359 siblings, the smallest ones only includes one sibling (Table 2, panel D).

2.1 Descriptive Statistics

Table 1 shows the number of families with different sibship sizes. Our final sample consists of 705,262 individuals (361,556 men and 343,706 women) from 434,203 families, defined through the mother. The mean number of siblings in each family is 1.6, and in our sample 32 percent of individuals are singletons (i.e. they have no siblings included in the sample – by contrast, we only have 6 percent 'true' singletons). We include singletons to increase the precision of the estimate of between-family variation, although our results are not sensitive to their inclusion or exclusion (see Section 3.2).

Table 2 presents the descriptive statistics of our outcomes of interest in panel A: 14.6 percent of the individuals in our sample are entrepreneurs at some point, and 8.3 percent are incorporated at least once, with the average number of years spent as entrepreneur and incorporated being 5.9 and 5.8, respectively. In panel B, we examine parental characteristics: 15 percent of mothers and 25 percent of fathers have been entrepreneurs at least once and only 3.2 percent of mothers and 6.4 percent of fathers have been incorporated. Mothers and fathers have relatively similar incomes and education levels, with fathers slightly more likely to be Swedish natives.

Panel C of Table 2 shows the father is missing for 11,222 individuals, or 1.6 percent of our sample; 2 percent of individuals are twins and 1.4 percent have been adopted by either mother, father, or both, leading to only 0.6 percent of households that include both biological

[9] We thus estimate parish correlations, but tighter or looser definitions of neighborhood (schools or statistical metropolitan areas) are unlikely to induce large changes in correlations (Lindahl, 2011; Nicoletti and Rabe, 2013).

and adoptive children. The average number of children is 2.8 per family, of which we capture in our sample 1.6 children per family on average. Mothers tend to give birth in their mid-20's, and are unlikely to conceive children with more than one man (only around 3 percent do so). Our family structure variable reveals that the lion's share of families consists of intact families – almost 70 percent. Single mothers represent the second most frequent type of family – 18.54 percent, with single fathers representing only 3.79 percent of families, mothers with a new husband 5 percent and fathers with a new wife 1.97 percent.

Panel D restates the main parish characteristics and also shows that on average 17.5 percent of individuals in a parish become entrepreneurs at some point, while 8 percent found an incorporated business.

3 Sibling Correlations

Self-employment, s_{if}, for sibling i from family f can be modeled as:

$$s_{if} = X'_{if}\beta + e_{if}, \tag{1}$$

where X'_{if} includes individuals' birth year and a gender dummy for individual i from family f. The residual term, e_{if}, is an individual-specific component representing a person's position in the overall distribution of entrepreneurship, whose population variance is given by σ_e^2. Following Solon (1999), the individual variance component, e_{if}, is assumed to be comprised of two linearly additive and independent variance components:

$$e_{if} = a_f + b_{if}. \tag{2}$$

The first part, a_f, is a permanent component shared by all siblings in family f. This is what makes siblings similar. The second component, b_{if}, is the permanent component unique to sibling i in family f. The variance of e_{if} can be expressed as the sum of the stationary population variances of the permanent family and individual components:

$$\sigma_e^2 = \sigma_a^2 + \sigma_b^2. \tag{3}$$

The share of the variance in an individual's long-run probability of being an entrepreneur (or in his or her innate propensity to choose self-employment over wage employment) that can be attributed to family background effects is:

$$\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2} \equiv \text{corr}(s_{if}, s_{i'f}). \tag{4}$$

This share coincides with the correlation in self-employment of randomly drawn pairs of siblings, which is why ρ is called a sibling correlation.

This sibling correlation can be thought of as an omnibus measure of the importance of family and community effects. It includes family-wide influences that are shared by siblings, such as parental entrepreneurship, parental income, parental aspirations, cultural inheritance, genes, etc. However, it also includes shared influences that are not directly experienced in the home, such as school, church, and neighborhood effects. Genetic traits not shared by siblings, differential treatment of siblings, time-dependent changes in neighborhoods, schools, etc., are captured by the individual component b_{if}. If such non-shared factors are relatively more important than shared factors for determining entrepreneurship, the variance of the family effects will be small relative to the variance of the individual effects and the sibling correlation will be low; in other words, the more important the effects that siblings share are, the larger the sibling correlation will be.

The existence of non-shared family factors, such as differential treatment based on birth order and/or gender, implies that the sibling correlation should be viewed as a lower bound on the importance of family-background and neighborhood effects.[10] We return to this important argument when we discuss a more general model of sibling similarities in Section 4.

An estimate of the sibling correlation in long-run self-employment, ρ, can be constructed using estimates of the between-family variation, σ_a^2, and the individual variation, σ_b^2. These can be obtained by estimating the following mixed-effects model:

$$s_{if} = \mathbf{X'}_{if}\beta + a_f + b_{if}. \tag{5}$$

When the outcome variable is continuous (e.g. entrepreneurial income), we estimate this model using Stata's *mixed* command under the assumption that the two random components are independent realizations from a multivariate normal distribution with mean zero and constant variance. The variance components are estimated using the restricted maximum likelihood method.

[10] Björklund and Jäntti (2012) discuss this issue at some length and provide a quantitative example for the case of birth order. In particular, they examine the size of the advantage that first born children have over their younger siblings in terms of cognitive and non-cognitive skills, height, schooling, and earnings. They find only minor effects, which we also confirm (see Section 4).

When the outcome variable is dichotomous (e.g. entrepreneur or incorporated), we reformulate equation (5) as a latent linear response model:

$$s_{if}^{*} = \mathbf{X}'_{if}\beta + a_f + b_{if}, \tag{6}$$

where we only observe $s_{if} = I(s_{if}^{*} > 0)$. We estimate equation (6) using Stata's *melogit* command under the assumption that the random effect a_f is a realization from a normal distribution with mean zero and constant variance, while the individual variance component, b_{if}, is drawn from the logistic distribution with mean zero and variance $\pi^2/3$.

3.1 Sibling Correlations: Results

Sibling correlations are reported in Table 3. In Column (1), we see that family background and community influences account for 23 percent of the choice to ever become an *Entrepreneur* and 34 percent of the choice to ever become *Incorporated.* Looking at stricter, or more intensive, measures of entrepreneurship, we see that 26 percent of the variation in *Entrepreneur* $\geq$ 4 *years* and 42 percent of the variation in *Incorporated* $\geq$ 5 *years* can be attributed to family-wide influences that siblings share. Becoming a *High income entrepreneur* on the other hand appears to be less influenced by family background.

Overall, these sibling correlations imply that family background does matter, especially for owning an incorporated firm for a longer period of time. At the same time, one cannot conclude that entrepreneurship is mostly determined at birth. In fact, it appears mostly to be due to individual factors that siblings do not share. At first glance, equality of opportunity in entrepreneurship – who becomes an entrepreneur, and who succeeds – is similar to that obtained for educational or earnings mobility, where sibling correlation estimates range between 0.1 and 0.3 for Scandinavian countries (Björklund and Jäntti, 2011). The sibling correlations in Table 3 also show us that shared family background is a more important determinant of entrepreneurship for men than for women and that outcomes for mixed sex siblings are less similar than those of same sex siblings.

3.2 Sensitivity Analyses

Column (5) of Table 3 displays the results of our first robustness check. Here, we exclude singletons from our analysis, as they only contribute to the precision of the standard error of our between-family variation estimates. Their exclusion has little impact on the estimated sibling correlations: all the coefficients vary within a 0.002 margin (except for years incorporated), and the standard errors are virtually identical with or without singletons.

In column (6), we define siblings as having the same father, as opposed to having the same mother. While the sample size is marginally reduced (since we are missing information on approximately 11,222 unknown fathers), the estimated sibling correlations are very close to the ones obtained when the family was defined by means of the mother (column (1)). Indeed, none of the father-based sibling correlations differs by more than 0.011 from the mother-based sibling correlations. The pattern of higher sibling correlations for males remains across these alternative specifications.[11]

3.3 Sibling Spacing

We inspect heterogeneity in sibling correlations by looking at different sibling spacings, from twins (with zero spacing), through siblings born at least 12 months apart in rolling intervals of 12 months, and up to sibling spacings of 120 months.[12] There are two competing expectations about the relationship between spacings and sibling correlations (Eriksson et al., 2016). On the one hand, siblings born closer together interact more intensively, which should lead to higher sibling correlations at low birth spacings; on the other hand, much older siblings can potentially act as stronger role models, and in phenomena like entrepreneurship, it may well be that it takes longer for the older sibling to establish him/herself as an entrepreneur. For this exercise, we restrict the non-twins to full siblings in families with two children in our sample.[13]

Results for entrepreneurship and incorporation in Figure 1 suggest that while twin correlations are higher than non-twin correlations, the latter do not display an evident relationship with birth spacing (this pattern is common across outcomes, see Figure A.1 in the Appendix). This is quite interesting, given that in the bulk of the sibling correlation literature sibling spacing tends to matter quite a lot. The outcomes of closely spaced siblings are typically much more similar than those of widely spaced siblings (see, e.g., Eriksson et al., 2016).

[11] Since these mixed-effects models do not have closed-form solutions and rely on numerical optimization techniques, as additional sensitivity analyses (not tabulated for brevity), we experiment with other estimation commands using slightly different optimization procedures. For example, we estimate the extensive margin models using the user-written command *gllamm* (Rabe-Hesketh et al., 2005) to find very similar results; we also estimate the intensive margin (linear) models using maximum likelihood instead of restricted maximum likelihood and the results are identical (an expected outcome given their near equivalence in large samples). For computational efficiency, we control for linear birth year effects in the fixed part of our models, although an alternative specification using more flexible birth year dummies produces identical results, and including maternal birth year dummies to control for parental cohort effects has no impact on the results.

[12] We omit spacings between 1 and 11 months, and larger than 120 months as these are quite rare. Labels in Figure 1 imply 12-month rolling intervals, i.e. the label 12 months covers spacings between 12 and 24 months.

[13] Reassuringly, the sibling correlations for these non-twins are the same as the overall sibling correlations estimated in column (1) of Table 3, suggesting they are representative of the entire sample.

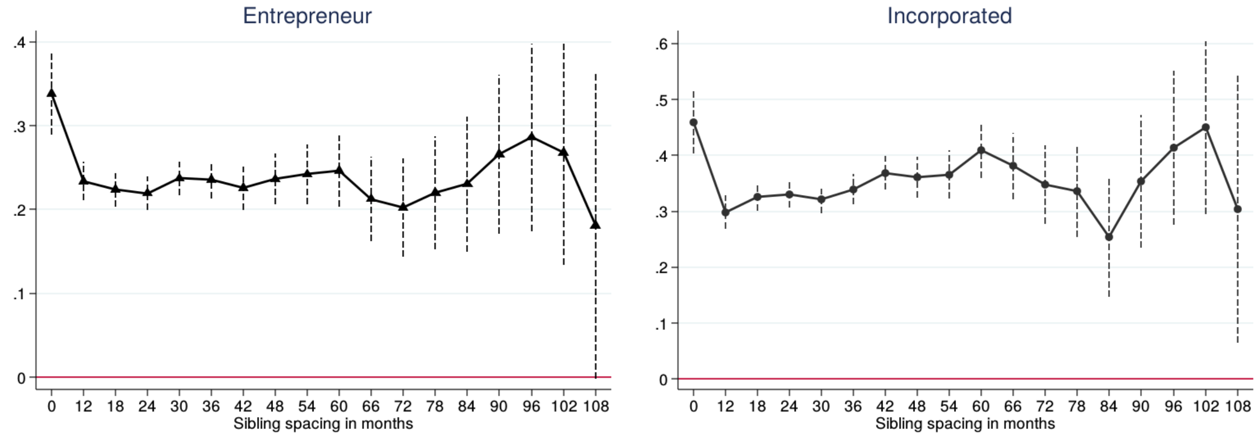

Figure 1: Comparison of twin and sibling correlations in entrepreneurship and incorporation by sibling spacing (in months).

What we can take away from this exercise is the following: (i) time-varying factors do not appear to be important, and (ii) close (day-to-day) interactions between siblings may not be important. These results also suggest that twins are more similar in entrepreneurial outcomes than non-twins, either because of genetic effects, more similar treatment by parents, or stronger inter-sibling interactions. We return more formally to genetic effects in Section 5.3, and potential peer effects in Section 5.4.

4 Sibling Differences

Sibling correlations are designed to measure sibling similarities in a given outcome. However, the estimation outlined above can only generate positive (or zero) estimates of sibling similarity, meaning that families can only make siblings alike – consistent with an assumption of parental equal concern for children across birth order and gender. In reality, families may actually act as an important source of inequality between siblings (Conley, 2004); that is, some things that families do – willingly or not – may increase the difference in outcomes between siblings. If this is the case, then the sibling correlation should be viewed as a lower bound on the importance of family background.

In a first and major contribution, Behrman and Taubman (1986) show that the model of parental resource allocation to children produces better predictions once the assumption of equal concern is relaxed and birth order effects and parental preferences are explicitly considered. While the relationship of birth order and outcomes is *a priori* ambiguous,[14] recent labor

[14] On the one hand, better biological endowment, parents' time availability, confluence effects in the presence of younger siblings, reciprocity and implicit old-age insurance, and parental preferences for observing achievement all favor lower-order births; parental experience with child rearing, a positive-sloping lifetime earnings

economics literature examining the impact of birth order on education and income generally reveals a significant premium associated with first-borns in developed economies (Black et al., 2005; Booth and Kee, 2009; Mechoulan and Wolff, 2015).

Early investments in children are sometimes reinforced by financial transfers during adulthood. While most bequests tend to be relatively equal, inter-vivos transfers display a much larger degree of inequality and may relieve financial constraints for some siblings relative to others (McGarry, 1999; Bernheim and Severinov, 2003; Behrman and Rosenzweig, 2004). These transfers appear to favor children closer to home in terms of caring for parents, who are more likely to be first-borns acting as implicit insurance during parental old age. In addition, parents are more likely to transfer resources to children involved in home production, farm or business work and to biological, rather than adoptive children (Light and McGarry, 2004).

An alternative hypothesis regarding birth order comes from psychology: even though siblings experience shared events, they have divergent interpretations of the same events, given their different ages or circumstances. Moreover, first born children tend to be associated with conformity, since they usually accept the roles that parents envision for them and embrace a more traditional outlook on life. All subsequent siblings then engage in a process of mutual differentiation or 'de-identification', whereby they adopt different roles within the family. Later born children tend to be more creative and less likely to conform to norms, which may make them more 'entrepreneurial' (Sulloway, 1996).

The (perhaps unintended) preferential treatment of first-borns, in parental attention and investments over time, may reduce the constraints these individuals face – either in skill, signaling ability or liquidity – and may facilitate their transition into entrepreneurship. Conversely, later-borns may be intrinsically more entrepreneurial, since they are more likely to be risk-taking and disruptive (Sulloway, 1996). In any case, families create differences between siblings that may have implications for entrepreneurship:[15] we thus expect our measure of family background to increase once birth order effects are accounted for.

In terms of gender, same-sex homophily suggests that daughters respond more to mothers' entrepreneurship and that sons respond more to fathers' entrepreneurship (Lindquist et al., 2015; Hoffmann et al., 2015). While this does not necessarily imply overall differential treatment, a pervasive preference for sons may still exist in a business context. Bennedsen et al. (2007) show that if the first born child of a family firm's CEO is male, his replacement is more likely to come from within the family rather than to be external. The interaction of gender, first

profile for parents, use of older siblings as assets, on the other hand, favor higher-order births.

[15] Early research by Robinson and Hunt (1992) finds no effect of birth order on entrepreneurship, but lacks methodological precision.

born and parental entrepreneurship may thus have an impact on the sibling correlations; its magnitude, however, is fairly limited, as Lindquist et al. (2015) show that less than 10 percent of firms are inherited.

In order to assess the potential effects of birth order and gender differences, we employ a two-step procedure, following Björklund and Jäntti (2012).[16] We estimate equations (5) and (6), and then predict the individual-level residuals. We subsequently regress these residuals on birth order (a dummy for being first born), gender and their interactions with parental entrepreneurship and incorporation. The R^2 from these regressions reveals how much of the individual-level variance is driven by differential treatment within the family and should thus be counted towards the explanatory power of family background. We compute a 'revised' estimator of the importance of family background, $\tilde{\rho}$:

$$\tilde{\rho} = \frac{\sigma_a^2 + R^2 \sigma_b^2}{\sigma_a^2 + \sigma_b^2}. \tag{7}$$

The difference $\tilde{\rho} - \rho$ reflects the importance of sibling gender and birth order when these characteristics lead to differences in outcomes between siblings.

4.1 Sibling Differences: Results

Results in Table 4 suggest the second-stage R^2 is smaller than 0.03, and that less than 2 percent of total variance should be transferred from the individual to the family level. Alternatively put, sibling correlations increase on average by 4 percent once differential treatment by gender and birth order within the family are accounted for. However, the confidence intervals of the original and the revised estimates are largely overlapping, suggesting that the 'bias' created by not controlling for sibling differences is rather small: the sibling correlation is thus a sufficiently tight bound on the importance of factors that individuals have little control over.

5 Mechanisms that Generate Sibling Similarities

In this section, we take a closer look at factors that make siblings similar and that are captured by sibling correlations, focusing on the effect of communities, the effect of specific (observable) parental characteristics, the effect of shared genes, and inter-sibling peer effects.

[16] Ideally, one would estimate a very general mixed effects model that includes many random effects and random interaction effects at the individual level. Unfortunately, this is not feasible in practice, particularly when estimating logit models.

5.1 Neighborhoods

Even though sibling correlations give a relatively accurate overview of how much family background influences entrepreneurship outcomes, they also capture other environments that siblings share, such as the community of residence. In his review of the determinants of entrepreneurship, Parker (2009) comments that "[a]ll major economies exhibit regional differences in rates of entrepreneurship" (p. 147), a note echoed by research into clusters of entrepreneurship (Glaeser et al., 2010). Giannetti and Simonov (2009) show that in Sweden the between-municipality variance is much larger than within-municipality variance in entrepreneurship and that a standard deviation increase in the proportion of entrepreneurs at a local labor market level leads to about 25 percent more entry into entrepreneurship.[17] Similarly, Guiso et al. (2015) find a positive effect of local firm density (in the individuals' region or province of residence at age 18) on entry into entrepreneurship. In contrast to Giannetti and Simonov (2009), Guiso et al. (2015) show that a higher firm density leads to higher income in entrepreneurship and the adoption of better management practices, which suggests that exposure to entrepreneurship aids learning.

Based on the literature on sibling correlations in education, incomes, and crime, we do not expect neighborhood effects to be large – they typically explain less than 10 percent of variance (Solon et al., 2000; Page and Solon, 2003a,b; Lindahl, 2011; Nicoletti and Rabe, 2013; Eriksson et al., 2016). Nevertheless, we deem it undesirable to disregard them, as they allow us to more accurately pin down the share of variance explained by the family. Therefore, we use data on the parish the individual resided in at age 15 to capture neighborhood effects. First, we substitute a neighborhood variance component (c_n, where n indexes parishes) for the family variance component and estimate the following two-level mixed effects model:

$$s_{in} = \mathbf{X'}_{in}\beta + c_n + b_{in}. \tag{8}$$

The neighborhood correlation then becomes:

$$\rho_{parish} = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_b^2}. \tag{9}$$

Alternatively, we can estimate three-level models, nesting individuals within families within

[17] This result is robust to the inclusion of controls for parental wealth and municipality fixed effects.

neighborhoods (the effect of this constraint is explained below):

$$s_{ifn} = \mathbf{X}'_{ifn}\beta + c_n + a_{fn} + b_{ifn}, \tag{10}$$

for individual i in family f in neighborhood n. In this model, we do not control for parental characteristics, as they are shared by parents and should be captured by the family within neighborhood correlation.

5.1.1 Neighborhoods: Results

Table 5, panel B shows the results of the two-level model for extensive and intensive outcome margins: the neighborhood correlation ranges between 0.012 and 0.047, meaning that shared neighborhood characteristics can explain at most 4.7 percent of entrepreneurial variance. Importantly, neighborhood correlations are small relative to our baseline sibling correlations (reproduced in panel A of Table 5), with an explanatory power of usually less than 12 percent of the sibling correlation; the sole exception being for *Years entrepreneur*, where neighborhoods explain 21.9 percent of the sibling correlation, consistent with the increased benefits of entrepreneurship status in more entrepreneurial neighborhoods (Giannetti and Simonov, 2009).

One concern with this method is family selection into neighborhoods (Solon et al., 2000). To account for this and obtain a tighter bound on neighborhood effects (Björklund and Jäntti, 2011), we add a set of family-level characteristics to the fixed part of equation (8): parental incomes, entrepreneurship, incorporation, education and immigrant status.[18] These new neighborhood correlation estimates are shown in panel C of Table 5 and are much smaller than those in panel B, on average by around 40 percent. Once we control for parental characteristics, neighborhood characteristics explain at most 7 percent of the sibling correlation, again with the exception of *Years entrepreneur*, where the corresponding value is 11.39 percent.

While the two-level models provide suggestive evidence of the limited scope for neighborhood effects in entrepreneurship, they only provide an indirect comparison of neighborhood and sibling correlations. The results in panel D of Table 5 show very similar neighborhood correlations to the ones in panel B, while the family within neighborhood correlations are on the same order of magnitude as the sibling correlations in panel A, revealing the consistency of our results.

[18] In Panel B of Table A.1 in the appendix, we show the neighborhood correlations controlling for parental characteristics separately in order to control for sorting. Parental entrepreneurship fares better in explaining entrepreneurship outcomes and parental incorporation does better in explaining incorporation outcomes; parental income seems to explain relatively more of the neighborhood correlation in *Entrepreneurial income*, but even then the magnitude of the effect is fairly minor (0.008 of 0.023).

Perhaps counter-intuitively, some of the family within neighborhood correlations are actually larger than the sibling correlations, but the explanation is fairly straightforward. One implication of the model in equation (10) is that we constrain families not to be spread across neighborhoods. In contrast, a large family may have some siblings in one neighborhood and other siblings in a different neighborhood:[19] it is then clear that siblings within the same neighborhood are more similar than siblings in different neighborhoods (i.e. the neighborhood effects capture some family non-intactness), which can explain the potential downward bias of the baseline sibling correlation.

An alternative explanation is that we record the parish of residence when the individual is 15 and allow the recording date to vary across individuals. This may create some measurement error in our definitions of neighborhood, which could account for the minor differences we observe. However, both Lindahl (2011) and Kunz et al. (2003) show that point-in-time measures of neighborhood are good proxies of long-run neighborhood characteristics.

Overall, we provide evidence that the scope for neighborhood effects in entrepreneurship is limited, usually less than 10 percent, very much in line with previous literature using income, education or crime as outcomes. This result puts into perspective the previous results of Giannetti and Simonov (2009) and Guiso et al. (2015), indicating that while there are some effects of community, most learning and role-modeling take place outside of the informal neighborhood environment. Thus, communities have only a minor role to play in reducing equality of opportunity in entrepreneurship, or differently put, entrepreneurship policy at the community level may have only a small role in fostering entrepreneurship.

5.2 Parental Characteristics

Which components of family background are mostly responsible for generating sibling similarities in entrepreneurship? We study this question by including potentially important family-wide variables, either one at a time or simultaneously, in the $\mathbf{X'}_{if}$ matrix of equation (5) (i.e. to the fixed part of our mixed effects model). For example, consider the inclusion of mothers' and fathers' entrepreneurship in $\mathbf{X'}_{if}$. These two additional control variables (fixed effects) should reduce the residual variation in the outcome variable and produce a lower estimate of the between-family variation, σ_a^{2*}, than the estimate produced without the added controls. Abstracting from measurement error, we can interpret the difference between these two estimates, $\sigma_a^2 - \sigma_a^{2*}$, as an upper bound on the amount of the variance in the family component that can be

[19] These models thus artificially expand the number of families from 434,203 to 459,789.

explained by parental entrepreneurship. It is viewed as an upper bound since it includes other factors affecting children's entrepreneurship that are correlated with parental entrepreneurship (for instance, education, occupation, residence).[20] This exercise also produces a new sibling correlation, ρ^*. From what we know about the relationship between parents' and children's entrepreneurship (Lindquist et al., 2015), we expect this new sibling correlation to be significantly lower, but still substantial in magnitude.

While this accounting exercise does not allow for a causal interpretation of the determinants of entrepreneurship, we can judge the 'importance' of any particular control variable by using as a metric the degree to which it lowers the sibling correlation after being included in the fixed part of our mixed-effects model (Mazumder, 2008; Björklund et al., 2010). We will use this metric to provide clues about what is potentially important and unimportant in explaining sibling similarities in entrepreneurship. Specifically, we will explore the potential roles played by: (i) parental education and income, (ii) parental entrepreneurship and incorporation, (iii) parents' immigration status, and (iv) family structure.

Previous research has suggested an important role of parental income and education (Lentz and Laband, 1990; Blanchflower and Oswald, 1998; Fairlie and Robb, 2007a); finding a large role for these variables would be consistent with the existence of capital constraints (Holtz-Eakin et al., 1994; Blanchflower and Oswald, 1998). Parental entrepreneurship and incorporation are likely to influence the occupational preferences of individuals, not only through the acquisition of general or specific business human or social capital, but also through role-modeling (Dunn and Holtz-Eakin, 2000; Fairlie and Robb, 2007b; Sørensen, 2007; Parker, 2009; Hoffmann et al., 2015; Lindquist et al., 2015). Ethnicity and parental immigration are also likely to play a role in entrepreneurship decisions – in terms of the location of new immigrants and their subsequent choice of business (Dunn and Holtz-Eakin, 2000; Edin et al., 2003; Andersson and Hammarstedt, 2010; Kerr and Mandorff, 2015). Finally, although family structure is potentially associated with personality developments affecting entrepreneurial decisions, it has been understudied as a determinant of entrepreneurship, mainly given a lack of reliable data. Previous studies find only a limited association of family structure with entrepreneurship (Dunn and Holtz-Eakin, 2000; Hout and Rosen, 2000; Tervo, 2006), but employ weak and simplistic measures of family structure (which we improve on as described in Section 2). By controlling for these observables one by one and then simultaneously, we can assess both their relative and their total contribution to entrepreneurial variance.

[20] In the presence of measurement error, this difference, $\sigma_a^2 - \sigma_a^{2*}$, is more correctly viewed as a downwardly biased estimate of the upper bound on the amount of the variance in the family component that can be accounted for by parental entrepreneurship.

5.2.1 Parental Characteristics: Results

We proceed to address these specific family factors that contribute to sibling similarities by adding a set of covariates to the fixed part of the models, as explained above. These covariates include: mother's and father's education, income, immigration, entrepreneurship and incorporation, family size, the mother's age at first birth and partner count, whether both biological and adoptive children were present in the household, and our dedicated measure of family structure. To simplify the exposition, we have performed a factor analysis,[21] generating six orthogonal factors that load onto (i) parental education and income, (ii) parental immigration, (iii) parental entrepreneurship, (iv) parental incorporation, (v) a composite measure of family structure, based on family size, mother's age at first birth and the mother's partner count, and (vi) the presence of different types of genetic siblings and our objective measure of family structure (these latter two factors explain little of the sibling correlation, and will be added together to assess the importance of family structure). We then add these factors separately and jointly in the fixed part of the model to obtain the new sibling correlations, ρ^*.

Since the factor analysis requires individuals to have information on *all* these variables, our sample size is slightly reduced (for extensive margin outcomes, for instance, it is reduced from 705,626 to 665,665 individuals). Therefore, we re-estimate the sibling correlations for this particular sample and report them in panel A of Table 6. While these sibling correlations are significantly different in a statistical sense from the baseline sibling correlations (given that our estimates are extremely precise), they are on the same order of magnitude as previously estimated.

Panel B of Table 6 shows that parental education and income explain less than 5 percent of entrepreneurial outcomes, with the unsurprising exception of entrepreneurial income, where the corresponding value is 11.38 percent. This means that while there is some degree of association in parent's incomes and the subsequent income an individual makes as an entrepreneur, the explanatory power is very low (the difference between the sibling correlation before and after controlling for education and incomes is 0.024, or 2.4 percentage points). While we do not possess wealth data, our results tentatively imply that capital constraints arguments building on parent's incomes as a determinant of entrepreneurial success lack strong empirical evidence (Holtz-Eakin et al., 1994; Blanchflower and Oswald, 1998; Hurst and Lusardi, 2004).

Having non-native parents does not seem to have a large separate impact on entrepreneurship outcomes, usually less than 1 percent, at first glance in contrast to Andersson and Ham-

[21] The results of this analysis, i.e. factor loadings, are given in Table A.2 in Appendix A.

marstedt (2010). However, since our factors are orthogonal, we can expect that higher rates of immigrant entrepreneurship to be captured by our entrepreneurship factor. The latter has a higher explanatory power, especially in the individual ever becoming an entrepreneur, in the likelihood that he remains an entrepreneur for more than the median number of years and for years as entrepreneur, with the factor explaining as much as 15 percent of outcomes. Conversely, parental incorporation has very small impacts on simple self-employment outcomes, but is a strong predictor of incorporation outcomes (incorporation, being incorporated for more than the median number of years, years incorporated), explaining as much as 16.4 percent of variation. The slightly larger effects found for incorporation than for entrepreneurship are consistent with the results in Lindquist et al. (2015).

Turning to our (composite and direct) measures of family structure, we find their explanatory power to be extremely limited, up to 1 percent. It does not appear likely, then, that family structure drives the sibling correlations we observe, and that economic, rather than purely sociological family factors are important for entrepreneurship outcomes. In that sense, our results echo those that Björklund et al. (2007) obtain for schooling and earnings.

Table 6, panel C shows the sibling correlations we obtain when we add the six factors pertaining to family characteristics jointly to the fixed part of the model. The explanatory power of family observables ranges between 9.91 percent for years incorporated and 20.52 percent for entrepreneurial income. In terms of *total* variance of entrepreneurial outcomes, back-of-the-envelope calculations show that only between 2 and 8 percent of variation can be explained by observable family characteristics – a very limited role indeed.

As a robustness check, we have also estimated the joint contribution of the separate variables to the sibling correlations (instead of the factors obtained through factor analysis), and results are slightly larger, but on the same order of magnitude. In this case, families explain at most 8 percent of the *total* variation in entrepreneurial outcomes.

5.3 Genes

As suggested by Björklund et al. (2005) and Conley and Glauber (2005), shared genetic endowment is also captured by sibling correlations. Despite failing to agree on the distribution of nature and nurture, the literature has established a definite role for genetic endowment in entrepreneurship (Nicolaou et al., 2008; Lindquist et al., 2015). In the Swedish context, we expect that around one third of the sibling correlation should be explained by genetic factors: in a study of adoptees, Lindquist et al. (2015) find that the parental influence in entrepreneurship is driven by nurture (roughly two thirds), rather than nature (roughly one third). By contrast, in their study of Swedish twins, Zhang et al. (2009) find a strong genetic effect and no effect of shared environment for women, but a large shared environment influence for men, with a zero genetic effect.

In the absence of knowledge about which pairs of twins are mono- or di-zygotic (i.e. MZ and, respectively, DZ), we impose a series of assumptions in order to identify correlations for these pairs of twins (i.e., we use the approach outlined in Nicoletti and Rabe, 2013).[22] The most important source of information comes from directly observable mixed sex twins, who must be DZ twins. Thus, we can directly estimate $\sigma^2_{DZ,FM}$, that is, the variance of the family component for DZ twins of mixed sexes (with subscripts F for female and M for male). In order to calculate the corresponding variances for same-sex DZ twins, $\sigma^2_{DZ,MM}$ and $\sigma^2_{DZ,FF}$, we make the following assumptions:

$$\sigma^2_{DZ,MM} = \sigma^2_{DZ,FM} + (\sigma^2_{NT,MM} - \sigma^2_{NT,FM}), \tag{11}$$

$$\sigma^2_{DZ,FF} = \sigma^2_{DZ,FM} + (\sigma^2_{NT,FF} - \sigma^2_{NT,FM}), \tag{12}$$

where NT denotes non-twins. Intuitively, we assume we can approximate gender differences in the variance of the family component for DZ twins reasonably well by gender differences in non-twin sibling pairs, using closely spaced non-twins (born between 12 and 24 months apart).[23]

In order to identify the corresponding expressions for MZ twins, we make use of the weak assumption that the variance of the family component for all same sex twins is a weighted average of corresponding variances for MZ and DZ twin pairs, with weights provided by their incidence in the population of same sex twins.[24] Denoting these proportions with P (and noting that $P_{MZ,MM} + P_{DZ,MM} = P_{MZ,FF} + P_{DZ,FF} = 1$), we use:

$$\sigma^2_{T,MM} = \sigma^2_{MZ,MM} \times P_{MZ,MM} + \sigma^2_{DZ,MM} \times P_{DZ,MM} \tag{13}$$

$$\sigma^2_{T,FF} = \sigma^2_{MZ,FF} \times P_{MZ,FF} + \sigma^2_{DZ,FF} \times P_{DZ,FF}. \tag{14}$$

Now we need to approximate these proportions in order to solve the MZ-variances from

[22] This approach has also been used by Björklund and Jäntti (2012).

[23] We also experiment with different non-twin pair spacings as robustness checks (12 to 18 months, 12 to 48 months) and obtain very similar results. Figure A.1 in Appendix A also shows that only small differences are to be expected. The figures also reveal that twin correlations are not higher just due to different sibling interaction patterns, as these seem to be relatively constant across the spacing distribution. The correlations at different spacings for the intensive margin produce noisier estimates, given their much smaller sample size.

[24] An implicit assumption is that the means of the family components for the two types of twins are identical.

equations (13) and (14). For this purpose, and using the fact that boys and girls are conceived with almost equal probabilities, we assume 50 percent of DZ twins have mixed sexes, 25 percent are DZ twin sisters and 25 percent are DZ twin brothers. From this assumption follows that we have the following numbers (N) of different twin types:

$$N_{DZ,MM} = N_{DZ,FF} = 0.5 \times N_{DZ,FM}, \tag{15}$$

$$N_{MZ,MM} = N_{T,MM} - N_{DZ,MM} \text{ and } N_{MZ,FF} = N_{T,FF} - N_{DZ,FF}, \tag{16}$$

where we have $N_{DZ,FM}$, $N_{T,MM}$ and $N_{T,FF}$ directly observable in our data. From these numbers, we can compute the proportions needed in equations (13) and (14). We estimate the variance of the error component for the individual in exactly the same way for continuous outcomes and maintain the usual $\pi^2/3$ individual-level variation for dichotomous outcomes. Once this information is available, we can compute and compare the sibling correlations separately for MZ and DZ twins for each of the two genders, i.e. $\rho_{MZ,FF}$, $\rho_{MZ,MM}$, $\rho_{DZ,FF}$, and $\rho_{DZ,MM}$.

In order to measure the relative contribution of genes to the sibling correlation, we compute an indicator of heritability by exploiting differences in shared genetic endowment between MZ and DZ twins (Guo and Wang, 2002; Björklund et al., 2005; Rabe-Hesketh et al., 2008). As sibling correlations represent the total contribution of shared factors to variation in the outcomes, we express them as a linearly additive function of common genes and common environment: $a_f = a_{genes} + a_{env}$. These correspond closely to the A and C factors included in the structural equation model approach used by Nicolaou et al. (2008). Since MZ twins share 100 percent of their genes, we decompose the sibling correlation into genetic variation – heritability – and environmental variation: $h^2 + c_T^2 = \rho_{MZ}$, where the subscript T indicates twins' common environment. Analogously, for DZ twins who share 50 percent of their genes, we obtain: $0.5h^2 + c_T^2 = \rho_{DZ}$, with the crucial assumption that the proportion of variance owed to shared environmental influences is the same for MZ and DZ twins. It is then simple to back out measures of heritability for each gender, $h_{MM}^2 = 2 \times (\rho_{MZ,MM} - \rho_{DZ,MM})$ and $h_{FF}^2 = 2 \times (\rho_{MZ,FF} - \rho_{DZ,FF})$.

To compute sibling correlations for MZ and DZ twins, we estimate a multilevel model to which we add a random coefficient for all the different and observable pairs (non twins, male, female and mixed gender, DZ mixed gender twins, twin males and females), suppress the constant at the pair level in order to estimate a variance parameter for each different pair, and add the interaction of controls with the different pairs in the fixed part of the model.

This regression is performed on a subsample, containing only twin pairs and families with two children, which allows us to use almost half of our sample. Based on the estimated variance parameters we then compute the MZ and DZ male and female sibling correlations and compute the indicators of heritability.

5.3.1 Genes: Results

In this section, we perform the decomposition suggested by Nicoletti and Rabe (2013), and later used by Björklund and Jäntti (2012), which exploits the information that twin pairs where one child is male and the other is female must be dizygotic twins. The results in Table 7 show that, overall, genes explain between 30 percent (for incorporation) and 50 percent (for entrepreneurship) of total variance. These results are fully in line with those in the literature, but there is no clear pattern for heritability estimates by gender:[25,26] for instance, female heritability in entrepreneurship is 0.628 (with a *p*-value of 0.001), while male heritability in being incorporated for more than 5 years is 0.127 (with a *p*-value of 0.614). Additionally, most heritability estimates are insignificant, especially for stricter definitions of entrepreneurship.

One important caveat is that while, ideally, this decomposition would only capture genetic influences, in practice it may also partially account for sibling peer effects. If genetically more similar pairs of twins also interact more intensively (or are treated more similarly by parents), then inter-sibling peer effects are also captured by the decomposition. For this reason, Zhang et al. (2009) control for twins' interaction intensity in their estimation of genetic influences (in their data, identical twins' interactions are 50 percent stronger than those of fraternal twins). Unfortunately, we do not have direct information on sibling interaction.

One would be tempted to consider that sibling correlations place an upper bound on family influences, and thus implicitly a *maximum maximorum* upper bound on genetic influences. As such, this upper bound (25 percent for entrepreneurship) would be lower than previously estimated: for instance, Nicolaou et al. (2008) and Nicolaou and Shane (2010) suggest that around 40 percent of the total entrepreneurial variation (in the UK and the US) is due to genetic influences. However, our results (as well as those previously obtained in the twins literature) only speak to sharing an *entire* genome – it is then unclear how to interpret the results at the

[25] The higher heritability of entrepreneurship compared to incorporation could perhaps be driven by an innate 'taste for entrepreneurship'. The higher heritability of females in entrepreneurship can be taken to suggest a stronger reliance on genes when the environment is less favorable, in line with Zhang et al. (2009).

[26] Results when using non-twins spaced 12 to 18 months, or 12 to 48 months are relatively similar, albeit on average slightly attenuated. One could consider the results we obtain as conservative, and the heritability estimates as upper bounds on the importance of genes, given violations of the equal environment assumption used to justify the equality of the environment components for MZ and DZ twins.

level of the entire population. Indeed, for two non-twin full siblings who share half their genes, it can be that they share *all* or *none* of the genes that influence entrepreneurship. It is thus difficult to compute the relationship between genetic effects and the sibling correlation itself – a criticism we levy against all similar models –, while we can certainly acknowledge a relatively strong impact of genes on total entrepreneurial variation. Having said that, this analysis also makes genes our number one candidate for the part of the sibling correlation that cannot be explained by (i) neighborhood effects, (ii) observable parental characteristics, and (iv) sibling peer effects.

5.4 Sibling Peer Effects

Consistent with the idea of individual liability to contextual influences, sibling correlations also capture inter-sibling interactions; while these could be treated as a nuisance in estimating the impact of shared family background, we consider such sibling peer effects to be an integral part of shared environments – and policies usually take into account multiplier effects generated by peer effects. In the entrepreneurship literature, peer effects have convincingly been identified within the workplace (Nanda and Sørensen, 2010) and within universities (Lerner and Malmendier, 2013; Kacperczyk, 2013), based on (quasi-) random assignment of employees to workplaces or students to classes. In addition, within-family role-modeling has been proposed as a mechanism for intergenerational transmission of entrepreneurship (Lindquist et al., 2015; Hoffmann et al., 2015), and we examine the role-modeling potential of sibling peers.

While we lack a formal randomization process, by exploiting differences in the timing of entrepreneurial entry for sibling pairs, we can gain important information about spillovers from one sibling to the other. A useful method for exploring (causal) peer effects has been proposed by Altonji et al. (2010), who apply it to the study of illegal substance abuse, and subsequently used by Eriksson et al. (2016) to look at criminal activity. The method relies on the relatively strong assumption that only older siblings can influence younger siblings and uses correlated random effects in a dynamic panel structure. While their method is intuitively applicable to situations where peer effects are likely to dominate other causes and where individuals are active when young, entrepreneurship represents an occupational choice with long term consequences, and it is not clear that older siblings necessarily engage in entrepreneurship earlier than younger siblings.[27]

Since this exercise focuses on explaining the variance of entrepreneurial outcomes due to the

[27] However, older siblings enter the labor market earlier, and are statistically more likely to become entrepreneurs before the younger siblings, especially at large birth spacings.

influence of sibling peers rather than on identifying causal effects, we take an agnostic approach to applying the Altonji et al. (2010) model. We estimate both the effects of the older sibling on the younger one, and the effect of the younger sibling on the older one, subsequently converting the results into correlations to assess the contribution of peer effects to the sibling correlation, as will become clear below.

We begin by estimating the raw association between sibling i's entrepreneurship (incorporation) at time t, $S_t^{i'}$, and sibling i''s entrepreneurship (incorporation) at time $t-1$, $S_{t-1}^{i'}$:

$$S_t^i = \beta_0 + \beta_1 S_{t-1}^{i'} + u_t^2, \tag{17}$$

where the family subscript f is suppressed. We then add the set of controls used in the accounting exercise, X^f, and age dummies age_t^i for the focal sibling i:

$$S_t^i = \beta_0 + \beta_1 S_{t-1}^{i'} + X^f + age_t^i + \epsilon_t^2. \tag{18}$$

We estimate equations (17) and (18) (corresponding to columns (1), (2), (6) and (7) in Tables 8 and A.3-A.6) by using the panel structure of our data, limiting the sample to families with two children.[28] We later split the sample into pairs of males, females, and mixed gender pairs, where the younger sibling is male or female. We use logistic regressions in order to maintain consistency with previous estimation techniques, and we report both odds ratios and (approximated) sibling correlations, as explained in the notes to Table 8.

Part of the effect of sibling i''s entrepreneurial status on sibling i's entrepreneurial status estimated in equation (18), however, may be due to correlated random family effects, rather than direct peer effects. Altonji et al. (2010) suggest the use of a correlated random effects regression to isolate the direct sibling effect, achieving causal inference by assuming one-directional causation (whereas our study does not attempt to obtain causality); they control for the sum of sibling i''s entrepreneurial status at time $t-1$ and $t+1$ to net out the unobservable family component. We can then write:

$$S_t^i = \beta_0 + \beta_1(S_{t-1}^{i'} + S_{t+1}^{i'}) + \lambda_0 S_{t-1}^{i'} + X^f + age_t^i + age_t^{i'} + \epsilon_t^2, \tag{19}$$

where the direct (lagged) sibling effect is captured by λ_0.[29] Similarly, we can also include a direct contemporaneous sibling influence by including sibling i''s entrepreneurial status at time

[28] Sibling correlations for this sample closely match those reported in Table 3; see also footnote 13.
[29] A detailed description of the assumptions and mechanics of this model is provided in Altonji et al. (2010).

t, $S_t^{i'}$, in conjunction with an expanded control for correlated random effects:

$$S_t^i = \beta_0 + \beta_1(S_{t-1}^{i'} + S_t^{i'} + S_{t+1}^{i'}) + \lambda_0 S_{t-1}^{i'} + \lambda_1 S_t^{i'} + X^f + age_t^i + age_t^{i'} + \epsilon_t^2, \tag{20}$$

where λ_1 is the estimate of the contemporaneous effect. This estimate should not be (over-) interpreted as a contemporaneous effect, but rather as a transitory and common shock to both siblings in the same family (e.g. the introduction of a tax credit for small businesses). Hence, we do not sum the lagged and contemporaneous sibling effect when analyzing the contribution of peers to the sibling correlation (in contrast to Eriksson et al. (2016), for instance). Results for equations 19 and 20 are given in columns (3), (4), (8) and (9) of Tables 8 and A.3-A.6, while columns (5) and (10) present results from a variation of equation (17), where the lagged sibling effect is replaced by the contemporaneous one (this equation being necessary for calibration purposes). We now turn to the results of this exercise, focusing less on technical aspects and more on implications for sibling correlations.

5.4.1 Sibling Peer Effects: Results

Table 8 shows the results of our sibling peer effects exercise on the sub-sample of sibling pairs, with panel A referring to entrepreneurship and panel B to incorporation; in columns (1) to (5) sibling i is the younger one, whereas in columns (6)-(10), sibling i is the older one. The results suggest a positive and significant (at 5 percent) impact of the older sibling's entrepreneurial status at time $t-1$ on the younger sibling's entrepreneurial status at time t (Table 8, column (3)); this translates into a sibling correlation $\rho = 0.014$ as given by the lagged sibling effect, representing 6.20 percent of the baseline sibling correlation, as seen in Table 9.

Conversely, the (non-causal) effect of younger sibling on the older one in entrepreneurship appears largely negative, producing negative correlations; this implies that peer effects may actually generate sibling dissimilarity. When we perform the exercise for each type of sibling pair, we note that the direct peer effects for this sub-sample are driven by the peer influence between brothers, which reflects the same pattern, with very similar magnitudes. This heightened interaction between brothers may be responsible for the higher sibling correlation in entrepreneurship we observe for males in Table 3. The estimates of peer effects among sisters are similar in magnitude, but not significant, which may imply a higher threshold for women with regards to entrepreneurial entry.

With regards to incorporation, most estimated peer effects are not significant. Notwithstanding the lack of significance, the pattern is seemingly reversed: the older sibling's entrepreneurial entry is negatively associated with the younger sibling's entry, while the younger sibling's entry is positively associated with the older sibling's entry. For both entrepreneurship and incorporation, if instead of estimates from columns (3) and (8) of Tables 8 and A.3-A.6 we use the estimates of columns (4) and (9), where we control for contemporaneous shocks, results are very similar, as Table 9 shows.

All in all, the timing of entrepreneurial entry and the subsequent peer influence are only significant (at 5 percent) for male pairs, and even then are smaller than 10 percent;[30] thus, peer effects are too small in magnitude to drive the sibling correlation, and at times may even create sibling dissimilarities. What exactly the mechanisms at play are, such as positive role-modeling, or an update of expectations regarding entrepreneurial success, is left for future research.

6 Discussion and Conclusion

The sibling correlation models we have applied to entrepreneurship outcomes reveal that there is a substantial degree of equality of opportunity in entrepreneurship, with less than 25 percent of outcomes explained by family variance for entrepreneurial entry, duration and incomes, with a corresponding value of 45 percent for incorporation outcomes. Our measures of equality of opportunity in entrepreneurship are similar to those obtained for incomes and earnings in Sweden (Björklund and Jäntti, 1997, 2012; Björklund et al., 2009, 2010), and our sibling correlations are (as expected) smaller than those found for height, cognitive and non-cognitive ability, grades or years of schooling. By most criteria, this result represents a positive phenomenon, since it ultimately means that barriers to entrepreneurship are small – and quite surmountable (although for incorporation this is true to a smaller extent).

We also purge our sibling correlation of neighborhood effects (and the sorting of families into neighborhoods) by applying both two-level and three-level models: only 10 percent of sibling correlations are explained in this way, consistent with, for instance, Solon et al. (2000) and Eriksson et al. (2016). This translates into an approximately 2.5 percent neighborhood effect on total entrepreneurial variance. While we do not deny positive effects of entrepreneurial neighborhoods on entrepreneurship, either in learning or favorable social status considerations (Giannetti and Simonov, 2009; Guiso et al., 2015), our results show that their total impact on entrepreneurial entry and attainment is likely to be very small.

For parental characteristics, parental entrepreneurship and incorporation tower over all

[30] The only exception is the effect of a younger sister on a younger brother, where the negative association is 12-14 percent of the sibling correlation in absolute terms, and only significant at 10 percent.

other observables and explain around 15 percent of the sibling correlations; parental education and incomes explain a further 5 percent of the sibling correlation. Altogether, these variables explain around 5 percent of total entrepreneurial variation, with implications for the literature on capital constraints in entrepreneurship. A positive relation of parents' income and wealth to individuals' entry has sometimes been taken as suggestive of capital constraints (Holtz-Eakin et al., 1994; Lindh and Ohlsson, 1996; Blanchflower and Oswald, 1998) – we consider their impact to be small and tentatively interpret this as limited evidence of capital constraints.

Interestingly, parental entrepreneurship is a prime explanatory force in individual entrepreneurship, but not incorporation, and parental incorporation explains best individual incorporation, but not entrepreneurship. This suggests that individual entrepreneurship (or, rather, self-employment) and incorporation are different aspects of entrepreneurship, in line with Henrekson and Sanandaji (2014), and that there are different transmission mechanisms depending on the type of entrepreneurial engagement of the parents.

Since siblings share not only the family environment, but also their genes, we also attempt to get at genetic effects by using the Nicoletti and Rabe (2013) decomposition: while we cannot reject the existence of genetic effects (up to 50 percent of total variation for some of the outcomes, but not always significant), we find it difficult to interpret the relationship between sibling correlations and genetic effects. This remains an open question for future research, although our results suggest that there is sufficient scope for learning how to become an entrepreneur. With regards to sibling peer effects, we find only weak (and non-causal) evidence that brothers influence each other, and the effect is limited in magnitude, i.e. less than 10 percent of the sibling correlations. This result may explain why we find no distinctive pattern for the relationship between birth spacing and sibling correlations. Ultimately, it appears unlikely that sibling peer-effects drive the sibling correlation.

Overall, our results indicate that family background – over which individuals have very little control, if any – does not explain the majority of variance in entrepreneurial entry, persistence and success, and that equality of opportunity in entrepreneurship (and entrepreneurial income) is compatible with that observed for employment income. Furthermore, birth circumstances do not predetermine entrepreneurship outcomes, and environmental influences account for most of the variation: learning to be a successful entrepreneur is not (entirely) conditioned by background and policy may have substantial leverage in this domain. For incorporation – a measure we consider to be more reflective of commitment to entrepreneurship –, the results are less favorable, with almost half the variation explained by factors the individuals has no control over. The results imply a clear distinction between entrepreneurship (self-employment) and incorporation, with entrepreneurship more likely to be driven by preferences, rather than ability (Henrekson and Sanandaji, 2014; Hurst and Pugsley, 2015).

Last but not least, a large fraction of our sibling correlations remains unaccounted for, i.e. it comprises unobserved family characteristics; these may include, but are not limited to, managerial or entrepreneurial ability (Lucas Jr., 1978), risk preferences (Dohmen et al., 2012), a wider set of values (Albanese et al., 2016), and even latent health (Ahlburg, 1998). Capturing such variation would be an interesting avenue for future research, although one could argue part of these effects is captured through the various observable parental characteristics we account for (e.g. parental risk preferences determine parental entrepreneurship). In addition, a future reconciliation of heritability and sibling correlations could shed more light on the importance of genes in generating sibling similarity.

Our paper is not without limitations. First and foremost, we cannot claim causality. In our context, however, causality is less important than partitioning entrepreneurial variance into the relevant sources. Second, since we measure the degree to which siblings are similar, we cannot make precise inferences regarding single-child families, although excluding or including singletons has no impact on our results. Moreover, we have sufficient flexibility in analyzing family characteristics, meaning that we capture a high degree of heterogeneity. Third, our results pertain to a highly developed economy, with specific cultural and economic traits; while we can assume that results would be similar in other Nordic countries – as in incomes or education (Solon, 1999; Björklund and Jäntti, 2011; Black and Devereux, 2011) –, they may not apply for the US (a country with higher inequality) or for developing countries. Last but not least, while it would be desirable to base policy on our results (for instance, investments in entrepreneurship for one generation would spill over to other generations, thereby generating a multiplier), it is also likely that the sibling correlations we obtain are also a product of policy. As Sweden is a country with noted egalitarian policies, it is perhaps less surprising to find a high degree of equality of opportunity in entrepreneurial entry and success, as well.

It would be interesting to extend the application of the model in entrepreneurship to other countries. For instance, as US earnings inequality estimates are constantly higher than those for Nordic countries (Solon, 1999; Björklund and Jäntti, 2011; Black and Devereux, 2011; Schnitzlein, 2014), equality of opportunity in entrepreneurial entry and success needs to be much higher if the US is to really be a 'land of opportunity'. As Björklund et al. (2009) have shown, sibling correlations can also be used to measure trends in mobility across successive cohorts and such an extension for entrepreneurship would be valuable.

References

Ahlburg, D. (1998). Intergenerational Transmission of Health. *American Economic Review*, 88(2):265–270.

Albanese, G., de Blasio, G., and Sestito, P. (2016). My Parents Taught Me. Evidence on the Family Transmission of Values. *Journal of Population Economics*, 29(2):571–592.

Altonji, J. G., Cattan, S., and Ware, I. (2010). Identifying Sibling Influence on Teenage Substance Use. *NBER Working Paper Series*, 16508.

Andersson, L. and Hammarstedt, M. (2010). Intergenerational Transmissions in Immigrant Self-Employment: Evidence from Three Generations. *Small Business Economics*, 34(3):261–276.

Becker, G. S. (1988). Family Economics and Macro Behavior. *American Economic Review*, 78(1):1–13.

Behrman, J. R. and Rosenzweig, M. R. (2004). Parental Allocations to Children: New Evidence on Bequest Differences among Siblings. *Journal of Political Economy*, 86(2):637–640.

Behrman, J. R. and Taubman, P. (1986). Birth Order, Schooling, and Earnings. *Journal of Labor Economics*, 4(3):S121–S145.

Bennedsen, M., Nielsen, K. M., Perez-Gonzalez, F., and Wolfenzon, D. (2007). Inside the Family Firm: The Role of Families in Succession Decisions and Performance. *Quarterly Journal of Economics*, 122(2):647–691.

Bernheim, B. D. and Severinov, S. (2003). Bequests as Signals: An Explanation for the Equal Division Puzzle. *Journal of Political Economy*, 111(4):733–764.

Björklund, A., Ginther, D. K., and Sundström, M. (2007). Family Structure and Child Outcomes in the USA and Sweden. *Journal of Population Economics*, 20(1):183–201.

Björklund, A. and Jäntti, M. (1997). Intergenerational Income Mobility in the United States. *American Economic Review*, 87(5):1009–1018.

Björklund, A. and Jäntti, M. (2011). Intergenerational Income Mobility and the Role of Family Background. In Nolan, B., Salverda, W., and Smeeding, T., editors, *The Oxford Handbook of Economic Inequality*, pages 1–21. Oxford University Press, Oxford.

Björklund, A. and Jäntti, M. (2012). How Important is Family Background for Labor-Economic Outcomes? *Labour Economics*, 19(4):465–474.

Björklund, A., Jäntti, M., and Lindquist, M. J. (2009). Family Background and Income During the Rise of the Welfare State: Brother Correlations in Income for Swedish Men Born 1932-1968. *Journal of Public Economics*, 93(5-6):671–680.

Björklund, A., Jäntti, M., and Solon, G. (2005). Influences of Nature and Nurture on Earnings Variation: A Report on a Study of Various Sibling Types in Sweden. In Bowles, S., Gintis, H., and Osborne, M., editors, *Unequal chances: Family Background and Economic Success*, pages 145–164. Russel Sage Foundation, New York.

Björklund, A., Lindahl, L., and Lindquist, M. J. (2010). What More Than Parental Income, Education and Occupation? An Exploration of What Swedish Siblings Get from Their Parents. *The B.E. Journal of Economic Analysis & Policy*, 10(1).

Black, S. E. and Devereux, P. J. (2011). Recent Developments in Intergenerational Mobility. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4B, pages 1487–1541. Elsevier B.V., Amsterdam.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2005). The More the Merrier? The Effect of Family Size and Birth Order on Children's Education. *Quarterly Journal of Economics*, 230(2):669–700.

Blanchflower, D. G. and Oswald, A. J. (1998). What Makes an Entrepreneur? *Journal of Labor Economics*, 16(1):26–60.

Bonett, D. G. (2007). Transforming Odds Ratios into Correlations for Meta-Analytic Research. *American Psychologist*, 62(3):254–255.

Booth, A. L. and Kee, H. J. (2009). Birth Order Matters: The Effect of Family Size and Birth Order on Educational Attainment. *Journal of Population Economics*, 22(2):367–397.

Conley, D. (2004). *The Pecking Order: Which Siblings Succeed and Why.* Pantheon Books, New York, NY.

Conley, D. and Glauber, R. (2005). Sibling Similarity and Difference in Socioeconomic Status: Life Course and Family Resource Effects. *NBER Working Paper Series*, 11320.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2012). The Intergenerational Transmission of Risk and Trust Attitudes. *The Review of Economic Studies*, 79(2):645–677.

Dunn, T. and Holtz-Eakin, D. (2000). Financial Capital, Human Capital, and the Transition to Self-Employment: Evidence from Intergenerational Links. *Journal of Labor Economics*, 18(2):282–305.

Edin, P. A., Fredriksson, P., and Åslund, O. (2003). Ethnic Enclaves and the Economic Success of Immigrants: Evidence from a Natural Experiment. *Quarterly Journal of Economics*, 118(1):329–357.

Eriksson, K. H., Hjalmarsson, R., Lindquist, M. J., and Sandberg, A. (2016). The Importance of Family Background and Neighborhood Effects as Determinants of Crime. *Journal of Population Economics*, 29(2):219–262.

Fairlie, R., Karlan, D., and Zinman, J. (2015). Behind the GATE Experiment: Evidence on Effects of and Rationales for Subsidized Entrepreneurship Training. *American Economic Review*, 7(2):125–161.

Fairlie, R. W. and Robb, A. M. (2007a). Families, Human Capital, and Small Business: Evidence From the Characteristics of Business Owners Survey. *Industrial and Labor Relations Review*, 60(2):225–245.

Fairlie, R. W. and Robb, A. M. (2007b). Why Are Black-Owned Businesses Less Successful than White-Owned Businesses? The Role of Families, Inheritances, and Business Human Capital. *Journal of Labor Economics*, 25(2):289–323.

Giannetti, M. and Simonov, A. (2009). Social Interactions and Entrepreneurial Activity. *Journal of Economics & Management Strategy*, 18(3):665–709.

Glaeser, E. L., Kerr, W. R., and Ponzetto, G. A. M. (2010). Clusters of Entrepreneurship. *Journal of Urban Economics*, 67(1):150–168.

Guiso, L., Pistaferri, L., and Schivardi, F. (2015). Learning Entrepreneurship from Other Entrepreneurs? *NBER Working Paper Series*, 21775.

Guo, G. and Wang, J. M. (2002). The Mixed or Multilevel Model for Behavior Genetic Analysis. *Behavior Genetics*, 32(1):37–49.

Henrekson, M. and Sanandaji, T. (2014). Small Business Activity Does Not Measure Entrepreneurship. *Proceedings of the National Academy of Sciences*, 111(5):1760–1765.

Hoffmann, A., Junge, M., and Malchow-Møller, N. (2015). Running in the Family: Parental Role Models in Entrepreneurship. *Small Business Economics*, 44(1):79–104.

Holtz-Eakin, D. (2000). Public Policy Toward Entrepreneurship. *Small Business Economics*, 15(4):283–291.

Holtz-Eakin, D., Joulfaian, D., and Rosen, H. S. (1994). Sticking it Out: Entrepreneurial Survival and Liquidity Constraints. *Journal of Political Economy*, 102(1):53–75.

Holtz-Eakin, D., Rosen, H. S., and Weathers, R. (2000). Horatio Alger Meets the Mobility Tables. *Small Business Economics*, 14(4):243–274.

Hout, M. and Rosen, H. (2000). Self-Employment, Family Background, and Race. *Journal of Human Resources*, 35(4):670–692.

Huber, L. R., Sloof, R., and Van Praag, M. (2014). The Effect of Early Entrepreneurship Education: Evidence from a Field Experiment. *European Economic Review*, 72:76–97.

Hurst, E. and Lusardi, A. (2004). Liquidity Constraints, Household Wealth, and Entrepreneurship. *Journal of Political Economy*, 112(2):319–347.

Hurst, E. G. and Pugsley, B. W. (2015). Wealth, Tastes, and Entrepreneurial Choice. *NBER Working Paper Series*, 21644.

Kacperczyk, A. J. (2013). Social Influence and Entrepreneurship: The Effect of University Peers on Entrepreneurial Entry. *Organization Science*, 24(3):664–683.

Kerr, W. R. and Mandorff, M. (2015). Social Networks, Ethnicity, and Entrepreneurship. *NBER Working Paper Series*, 21597.

Kunz, J., Page, M. E., and Solon, G. (2003). Are Point-in-Time Measures of Neighborhood Characteristics Useful Proxies for Children's Long-Run Neighborhood Environment? *Economics Letters*, 79(2):231–237.

Lelarge, C., Sraer, D., and Thesmar, D. (2010). Entrepreneurship and Credit Constraints: Evidence from a French Loan Guarantee Program. In Lerner, J. and Schoar, A., editors, *International Differences in Entrepreneurship*, pages 243–272. Chicago University Press, Chicago, IL.

Lentz, B. F. and Laband, D. N. (1990). Entrepreneurial Success and Occupational Inheritance among Proprietors. *The Canadian Journal of Economics*, 23(3):563–579.

Lerner, J. (2009). *Boulevard of Broken Dreams: Why Public Efforts to Boost Entrepreneurship and Venture Capital Have Failed – and What to Do About It.* Princeton University Press, Princeton, NJ.

Lerner, J. and Malmendier, U. (2013). With a Little Help from My (Random) Friends: Success and Failure in Post-Business School Entrepreneurship. *Review of Financial Studies*, 26(10):2411–2452.

Levine, R. and Rubinstein, Y. (2013). Smart and Illicit: Who Becomes an Entrepreneur and Does It Pay? *NBER Working Paper Series*, 19276.

Light, A. and McGarry, K. (2004). Why Parents Play Favorites: Explanations for Unequal Bequests. *American Economic Review*, 94(5):1669–1681.

Lindahl, L. (2011). A Comparison of Family and Neighborhood Effects on Grades, Test Scores, Educational Attainment and Income-Evidence from Sweden. *Journal of Economic Inequality*, 9(2):207–226.

Lindh, T. and Ohlsson, H. (1996). Self-Employment and Windfall Gains: Evidence from the Swedish Lottery. *Economic Journal*, 106(439):1515–1526.

Lindquist, M. J., Sol, J., and van Praag, C. M. (2015). Why Do Entrepreneurial Parents Have Entrepreneurial Children? *Journal of Labor Economics*, 33(2):269–296.

Lucas Jr., R. E. (1978). The Size Distribution of Business Firms. *The Bell Journal of Economics*, 9(2):508–523.

Mazumder, B. (2008). Sibling Similarities and Economic Inequality in the US. *Journal of Population Economics*, 21(3):685–701.

McGarry, K. (1999). Inter Vivos Transfers and Intended Bequests. *Journal of Public Economics*, 73(3):321–351.

Mechoulan, S. and Wolff, F.-C. (2015). Intra-Household Allocation of Family Resources and Birth Order: Evidence from France Using Siblings Data. *Journal of Population Economics*, 28(4):937–964.

Nanda, R. and Sørensen, J. B. (2010). Workplace Peers and Entrepreneurship. *Management Science*, 56(7):1116–1126.

Nicolaou, N. and Shane, S. (2010). Entrepreneurship and Occupational Choice: Genetic and Environmental Influences. *Journal of Economic Behavior & Organization*, 76(1):3–14.

Nicolaou, N., Shane, S., Cherkas, L., Hunkin, J., and Spector, T. D. (2008). Is the Tendency to Engage in Entrepreneurship Genetic? *Management Science*, 54(1):167–179.

Nicoletti, C. and Rabe, B. (2013). Inequality in Pupils' Test Scores: How Much do Family, Sibling Type and Neighbourhood Matter? *Economica*, 80(318):197–218.

Oosterbeek, H., van Praag, M., and Ijsselstein, A. (2010). The Impact of Entrepreneurship Education on Entrepreneurship Skills and Motivation. *European Economic Review*, 54(3):442–454.

Page, M. E. and Solon, G. (2003a). Correlations between Brothers and Neighboring Boys in Their Adult Earnings: The Importance of Being Urban. *Journal of Labor Economics*, 21(4):831–855.

Page, M. E. and Solon, G. (2003b). Correlations Between Sisters and Neighbouring Girls in Their Subsequent Income as Adults. *Journal of Applied Econometrics*, 18(5):545–562.

Parker, S. (2009). *The Economics of Entrepreneurship.* Cambridge University Press, Cambridge, UK.

Rabe-Hesketh, S., Skrondal, A., and Gjessing, H. K. (2008). Biometrical modeling of twin and family data using standard mixed model software. *Biometrics*, 64(1):280–288.

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2005). Maximum Likelihood Estimation of Limited and Discrete Dependent Variable Models with Nested Random Effects. *Journal of Econometrics*, 128(2):301–323.

Robinson, P. B. and Hunt, K. H. (1992). Entrepreneurship and Birth Order: Fact or Folklore. *Entrepreneurship & Regional Development*, 4(3):287–298.

Schnitzlein, D. D. (2014). How Important Is the Family? Evidence from Sibling Correlations in Permanent Earnings in the USA, Germany, and Denmark. *Journal of Population Economics*, 27(1):69–89.

Shane, S. (2009). Why Encouraging More People to Become Entrepreneurs Is Bad Public Policy. *Small Business Economics*, 33(2):141–149.

Solon, G. (1999). Intergenerational Mobility in the Labor Market. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3, pages 1761–1800. Elsevier N.V., Amsterdam.

Solon, G., Page, M. E., and Duncan, G. J. (2000). Correlations between Neighboring Children in Their Subsequent Educational Attainment. *Review of Economics and Statistics*, 82(3):383–392.

Sørensen, J. B. (2007). Closure and Exposure: Mechanisms in the Intergenerational Transmission of Self-Employment. *Research in the Sociology of Organizations: The Sociology of Entrepreneurship*, 25:83–124.

Sulloway, F. J. (1996). *Born to Rebel*. Pantheon Books, New York, NY.

Tervo, H. (2006). Regional Unemployment, Self-Employment and Family Background. *Applied Economics*, 38(9):1055–1062.

van Praag, C. M. and Versloot, P. H. (2007). What Is the Value of Entrepreneurship? A Review of Recent Research. *Small Business Economics*, 29(4):351–382.

Zhang, Z., Zyphur, M. J., Narayanan, J., Arvey, R. D., Chaturvedi, S., Avolio, B. J., Lichtenstein, P., and Larsson, G. (2009). The Genetic Basis of Entrepreneurship: Effects of Gender and Personality. *Organizational Behavior and Human Decision Processes*, 110(2):93–107.

Table 1: **Number of Families with *N* Children**

N Children	No. of Families	%	No. of Individuals	%
1	227,519	52.40	227,519	32.26
2	154,051	35.48	308,102	43.69
3	43,026	9.91	129,078	18.30
4	7,881	1.82	31,524	4.47
5	1,401	0.32	7,005	0.99
6	255	0.06	1,530	0.22
7	56	0.01	392	0.06
8	14	0.00	112	0.02
Total	434,203	100.00	705,262	100.00

All children of the same mother are defined as belonging to the same family. In panel B, which includes all siblings, the largest family contains 18 siblings.

Table 2: **Descriptive Statistics**

	Mean	S.D.	N	Min	Max
A. Entrepreneurial outcomes					
Entrepreneur	0.146	(0.353)	705,262	0	1
Incorporated	0.083	(0.276)	705,262	0	1
Entrepreneur $\geq 4y$	0.076	(0.265)	705,262	0	1
Incorporated $\geq 5y$	0.042	(0.201)	705,262	0	1
High income entrepreneur	0.073	(0.260)	705,262	0	1
Years entrepreneur [a]	5.290	(5.800)	102,921	1	28
Entrepreneurial income [a]	13.992	(1.192)	102,921	0	19.9
Years incorporated [a]	5.785	(4.604)	58,414	1	20
B. Parental characteristics					
Mother Entrepreneur	0.152	(0.359)	705.262	0	1
Father Entrepreneur	0.251	(0.434)	694,040	0	1
Mother Incorporated	0.032	(0.175)	705,262	0	1
Father Incorporated	0.064	(0.244)	694,040	0	1
Mother's log income	11.596	(0.834)	703,369	0	17.1
Father's log income	12.173	(0.667)	690,223	0	17.3
Mother's years of schooling	10.044	(2.799)	693,138	7	19
Father's years of schooling	9.985	(3.045)	671,125	7	19
Mother immigrant	0.112	(0.315)	705,262	0	1
Father immigrant	0.093	(0.291)	694,040	0	1
C. Family structure					
Male	0.513	(0.499)	705,262	0	1
Twins	0.020	(0.143)	705,262	0	1
Adopted	0.014	(0.119)	705,262	0	1
Father unknown	0.016	(0.125)	705,262	0	1
Family size, total [b]	2.801	(1.306)	434,203	1	18
Family size, in sample [b]	1.624	(0.773)	434,203	1	8
Mother's partner count [b]	1.031	(0.182)	434,203	1	5
Adoptive and biological children [b]	0.006	(0.077)	434,203	0	1
Mother's age at first birth [b]	23.243	(4.423)	434,203	13	55
Family structure at age 15 [b]					
Both parents present	69.31%				
Single mother	18.54%				
Single father	3.79%				
Mother with new husband	5.00%				
Father with new wife	1.97%				
Missing	1.39%				
D. Neighborhood characteristics					
Parish size	262.368	(472.045)	2,650	1	5,359
Ever entrepreneur	0.175	(0.380)	2,650	0	1
Ever incorporated	0.080	(0.271)	2,650	0	1

[a] Conditional on entrepreneurship and, respectively, incorporation.

[b] Variables calculated at the family level to avoid overweighting large families.

Table 3: **Baseline Sibling Correlation Estimates**

	(1) All	(2) Male	(3) Female	(4) Mixed	(5) All, excl. singles	(6) All, father
A. Extensive margin						
Entrepreneur	0.232	0.318	0.232	0.198	0.232	0.224
	(0.004)	(0.008)	(0.011)	(0.005)	(0.004)	(0.004)
Incorporated	0.341	0.399	0.347	0.306	0.341	0.331
	(0.005)	(0.008)	(0.015)	(0.007)	(0.005)	(0.005)
B. Stricter definitions						
Entrepreneur $\geq 4y$	0.262	0.365	0.245	0.219	0.262	0.252
	(0.005)	(0.010)	(0.016)	(0.007)	(0.005)	(0.005)
Incorporated $\geq 5y$	0.424	0.477	0.448	0.380	0.424	0.413
	(0.006)	(0.010)	(0.021)	(0.009)	(0.006)	(0.007)
High income entrepreneur	0.175	0.215	0.193	0.153	0.175	0.175
	(0.006)	(0.011)	(0.017)	(0.008)	(0.006)	(0.006)
Individuals	705,262	218,579	201,800	284,883	477,743	694,040
Families	434,203	163,583	153,211	117,409	206,684	424,391
C. Continuous outcomes						
Years entrepreneur	0.214	0.315	0.185	0.167	0.214	0.209
	(0.008)	(0.013)	(0.028)	(0.011)	(0.009)	(0.008)
Entrepreneurial income	0.219	0.313	0.153	0.183	0.217	0.214
	(0.008)	(0.014)	(0.027)	(0.011)	(0.009)	(0.008)
Individuals	102,091	38,396	22,692	41,833	21,766	101,307
Families	91,513	34,939	21,509	35,056	10,358	89,954
Years incorporated	0.387	0.467	0.396	0.312	0.435	0.384
	(0.009)	(0.013)	(0.032)	(0.014)	(0.011)	(0.009)
Individuals	58,414	25,933	9,244	23,237	10,096	57,823
Families	53,160	23,754	8,902	20,504	4,842	52,610

Standard errors in parentheses. Family defined by mother in columns (1)-(5), and by father in column (6). Every reported sibling correlation (and its standard error) comes from a different model, and all models include gender and birth year dummies as fixed covariates.

Table 4: **Sibling Differences**

	(1) Entrepreneur	(2) Incorporated	(3) Entrepreneur $\geq 4y$	(4) Incorporated $\geq 5y$	(5) High income entrepreneur	(6) Years entrepreneur	(7) Entrepreneurial income	(8) Years incorporated
	A. Baseline sibling correlations							
ρ	0.232	0.341	0.262	0.424	0.175	0.214	0.219	0.387
	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)	(0.008)	(0.008)	(0.009)
	B. Second step							
R^2	0.013	0.018	0.011	0.016	0.004	0.024	0.013	0.030
	C. Revised measure of family influences [a]							
$\tilde{\rho}$	0.242	0.355	0.270	0.432	0.178	0.233	0.228	0.405
	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)	(0.008)	(0.008)	(0.009)

Standard errors in parentheses.

[a] A Wald test suggests that ρ and $\tilde{\rho}$ are significantly different, but the absolute difference between them is usually less than 1.9 percent of total variance, or a maximum of 9 percent of the sibling correlation, with an average around 4 percent. The confidence intervals are largely overlapping.

[b] We obtain very similar results if we restrict the sample to i) complete families (for which the first born does not lie outside our birth window) and ii) complete families with at least two children (to ensure our results are not driven by lack of variation in the first born variable).

Table 5: **Neighborhood Correlations**

	(1) Entrepreneur	(2) Incorporated	(3) Entrepreneur $\geq 4y$	(4) Incorporated $\geq 5y$	(5) High income entrepreneur	(6) Years entrepreneur	(7) Entrepreneurial income	(8) Years incorporated
A. Baseline sibling correlations								
	0.232	0.341	0.262	0.424	0.175	0.214	0.219	0.387
	(0.004)	(0.005)	(0.005)	(0.007)	(0.006)	(0.008)	(0.008)	(0.009)
B. Neighborhood correlations (two-level)								
	0.028	0.020	0.043	0.024	0.012	0.047	0.023	0.027
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)	(0.002)
	11.98%	**5.84%**	**16.45%**	**5.65%**	**6.96%**	**21.90%**	**10.57%**	**7.02%**
C. Controlling for parental characteristics								
	0.016	0.010	0.024	0.014	0.007	0.024	0.011	0.017
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
	6.91%	**3.00%**	**9.27%**	3.21%	**4.07%**	**11.39%**	**4.81%**	**4.33%**
D. Neighborhood correlation (three-level)								
Neighborhood correlation	0.025	0.015	0.036	0.013	0.011	0.043	0.021	0.021
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.003)	(0.002)	(0.002)
Family|neighborhood correlation	0.236	0.303	0.250	0.354	0.183	0.227	0.223	0.398
	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)	(0.008)	(0.009)	(0.010)

Standard errors in parentheses. The three-level models for extensive margin outcomes are estimated using Stata's *xtmelogit* command (*melogit* fails to obtain feasible starting values – most likely because the neighborhood variances are close to zero, i.e. the parameter boundary). Numbers in bold represent the contribution of neighborhoods to the sibling correlation.

Table 6: **Accounting Exercise: Family Factors Generating Sibling Similarities**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Entrepreneur	Incorporated	Entrepreneur $\geq 4y$	Incorporated $\geq 5y$	High income entrepreneur	Years entrepreneur	Entrepreneurial income	Years incorporated
A. Sample sibling correlations								
	0.229	0.340	0.257	0.423	0.178	0.212	0.211	0.394
	(0.004)	(0.005)	(0.006)	(0.007)	(0.006)	(0.008)	(0.009)	(0.009)
B. Adding parental characteristics separately								
Education, income	0.229	0.334	0.257	0.422	0.172	0.201	0.187	0.389
	(0.004)	(0.005)	(0.006)	(0.007)	(0.006)	(0.008)	(0.009)	(0.009)
	0.11%	**1.75%**	**0.18%**	**0.23%**	**3.34%**	**4.87%**	**11.38%**	**1.47%**
Immigration	0.229	0.339	0.257	0.421	0.178	0.212	0.208	0.393
	(0.004)	(0.005)	(0.006)	(0.007)	(0.006)	(0.008)	(0.009)	(0.009)
	0.10%	**0.11%**	**0.10%**	**0.26%**	**0.00%**	**0.00%**	**1.29%**	**0.22%**
Entrepreneurship	0.199	0.334	0.219	0.416	0.168	0.194	0.199	0.393
	(0.004)	(0.005)	(0.006)	(0.007)	(0.006)	(0.008)	(0.009)	(0.009)
	12.85%	**1.83%**	**14.97%**	**1.57%**	**5.64%**	**8.26%**	**5.69%**	**0.43%**
Incorporation	0.228	0.285	0.257	0.353	0.175	0.210	0.206	0.362
	(0.004)	(0.005)	(0.006)	(0.007)	(0.006)	(0.008)	(0.009)	(0.010)
	0.24%	**16.16%**	**0.01%**	**16.38%**	**1.37%**	**0.87%**	**2.39%**	**8.07%**
Family structure	0.229	0.337	0.257	0.420	0.177	0.210	0.210	0.394
	(0.004)	(0.005)	(0.006)	(0.007)	(0.006)	(0.008)	(0.009)	(0.009)
	0.00%	**0.72%**	**0.04%**	**0.71%**	**0.12%**	**0.82%**	**0.53%**	**0.18%**
C. Adding parental characteristics jointly								
	0.198	0.271	0.217	0.345	0.159	0.182	0.168	0.355
	(0.004)	(0.005)	(0.006)	(0.007)	(0.006)	(0.008)	(0.009)	(0.010)
	13.31%	**20.39%**	**15.51%**	**18.44%**	**10.64%**	**13.77%**	**20.52%**	**9.91%**

Standard errors in parentheses. Numbers in bold represent the contribution of parental characteristics to the sibling correlation.

Table 7: **Twin Correlations for DZ and MZ Twins**

	(1) Entrepreneur	(2) Incorporated	(3) Entrepreneur $\geq 4y$	(4) Incorporated $\geq 5y$	(5) High income entrepreneur
A. Males					
ρ_{MZ}	0.507	0.637	0.639	0.597	0.487
	(0.068)	(0.064)	(0.067)	(0.100)	(0.091)
ρ_{DZ}	0.338	0.455	0.385	0.533	0.299
	(0.035)	(0.039)	(0.048)	(0.046)	(0.057)
Heritability:	**0.338**	**0.364**	**0.506**	**0.127**	**0.376**
	(0.173)	(0.170)	(0.186)	(0.252)	(0.251)
	[0.051]	[0.033]	[0.007]	[0.614]	[0.133]
B. Females					
ρ_{MZ}	0.546	0.493	0.416	0.613	0.423
	(0.065)	(0.128)	(0.133)	(0.134)	(0.125)
ρ_{DZ}	0.232	0.360	0.284	0.339	0.219
	(0.048)	(0.060)	(0.069)	(0.108)	(0.076)
Heritability:	**0.628**	**0.267**	**0.264**	**0.548**	**0.408**
	(0.185)	(0.318)	(0.345)	(0.396)	(0.341)
	[0.001]	[0.401]	[0.444]	[0.166]	[0.232]

Standard errors in parentheses, and p-values in square brackets. Twin correlations computed using auxiliary regressions involving non-twins born 12 to 24 months apart (results with different spacings are similar). Results not reported for intensive margin outcomes, given small sample sizes and extremely noisy results.

Table 8: **Peer Effects Exercise, All Sibling Types**

	A. Entrepreneurship									
	Old on young [$\phi_{t-1} = 0.076, \phi_t = 0.074$]					Young on old [$\phi_{t-1} = 0.077, \phi_t = 0.074$]				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
OR($S^{i'}_{t-1}$)	3.917[a]	2.751[a]	1.098[a]	1.100[a]		3.824[a]	2.692[a]	0.922[a]	0.907[a]	
ρ	0.230	0.166	0.014	0.014		0.230	0.165	-0.013	-0.015	
OR($S^{i'}_{t}$)				1.247[a]	4.081[a]				1.063[a]	4.081[a]
ρ				0.033	0.230				0.009	0.230
	B. Incorporation									
	Old on young [$\phi_{t-1} = 0.070, \phi_t = 0.068$]					Young on old [$\phi_{t-1} = 0.070, \phi_t = 0.068$]				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
OR($S^{i'}_{t-1}$)	7.802[a]	4.584[a]	0.936	0.968		7.942[a]	4.583[a]	1.052	1.070[b]	
ρ	0.335	0.239	-0.009	-0.005		0.335	0.237	0.007	0.010	
OR($S^{i'}_{t}$)				1.649[a]	8.412[a]				1.632[a]	8.412[a]
ρ				0.070	0.335				0.069	0.335
Family background		Yes	Yes	Yes			Yes	Yes	Yes	
Age dummies, i		Yes	Yes	Yes			Yes	Yes	Yes	
Age dummies, i'			Yes	Yes				Yes	Yes	
$S^{i'}_{t-1} + S^{i'}_{t+1}$			Yes					Yes		
$S^{i'}_{t-1} + S^{i'}_{t} + S^{i'}_{t+1}$				Yes					Yes	

[a] Effect significant at 5%. [b] Effect significant at 10%. Results are unchanged with family-clustered standard errors. In columns (1)-(5), sibling i is the younger sibling; in columns (6)-(10), sibling i is the older sibling. Odds ratios (OR) estimated using logistic regressions. Family background variables are those used in the accounting exercise: parental education, income, immigration, entrepreneurship, incorporation, and several family structure proxies. Sibling correlations in columns (1), (5), (6) and (10) are estimated using Stata's *xtlogit* command; those in columns (2)-(4) and (7)-(9) are approximated using the following formula (Bonett, 2007): $\rho \approx (OR^{\phi} + 1)(OR^{\phi} - 1)$, where ϕ is calibrated using the odds ratios and sibling correlations estimated in columns (1) and (6) for lagged effects (ϕ_{t-1}), and (5) and (10) for contemporaneous effects (ϕ_t).

Table 9: **Upper Bounds on Peer Effects Contribution to Sibling Correlations**

	Effect on younger sibling		Effect on older sibling	
	(1)	(2)	(3)	(4)
A. Entrepreneurship				
All sibling types	6.20% [a]	6.31% [a]	-5.45% [a]	-6.53% [a]
Males	8.29% [a]	9.28% [a]	-5.16% [a]	-5.38% [a]
Females	8.10%	7.32%	-4.92%	-7.87%
Mixed (younger brother)	-7.26%	-5.12%	-0.23%	-2.33%
Mixed (younger sister)	5.27%	1.82%	-12.85% [b]	-14.14% [b]
B. Incorporation				
All sibling types	-2.77%	-1.36%	2.11%	2.83% [b]
Males	-0.61%	0.86%	0.20%	1.45%
Females	-8.93% [b]	-7.12%	1.61%	1.92%
Mixed (younger brother)	-5.66%	-4.55%	7.30%	7.51%
Mixed (younger sister)	-5.95%	-4.22%	7.54%	7.11%
Contemporaneous effect	No	Yes	No	Yes

[a] Effect significant at 5%. [b] Effect significant at 10%.

The percentages represent the share of the sibling correlation explained by the lagged entrepreneurship or incorporation variable of the older sibling, columns (1) and (2), and the younger sibling, columns (3) and (4), once controls are added and correlated random effects are accounted for. For the full set of results see Table 8 and appendix Tables A.3-A.6 (the results in this table are based on columns (3), (4), (8) and (9) in those tables.

A Appendix: Additional Figures and Tables

This section contains tables with additional information.

Figure A.1 compares twin correlations with the sibling correlations of siblings spaced 10 to 120 months apart, at the extensive and the intensive margin.

Table A.1 discusses the relative contributions of parental characteristics (education, incomes, entrepreneurship, incorporation, immigration) to neighborhood correlations.

Table A.2 shows the results of the factor analysis (principal components), with the definition of six orthogonal factors used in the accounting exercise in section 5.2.

Tables A.3-A.6 show the results of Altonji-style peer effects exercises for entrepreneurship and incorporation, by each type of the sibling pair (males, females, mixed gender with younger brother or younger sister).

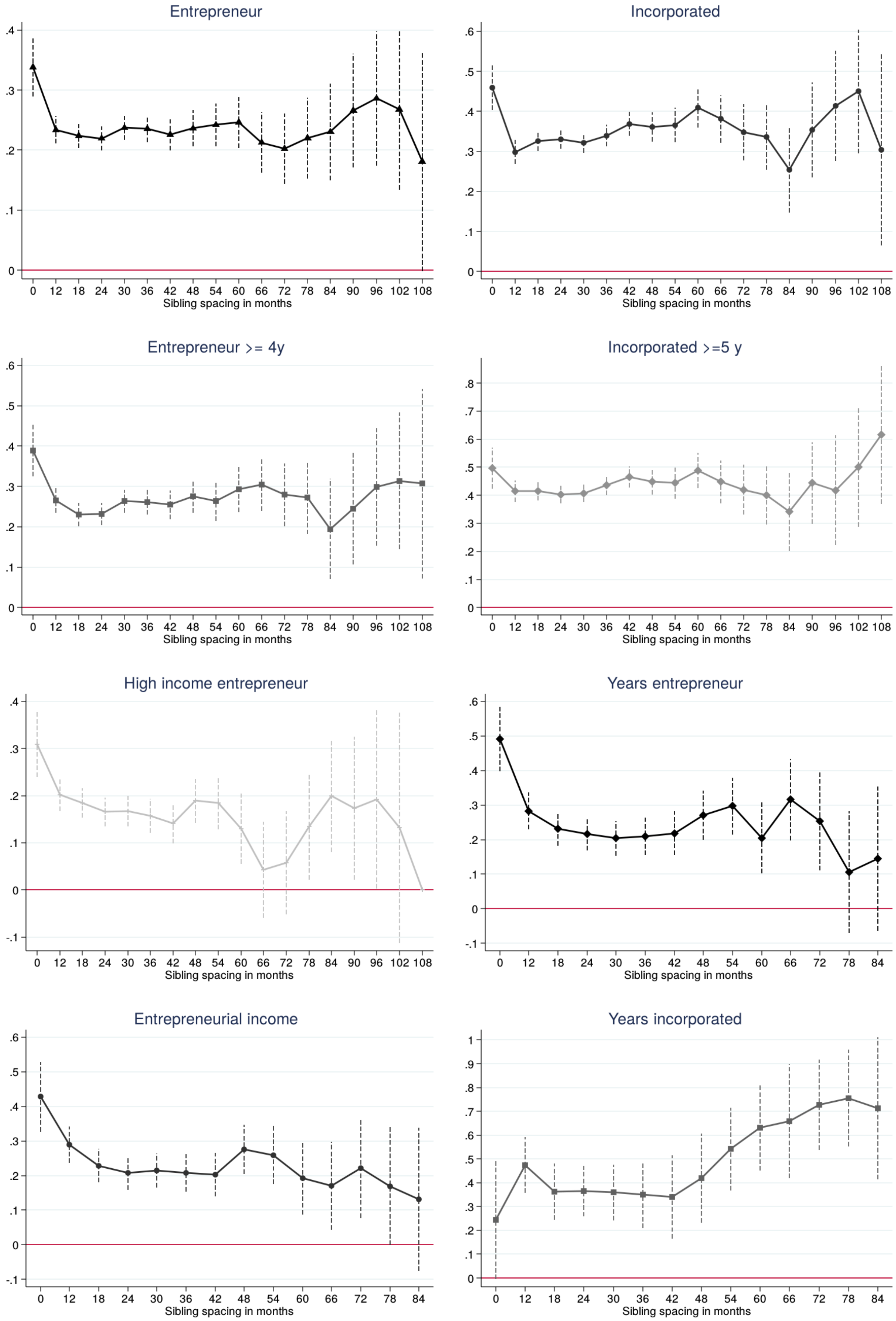

Figure A.1: Comparison of twin and sibling correlations in entrepreneurship outcomes.

Table A.1: **Neighborhood Correlations Controlling for Parental Characteristics**

	(1) Entrepreneur	(2) Incorporated	(3) Entrepreneur $\geq 4y$	(4) Incorporated $\geq 5y$	(5) High income entrepreneur	(6) Years entrepreneur	(7) Entrepreneurial income	(8) Years incorporated
A. Neighborhood correlations (two-level)								
	0.028	0.020	0.043	0.024	0.012	0.047	0.023	0.027
	(0.0014)	(0.0012)	(0.0021)	(0.0018)	(0.0009)	(0.0029)	(0.0018)	(0.0024)
B. Controlling for parental characteristics								
Parental incomes	0.027	0.016	0.041	0.024	0.011	0.041	0.015	0.027
	(0.0014)	(0.0011)	(0.0021)	(0.0018)	(0.0010)	(0.0028)	(0.0015)	(0.0025)
Parental entrep.	0.017	0.019	0.025	0.021	0.010	0.032	0.017	0.024
	(0.0010)	(0.0012)	(0.0015)	(0.0017)	(0.0008)	(0.0024)	(0.0015)	(0.0023)
Parental incorporation	0.028	0.014	0.043	0.017	0.012	0.047	0.023	0.024
	(0.0014)	(0.0010)	(0.0021)	(0.0015)	(0.0009)	(0.0029)	(0.0018)	(0.0023)
Parental education	0.028	0.018	0.043	0.023	0.010	0.041	0.018	0.021
	(0.0014)	(0.0012)	(0.0022)	(0.0018)	(0.0009)	(0.0028)	(0.0017)	(0.0022)
Parental immigration	0.028	0.020	0.044	0.023	0.012	0.047	0.024	0.026
	(0.0014)	(0.0012)	(0.0022)	(0.0017)	(0.0009)	(0.0029)	(0.0019)	(0.0024)
All of the above	0.016	0.010	0.024	0.014	0.007	0.024	0.011	0.017
	(0.0009)	(0.0009)	(0.0015)	(0.0014)	(0.0007)	(0.0022)	(0.0012)	(0.0020)

Standard errors in parentheses.

Table A.2: **Factor Analysis Results (Factor Loadings)**

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Mother's income	**0.26038**	0.05063	0.04426	0.00806	0.01269	-0.37705
Father's income	**0.28765**	-0.08183	0.14311	-0.14395	-0.00058	0.18772
Mother's education	**0.41778**	0.03411	-0.09935	0.08825	0.06447	-0.08599
Father's education	**0.41863**	0.03456	-0.07297	0.01235	0.06101	0.11373
Mother immigrant	0.01564	**0.56289**	0.01380	0.00903	-0.02951	0.00609
Father immigrant	0.03506	**0.56468**	0.01674	0.02028	0.00234	-0.00856
Mother entrepreneur	0.02035	0.01653	-0.03664	**0.59988**	0.00109	0.01479
Father entrepreneur	0.00188	0.00746	-0.02519	**0.60703**	0.00459	-0.03401
Mother incorporated	-0.06127	0.02241	**0.57101**	-0.02649	0.00231	0.00953
Father incorporated	-0.03582	0.00614	**0.57040**	-0.02879	0.01594	-0.00292
Number of children	0.03739	0.01555	-0.03853	0.00967	**0.56291**	0.29455
Mother's partner count	0.10306	-0.05109	-0.03711	0.00022	**0.40922**	-0.19288
Mother's age at first birth	-0.02038	0.00926	-0.08852	0.01675	**-0.52422**	0.16073
Adoptive and biological children	0.13065	-0.02834	-0.05932	-0.01949	0.17110	**0.34379**
Family structure	0.03447	0.02152	0.03121	0.01391	0.01615	**0.65581**
Factor definition	Education and income	Immigration	Incorporation	Entrepreneurs	Family Structure (1)	Family Structure (2)

Main factor loadings appear in bold.

Table A.3: **Peer Effects Exercise, Males**

	A. Entrepreneurship									
	Old on young $[\phi_{t-1} = 0.077, \phi_t = 0.075]$					Young on old $[\phi_{t-1} = 0.079, \phi_t = 0.075]$				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
OR($S^{i'}_{t-1}$)	5.865[a]	4.112[a]	1.181[a]	1.205[a]		5.702[a]	3.984[a]	0.901[a]	0.897[a]	
ρ	0.315	0.244	0.026	0.029		0.315	0.243	-0.016	-0.017	
OR($S^{i'}_{t}$)				1.524[a]	6.208[a]				1.172	6.208[a]
ρ				0.033	0.315				0.009	0.315
	B. Incorporation									
	Old on young $[\phi_{t-1} = 0.075, \phi_t = 0.073]$					Young on old $[\phi_{t-1} = 0.075, \phi_t = 0.073]$				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
OR($S^{i'}_{t-1}$)	9.542[a]	5.754[a]	0.984	1.023		9.646[a]	5.730[a]	1.006	1.040	
ρ	0.405	0.302	-0.002	0.003		0.405	0.300	0.001	0.006	
OR($S^{i'}_{t}$)				1.945[a]	10.394[a]				1.821[a]	10.394[a]
ρ				0.102	0.405				0.091	0.405
Family traits		Yes	Yes	Yes			Yes	Yes	Yes	
Age dummies, i		Yes	Yes	Yes			Yes	Yes	Yes	
Age dummies, i'			Yes	Yes				Yes	Yes	
$S^{i'}_{t-1} + S^{i'}_{t+1}$			Yes					Yes		
$S^{i'}_{t-1} + S^{i'}_{t} + S^{i'}_{t+1}$				Yes					Yes	

See notes to Table 8.

Table A.4: **Peer Effects Exercise, Females**

	A. Entrepreneurship									
	Old on young $[\phi_{t-1} = 0.078, \phi_t = 0.076]$					Young on old $[\phi_{t-1} = 0.080, \phi_t = 0.076]$				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
OR($S^{i'}_{t-1}$)	3.990[a]	2.906[a]	1.132	1.120		3.849[a]	2.813[a]	0.928	0.887	
ρ	0.242	0.182	0.020	0.018		0.242	0.181	-0.012	-0.019	
OR($S^{i'}_{t}$)				1.226	4.169[a]				1.098	4.169[a]
ρ				0.024	0.242				0.014	0.242
	B. Incorporation									
	Old on young $[\phi_{t-1} = 0.062, \phi_t = 0.060]$					Young on old $[\phi_{t-1} = 0.061, \phi_t = 0.060]$				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
OR($S^{i'}_{t-1}$)	11.005[a]	6.870[a]	0.776[b]	0.818		11.509[a]	7.107[a]	1.047	1.056	
ρ	0.345	0.269	-0.031	-0.025		0.345	0.268	0.006	0.007	
OR($S^{i'}_{t}$)				1.428[a]	11.990[a]				1.459[b]	11.990[a]
ρ				0.043	0.345				0.046	0.345
Family traits		Yes	Yes	Yes			Yes	Yes	Yes	
Age dummies, i		Yes	Yes	Yes			Yes	Yes	Yes	
Age dummies, i'			Yes	Yes				Yes	Yes	
$S^{i'}_{t-1} + S^{i'}_{t+1}$			Yes					Yes		
$S^{i'}_{t-1} + S^{i'}_{t} + S^{i'}_{t+1}$				Yes					Yes	

See notes to Table 8.

Table A.5: **Peer Effects Exercise, Mixed (Younger Brother)**

	A. Entrepreneurship									
	Old on young [$\phi_{t-1} = 0.084, \phi_t = 0.081$]					Young on old [$\phi_{t-1} = 0.084, \phi_t = 0.081$]				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
OR($S_{t-1}^{i'}$)	2.492[a]	1.870[a]	0.930	0.950		2.509[a]	1.893[a]	0.998	0.977	
ρ	0.167	0.111	-0.012	-0.009		0.167	0.113	0.000	-0.004	
OR($S_t^{i'}$)				1.002	2.579[a]				1.011	2.579[a]
ρ				0.000	0.167				0.002	0.167
	B. Incorporation									
	Old on young [$\phi_{t-1} = 0.069, \phi_t = 0.067$]					Young on old [$\phi_{t-1} = 0.068, \phi_t = 0.067$]				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
OR($S_{t-1}^{i'}$)	5.491[a]	3.051[a]	0.896	0.916		5.587[a]	3.121[a]	1.150	1.155	
ρ	0.264	0.166	-0.015	-0.012		0.264	0.168	0.019	0.020	
OR($S_t^{i'}$)				1.351[a]	5.792[a]				1.309[a]	5.792[a]
ρ				0.035	0.264				0.037	0.264
Family traits		Yes	Yes	Yes			Yes	Yes	Yes	
Age dummies, i		Yes	Yes	Yes			Yes	Yes	Yes	
Age dummies, i'			Yes	Yes				Yes	Yes	
$S_{t-1}^{i'} + S_{t+1}^{i'}$			Yes					Yes		
$S_{t-1}^{i'} + S_t^{i'} + S_{t+1}^{i'}$				Yes					Yes	

See notes to Table 8.

Table A.6: **Peer Effects Exercise, Mixed (Younger Sister)**

	A. Entrepreneurship									
	Old on young $[\phi_{t-1} = 0.078, \phi_t = 0.078]$					Young on old $[\phi_{t-1} = 0.082, \phi_t = 0.078]$				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$\text{OR}(S^{i'}_{t-1})$	2.674^a	2.012^a	1.057	1.020		2.572^a	1.939^a	0.876^b	0.864^b	
ρ	0.167	0.116	0.009	0.003		0.167	0.114	-0.021	-0.024	
$\text{OR}(S^{i'}_{t})$				1.012	2.703^a				0.864	2.703^a
ρ				0.002	0.167				-0.022	0.167
	B. Incorporation									
	Old on young $[\phi_{t-1} = 0.073, \phi_t = 0.070]$					Young on old $[\phi_{t-1} = 0.072, \phi_t = 0.070]$				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$\text{OR}(S^{i'}_{t-1})$	4.949^a	3.040^a	0.898	0.926		5.111^a	3.126^a	1.147	1.138	
ρ	0.264	0.177	-0.016	-0.011		0.264	0.178	0.020	0.019	
$\text{OR}(S^{i'}_{t})$				1.273^b	5.286^a				1.518^a	5.286^a
ρ				0.035	0.264				0.060	0.264
Family traits		Yes	Yes	Yes			Yes	Yes	Yes	
Age dummies, i		Yes	Yes	Yes			Yes	Yes	Yes	
Age dummies, i'			Yes	Yes				Yes	Yes	
$S^{i'}_{t-1} + S^{i'}_{t+1}$			Yes					Yes		
$S^{i'}_{t-1} + S^{i'}_{t} + S^{i'}_{t+1}$				Yes					Yes	

See notes to Table 8.