

TI 2016-077/VII
Tinbergen Institute Discussion Paper



On the Origins of Entrepreneurship: Evidence from Sibling Correlations

*Matthew J. Lindquist*¹

*Joeri Sol*²

C. Mirjam van Praag^{2,3,4}

*Theodor Vladasel*³

¹ *Stockholm University, Sweden;*

² *Faculty of Economics and Business, University of Amsterdam, the Netherlands;*

³ *Copenhagen Business School, Denmark;*

⁴ *Tinbergen Institute, the Netherlands.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

On the Origins of Entrepreneurship: Evidence from Sibling Correlations*

Matthew J. Lindquist Joeri Sol C. Mirjam van Praag
Theodor Vladasel

September 29, 2016

Abstract

Promoting entrepreneurship has become an increasingly important part of the policy agenda in many countries. The success of such policies, however, rests in part on the assumption that entrepreneurship outcomes are not fully determined at a young age by factors that are unrelated to current policy. We test this assumption and assess the importance of family background and neighborhood effects as determinants of entrepreneurship, 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 (i.e., family background and neighborhood effects). The average is 28 percent. Hence, entrepreneurship is far less than fully determined at a young age. Our estimates increase only a little when allowing for differential treatment within families by gender and birth order. We then investigate a comprehensive set of mechanisms that explain sibling similarities. Parental entrepreneurship plays a large role in explaining sibling similarities, as do shared genes. We show that neighborhood effects matter, but are rather small, particularly when compared with the overall importance of family factors. Sibling peer effects, and parental income and education matter even less.

JEL Classification: D13, J62, L26.

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

* Lindquist: SOFI, Stockholm University, matthew.lindquist@sofi.su.se; Sol: University of Amsterdam, J.Sol@uva.nl; van Praag: Copenhagen Business School, University of Amsterdam, Tinbergen Institute, CEPR and IZA,.mvp.ino@cbs.dk; Vladasel: Copenhagen Business School, tvl.ino@cbs.dk, corresponding author. We thank seminar and workshop participants at the University of Bordeaux, University Institute Lisbon, Stockholm University, Carlos III University of Madrid, Barcelona Graduate School of Economics, University of Sevilla, DRUID16 and the Third CEPR Workshop on Entrepreneurship Economics, as well as Anders Björklund, Saul Estrin and Markus Jääntti for helpful comments.

1 Introduction

Entrepreneurship has been hailed as an avenue for upward social mobility and a driver of innovation, job creation, and growth (Holtz-Eakin et al., 2000; Acemoglu et al., 2013; Decker et al., 2014); policies aimed at encouraging successful entrepreneurship have therefore been adopted in many countries (Acs et al., 2016). 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 and training curricula, from primary school (Huber et al., 2014), through tertiary education (Oosterbeek et al., 2010), and to active labor market programs (Fairlie et al., 2015).

The goals of these programs are usually to increase participants' levels of entrepreneurship relevant skills, such as business planning or creativity, or alternatively, to make them familiar with the possibility of careers in entrepreneurship. The potential success of policy measures and education programs, however, rests in part on the assumption that individuals' entrepreneurship selection and performance are not predetermined. 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 and education efforts 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 and succeeding in entrepreneurship? If not, then what family- and/or community-wide factors do children and young adults face that may later limit – or promote – their opportunities of becoming successful 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 entrepreneurship relevant skills (Huber et al., 2014) and future entrepreneurial performance (Elert et al., 2015) of relatively young pupils (in their early teens), but not of individuals in tertiary education (Oosterbeek et al., 2010) or adults (Fairlie et al., 2015). Moreover, strong intergenerational associations in entrepreneurship have attracted considerable attention; while part of this relationship has been shown to be genetic (Nicolaou et al., 2008), parental role-modeling appears to be the main driver of the intergenerational association in entrepreneurship (Sørensen, 2007; Lindquist et al., 2015). Additionally, exposure to a dense entrepreneurial environment during formative years also increases the likelihood of entry into entrepreneurship (Guiso et al., 2015). More generally, family influences are critical in human capital formation and occupational choices (Becker, 1988; Björklund and Jäntti, 2012).

Going beyond 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 intergenerational associations (Solon, 1999; Björklund and Jäntti, 2012).

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 outcomes: we recognize that self-employment and incorporation may capture different aspects of entrepreneurial engagement. Moreover, we analyze not only to what extent youth environment affects the decision to *become* an entrepreneur, but also to what extent it affects the ability to *survive* and *thrive* as an entrepreneur.

Our main results show that close to 25 percent of the variance in individuals’ decisions to become self-employed is explained by family background and community influences; for incorporation this is close to 35 percent. These percentages are slightly higher when we consider above median years of self-employment and incorporation. Brother correlations are always larger than sister correlations, with the largest correlation for males with above median years of incorporation being close to 50 percent. Mixed sibling correlations are consistently smaller than same-sex correlations.

Sibling correlations may be viewed as lower bounds on the importance of family background, since families may also generate differences between siblings if parents treat children differently based on, for example, birth order and/or gender (see also Conley, 2004; Björklund and Jäntti, 2012).¹ As differential treatment by gender or birth order may impact 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.

¹ 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.

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) sibling peer effects, and (iv) genes.

To investigate the role of neighborhood factors we estimate neighborhood correlations in entrepreneurship (see, e.g., [Solon et al., 2000](#)). Less than 10 percent of sibling correlations or 2.5 percent of total variance can be explained by neighborhood effects. 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 the scope for such effects in our setting is rather limited.

To assess the role of parental characteristics in generating sibling similarities we run an accounting exercise, where we re-estimate our sibling correlations controlling for observable parental characteristics ([Mazumder, 2008](#); [Björklund et al., 2010](#)). Our results show that parental entrepreneurship status is 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. Interestingly, parental self-employment is a prime explanatory force in individual self-employment, but not incorporation, and parental incorporation explains best individual incorporation, but not self-employment. This suggests that self-employment and incorporation are different aspects of entrepreneurship (in line with [Levine and Rubinstein, 2013](#); [Henrekson and Sanandaji, 2014](#); [van Praag and Raknerud, 2014](#); [Åstebro and Tåg, 2015](#); [Guzman and Stern, 2015](#)), and that there are different transmission mechanisms depending on the type of entrepreneurial engagement of the parents. 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.

We examine how correlations vary with the age difference between siblings to get a first indication of potential peer effects. Unlike the bulk of the sibling correlation literature, where closely spaced siblings tend to have more similar outcomes than widely spaced siblings (see, e.g., [Eriksson et al., 2016](#)), we find that sibling correlations are unaffected by birth spacing. We then compute an upper bound on the potential contribution of sibling peer effects to our sibling correlations by performing a correlated random effects exercise ([Altonji et al., 2016](#)). Our estimated peer effects are generally non-significant and very small in magnitude – less than 10 percent of the sibling correlation could potentially be explained by sibling peer effects.

Lastly, to investigate the role of genes (in general) and shared genes (in particular), 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. \(2008\)](#). The genes that non-

twin siblings share may account for a large part of sibling correlation in entrepreneurship; on average, between 56 and 78 percent of the sibling correlations in self-employment may be due to shared genes, and between 38 and 46 percent for incorporation, although not all our estimates of heritability are significant.

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 methods generate a novel measure of the extent to which entrepreneurial entry and success are predetermined by genetic and domestic factors when people are young. Our approach allows us to investigate a quite comprehensive set of potential mechanisms in order to ascertain the channels through which family environment influences children’s entrepreneurial choices and success. In doing so, we put many previous results in the literature – regarding the role of neighborhoods, parental income, parental entrepreneurship, and genes – into perspective. Unlike previous studies, we are able to investigate the relative importance of various mechanisms within a unified empirical framework. We conclude that parental entrepreneurship and genes are the two most important factors generating sibling similarities in entrepreneurship.

The remainder of this paper is structured as follows. Section 2 describes the data 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) sibling peer effects, and (iv) genes. 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 of their brothers and sisters, where siblings are defined as those having the same biological or adoptive mother. 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. 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.²

Consistent with the Swedish tax authorities, we define individuals as self-employed when they derive the majority of their taxable labor income 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, 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).

We characterize various types of entrepreneurs.^{3,4,5} Our first measure of entrepreneurship, *Self-employed*, is a dichotomous 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. In any given year, no individual is classified as both self-employed and incorporated.⁶ We add two stricter definitions of entrepreneurship by using the median of the number of years individuals spend in self-employment (incorporation) as a threshold in labeling an individual as an self-employed (incorporated). In our sample, *Self-employed* $\geq 4y$ takes a value of one for individuals who have been self-employed for at least of 4 years, and *Incorporated* $\geq 5y$ equals one for individuals incorporated for a minimum of 5 years.⁷ We define a *High income self-employed* as an individual whose self-employment income (defined below) is above the median.

² We impose these sample restrictions such that we are able to observe the occupational choices of the individuals in all cohorts and their parents for a number of years. In Appendix Table C.1, column (5), we show that our estimated sibling correlations do not change if we focus only on the smaller sample of complete families.

³ 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.

⁴ Farmers are included in Statistics Sweden's definition of business owners, since farms are typically run as companies (either incorporated or unincorporated). 16 percent of our sample of self-employed are farmers, while 4 percent of the incorporated are farmers. Sibling correlations are the same for siblings with parents who were farmers and for siblings whose parents were not farmers.

⁵ 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.

⁶ Individuals who are *Incorporated* are not a subsample of the *Self-employed*. In any given year, these two variables are mutually exclusive. Some individuals have been incorporated, but not self-employed, and vice versa, and other individuals have experienced both types of entrepreneurship at some point in their careers.

⁷ Note that our measure of duration could, in principle, consist of separate spells.

We also create several continuous measures of entrepreneurship. We start with the number of years individuals have been self-employed or incorporated: *Years self-employed* and *Years incorporated*. These variables represent established measures of entrepreneurial duration used in the literature. We also study income from self-employment 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 ‘self-employment income’. We then take the average across all years of self-employment income and then take the log of this average. This is our measure of *Self-employment income*.

We have also created a set of family-wide background variables to use in our accounting exercise. We define parental self-employment 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 old minimum, seven-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 (including those born before 1960 that do not appear in our sample of siblings). Similarly, we 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 at age 15, possibly varying 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

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

⁹ We thus estimate parish correlations, but tighter definitions of neighborhood (schools or statistical metropolitan areas) are unlikely to induce large changes in these correlations (Lindahl, 2011).

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. 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.

Table 2 presents the descriptive statistics of our outcomes of interest in panel A: 14.6 percent of the individuals in our sample are self-employed at some point, and 8.3 percent are incorporated at least once, with the average number of years spent in self-employment and incorporation 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 self-employed 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; fathers are slightly more likely to be Swedish natives.

Panel C of Table 2 shows the father is missing for 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 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 family type (18.54 percent), followed by mothers with a new husband (5 percent), single fathers (3.79 percent), and fathers with a new wife (1.97 percent). This variable is missing for 1.39 percent of our sample.

Panel D shows that our average parish, out of a total of 2,650 parishes, comprises 262 individuals, while the largest includes 5,359 individuals.¹⁰ On average, 17.5 percent of individuals in a parish become self-employed at some point, while 8 percent found an incorporated business.

In Table 3, we examine differences in the observable characteristics of employees, self-employed and incorporated. Employees are defined as those labor market participants who have never been self-employed or incorporated. The incorporated have (on average) higher

¹⁰ These numbers reflect the size of our sample in each parish, not true parish size. A Swedish parish is roughly similar in terms of size to a Census tract in the United States, with a median parish size of just under 3,000 inhabitants (in the year 2000).

incomes and more education than the other two groups, while the self-employed have (on average) lower incomes and less education. The parents of incorporated entrepreneurs also have higher educations and more income (on average) than the parents of the other two groups. Furthermore, we see that the self-employed are more likely to have parents who were self-employed (but not incorporated) than those in the other two groups. Similarly, we see that incorporated entrepreneurs are more than three times as likely to have parents who were incorporated compared with the other two categories. Clearly, incorporated and self-employed are different types of entrepreneurs in terms of their observable characteristics and in terms of their family backgrounds, including the type of entrepreneurial experiences they were exposed to as children. These differences are very much in line with those noted by [Levine and Rubinstein \(2013\)](#); also, the differences between employees and the incorporated are larger than those between employees and the self-employed, implying that employees and the self-employed are more substitutable than employees and the incorporated. This is expected to translate into larger sibling correlations in incorporation than in self-employment.

3 Sibling Correlations

Entrepreneurship, E_{if} , for sibling i from family f can be modeled as:

$$E_{if} = X'_{if}\beta + \epsilon_{if}, \quad (1)$$

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

$$\epsilon_{if} = a_f + b_{if}. \quad (2)$$

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 ϵ_{if} can be expressed as the sum of the stationary population variances of the permanent family and individual components:

$$\sigma_\epsilon^2 = \sigma_a^2 + \sigma_b^2. \quad (3)$$

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 entrepreneurship over wage employment) that can be attributed to family background effects is:

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

This share coincides with the correlation in entrepreneurship 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 of factors 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.¹¹ We return to this argument when we discuss sibling differences in Section 4.

An estimate of the sibling correlation in long-run entrepreneurship, 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:

$$E_{if} = \mathbf{X}'_{if}\beta + a_f + b_{if}. \quad (5)$$

When the outcome variable is continuous (e.g. self-employment 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 con-

¹¹ Björklund and Jäntti (2012) discuss this issue 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.

stant variance. The variance components are estimated using restricted maximum likelihood. When the outcome variable is dichotomous (e.g. *Self-employed* or *Incorporated*), we reformulate equation (5) as a latent linear response model:

$$E_{if}^* = \mathbf{X}'_{if}\beta + a_f + b_{if}, \quad (6)$$

where we only observe $E_{if} = I(E_{if}^* > 0)$. 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 $\pi^2/3$.

Results Sibling correlations are reported in Table 4. In column (1), we see that family background and community influences account for 23 percent of the likelihood of ever becoming *Self-employed* and 34 percent of the likelihood of ever becoming *Incorporated*. Looking at stricter measures of entrepreneurship, we see that 26 percent of the variation in *Self-employed* $\geq 4y$ and 42 percent of the variation in *Incorporated* $\geq 5y$ can be attributed to family-wide influences that siblings share. Becoming a *High income self-employed*, however, appears to be less influenced by family background. Columns (2)-(4) also show 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.¹²

Overall, these sibling correlations imply that family background is quite important, especially for explaining the likelihood of becoming an entrepreneur according to stricter measures of entrepreneurship, such as owning an incorporated firm for a longer period of time. They are of similar magnitudes as the sibling correlations for earnings and education in Sweden (see, e.g., Björklund and Jäntti, 2012). Yet, one cannot conclude that entrepreneurship is mostly determined when young, since the majority of the variance in our outcome variables can be attributed to factors that siblings do not share.

Sensitivity Analyses Sensitivity analyses are reported in Table C.1 in Appendix C. As a first robustness check, 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 small margin and the standard errors are virtually identical with or without singletons. Additionally,

¹² The higher sibling correlation we observe for males in Table 4 may be due to heightened interaction between brothers, see Section 5.3.

we define siblings as having the same father: the estimated sibling correlations are very close to the ones for siblings having the same mother. The pattern of higher sibling correlations for males remains across these alternative specifications. Sibling correlations for two-child families, closely spaced siblings, and complete families reveal a consistent picture. A placebo test for ‘fake’ families yields insignificant ‘sibling’ correlations.¹³

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.

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 the 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

¹³ These mixed-effects models do not have closed-form solutions and rely on numerical optimization techniques. In additional sensitivity analyses (not tabulated for brevity), we experiment with 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. 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.

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

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’ (Suloway, 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 (Suloway, 1996). In any case, families create differences between siblings that may have implications for entrepreneurship: we thus expect our measure of family background to increase once birth order effects are accounted for.¹⁵

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 born and parental entrepreneurship may thus have an impact on the sibling correlations; its magnitude, however, is fairly limited, as Lindquist et al. (2015, footnote 2) show that less than 10 percent of firms in Sweden are inherited.

In order to assess the potential effects of birth order and gender differences, we employ a two-step procedure, following Björklund and Jääntti (2012).¹⁶ We estimate equations (5) and (6), and then predict the individual-level residuals. We subsequently regress these residuals

¹⁵ Early research by De Wit and Van Winden (1989) and Robinson and Hunt (1992) finds no birth order effects in entrepreneurship.

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

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, $\tilde{\rho}$:

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

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

Results Estimates in Table 5 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 family background if we (for the moment) disregard the potential role played by the non-shared genes of siblings (more on this below). It appears that differential treatment of siblings by parents according to either gender or birth order is small or hardly influential in entrepreneurship.

5 Accounting for Sibling Similarities

What is it that makes the outcomes of siblings so similar? In this section, we investigate the extent to which our sibling correlation can be accounted for by (i) neighborhoods, (ii) parental characteristics, (iii) sibling peer effects, and (iv) genes. While this accounting exercise does not allow for a causal interpretation of the determinants of entrepreneurship, it provides clues about what is potentially important in explaining sibling similarities in entrepreneurship.

5.1 Neighborhoods

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 Si-](#)

monov (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 in the local labor market is associated with about 25 percent more entry into entrepreneurship. 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) also 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 when young aids learning.

As discussed above, the sibling correlations in Table 4 include these kinds of neighborhood and community effects. What share of the sibling correlation can be accounted for by community influences that are experienced outside of the home, but still shared by siblings? To answer this question, we estimate neighborhood correlations in entrepreneurship, using data on the parish the individual resided in at age 15. We then 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:

$$E_{in} = \mathbf{X}'_{in}\beta + c_n + b_{in}, \quad (8)$$

where we include our full set of parental characteristics in \mathbf{X}'_{in} in order to correct for parental sorting into neighborhoods (Solon et al., 2000). The neighborhood correlation then becomes:

$$\rho_n = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_b^2}. \quad (9)$$

Results Panel B in Table 6 shows the results for extensive and intensive outcome margins: the neighborhood correlation ranges between 0.007 and 0.024, meaning that shared neighborhood characteristics can explain at most 2.4 percent of entrepreneurial variance. These neighborhood correlations are small relative to our baseline sibling correlations (reproduced in panel A of Table 6), with an explanatory power of usually less than 5 percent of the sibling correlations in incorporation, but up to 11 percent of the sibling correlations in self-employment. Overall, we provide evidence that the scope for neighborhood effects in entrepreneurship in general, and incorporation in particular, is rather limited, usually less than 11 percent. This result is in line with previous literature using income, education or crime as outcomes: neighborhood effects 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). Our results also help put

into perspective the previous results of [Giannetti and Simonov \(2009\)](#) and [Guiso et al. \(2015\)](#). Although we do find some effects of community, most learning and role-modeling appear to take place closer to home (within the family). Thus entrepreneurship policy at the community level may have only a small role in fostering entrepreneurship among youths.

5.2 Parental Characteristics

Which parental characteristics 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 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).¹⁷ 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.

The degree to which any particular control variable lowers the sibling correlation after being included in the fixed part of the mixed-effects model provides a metric for judging its importance in explaining sibling similarities ([Mazumder, 2008](#); [Björklund et al., 2010](#)), but does not allow for a causal interpretation. Specifically, we explore the potential roles played by: (i) parental education and income, (ii) parental self-employment 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 self-employment and incorporation are likely to influence the occupational preferences of individuals, not only through the acquisition

¹⁷ 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.

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; Colombier and Masclet, 2008; 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 (De Wit and Van Winden, 1989; Dunn and Holtz-Eakin, 2000; Hout and Rosen, 2000; Hundley, 2006; Tervo, 2006). Controlling for these observables one by one and then jointly, we can assess both their relative and their total contribution to entrepreneurial variance.

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, self-employment 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,¹⁸ generating six orthogonal factors that load onto (i) parental education and income, (ii) parental immigration, (iii) parental self-employment, (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 7. While these sibling correlations are significantly different in a statistical sense from the baseline sibling correlations (given our large sample), they have the same order of magnitude as before.

¹⁸ The results of this analysis, i.e. factor loadings, are given in Table C.2 in Appendix C.

Panel B of Table 7 shows that parental education and income explain less than 5 percent of entrepreneurial outcomes, with the exception of *Self-employment 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).¹⁹

Having non-native parents does not have a large separate impact on entrepreneurship outcomes, usually less than 1 percent, at first glance in contrast to Andersson and Hammarstedt (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 *Self-employed*, *Self-employed $\geq 4y$* , and *Years self-employed*, with the factor explaining as much as 15 percent of sibling similarities. Conversely, parental incorporation has a very small impact on self-employment outcomes, but is a strong predictor of incorporation outcomes, explaining as much as 16.4 percent of variation.²⁰ The slightly larger effects found for incorporation than for self-employment 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 7, 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.²¹ 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.

¹⁹ To capture non-linearities in the contribution of parental wealth, we also experimented with a dummy for the family being in the top 5 percent of capital incomes (capital income is the difference between earnings and income). This results in a separate factor, which only explains about 2 percent of the sibling correlation in *Ever incorporated* and *High income self-employed*, and 4 percent in *Self-employment income*. The total explanatory power of the 7 factors is, however, unchanged from that reported in Table 7. In addition, using dummies for quintiles of parental socio-economic status (instead of the continuous factor) adds little explanatory power.

²⁰ The same pattern is observed when the sample is split by sibship gender, see Table C.3 in Appendix C.

²¹ Controlling for all the interactions of the six factors only adds about 1.5 percent more explanatory power.

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 have the same order of magnitude. In this case, observable family characteristics explain at most 8 percent of the *total* variation in entrepreneurial outcomes.

To sum up: parental education and income, family structure, and immigrant status account for very little of the sibling correlations in entrepreneurship; parental self-employment, however, explains a large share of the sibling correlations in self-employment (but not incorporation); parental incorporation explains a large share of the sibling correlations in incorporation (but not self-employment).

5.3 Sibling Peer Effects

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. In the entrepreneurship literature, peer effects have been convincingly 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). Here, we assess the potential role of sibling peer effects in generating sibling correlations.

We first examine sibling correlations at different birth spacings based on month of birth data, from twins (zero spacing), through siblings born at least 12 months apart in rolling intervals of 12 months, and up to sibling spacings of 120 months.²² 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. Also, closely spaced siblings may share a more similar family environment while growing up. 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. Thus, sibling correlations may increase or decrease with sibling spacing.

²² 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. In addition, we restrict the non-twins to full siblings in families with two children in our sample. Sibling correlations for this sample are the same as the baseline sibling correlations. Compare column (1) in Table 4 to column (3) in Appendix Table C.1.

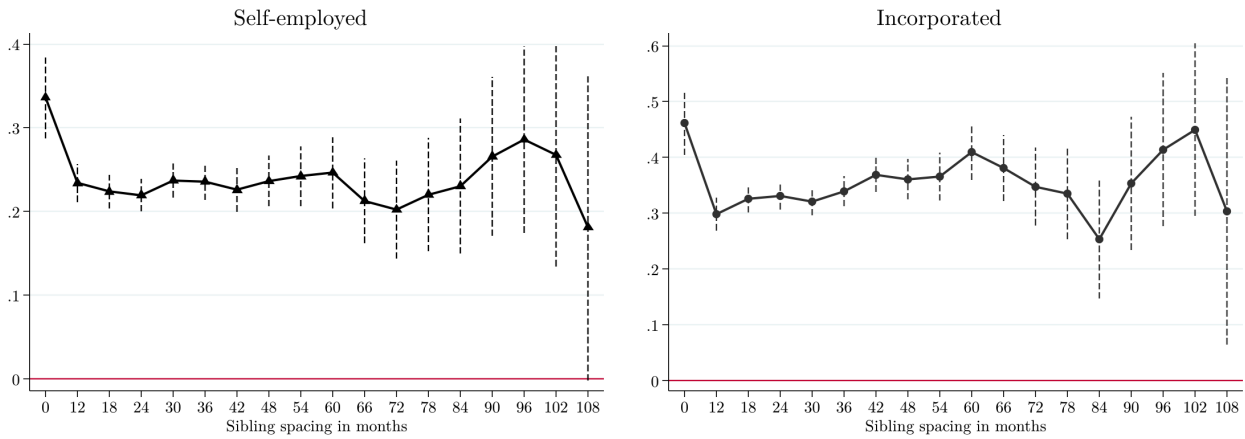


Figure 1: Twin and sibling correlations in self-employment and incorporation by sibling spacing.

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 C.1 in Appendix C). 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). However, entrepreneurship may often materialize when individuals have long ago left the household. Many of the outcomes studied earlier, such as education and crime, are realized at an earlier stage. This may explain the smaller role of sibling spacing in our case.

What we can take away from this exercise is the following: (i) time-varying, family-wide 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 5.4, and turn to a second peer effects exercise next.

While we lack a formal randomization process, by exploiting differences in the timing of entrepreneurial entry for sibling pairs, we may gain information about spillovers from one sibling to the other. A useful method for exploring such peer effects has been proposed by Altonji et al. (2016), 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 assumptions that only older siblings can influence younger siblings and that parental influences are not a mediating channel. 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.²³

Since our exercise focuses on explaining the variance of entrepreneurial outcomes due to the influence of sibling peers rather than on identifying causal effects, we take an agnostic approach to applying the [Altonji et al. \(2016\)](#) model. We estimate both the effect 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 ([Bonett, 2007](#)). A more detailed description of our empirical model is given in Appendix A.

Results Table 8 summarizes the results of our sibling peer effects exercise on the subsample of sibling pairs, with panel A referring to self-employment and panel B to incorporation. Column (1) shows how much the impact of the older sibling’s entrepreneurship status at time $t-1$ on the younger sibling’s entrepreneurship status at time t contributes (at most) to the baseline sibling correlation, and column (2) does so while controlling for a contemporaneous effect. Columns (3) and (4) do the same for the impact of the younger sibling on the older sibling.

The lagged effect of the older sibling’s self-employment on the younger sibling represents at most 6.20 to 6.31 percent of the baseline sibling correlation. Conversely, the effect of the younger sibling on the older one in self-employment appears largely negative; this implies that peer effects may actually generate sibling dissimilarity. When we disaggregate by type of sibling pair, we note that the direct peer effects for the subsample are driven by peer influences between brothers, which reflect the same pattern, with very similar magnitudes. With regards to other sibling types and incorporation, most estimated peer effects are not significant.²⁴

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 contribute less than 10 percent to our sibling correlations. Thus, peer effects are too small in magnitude to drive the sibling correlation, and at times may even create sibling dissimilarities.

5.4 Genes

As noted by [Björklund et al. \(2005\)](#) and [Conley and Glauber \(2005\)](#), sibling correlations also capture shared genetic endowment. Despite disagreement on the distribution of nature and

²³ 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.

²⁴ An 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 significant at 10 percent. However, applying a Bonferroni correction for testing multiple hypotheses would render all estimates insignificant.

nurture, the literature has established a definite role for genetic endowment in entrepreneurship (Nicolaou et al., 2008; Lindquist et al., 2015). In a study of Swedish adoptees, Lindquist et al. (2015) find that the intergenerational association in entrepreneurship is driven by nurture (roughly two thirds), rather than nature (roughly one third). By contrast, in their Swedish twins study, 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.

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 self-employment) 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, the results from the twins literature only speak to sharing an *entire* genome – it is then unclear how to interpret the results at the level of the entire population. Indeed, for two non-twin full siblings who share (on average) 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 exact relationship between genetic effects and the sibling correlation itself. We can, however, place an upper bound on the contribution of shared genes to the sibling correlation.

To do this, we first need to calculate sibling correlations for mono- (MZ) and di-zygotic (DZ) twins. We can identify all twins in our data (via month of birth). But, unfortunately, we have no indicator of their zygosity (other than knowing that non same-sex twins are di-zygotic). In the absence of knowledge about which pairs of twins are mono- or di-zygotic, 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).²⁵ In order to estimate the contribution of genes to sibling correlations, we assume that: (i) gender differences for DZ twins can be approximated by gender differences in closely spaced non-twin sibling pairs, (ii) the variance of the family component for all same sex twins is a weighted average of corresponding variances for MZ and DZ twin pairs, and (iii) boys and girls are conceived with almost equal probabilities.

One important caveat we share with most other twin studies 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

²⁵ This approach has also been used by Björklund and Jäntti (2012). See Appendix B for a more detailed exposition of our empirical model.

in their estimation of genetic influences (in their data, interactions are 50 percent stronger for identical twins). Unfortunately, we do not have direct information on sibling interaction. However, we did not find any evidence for strong sibling peer effects in the previous section.

Results The results in Table 9 show that, overall, genes can potentially explain between 26 percent and 63 percent of total variance in entrepreneurship.²⁶ While we can acknowledge an impact of genes on total entrepreneurial variation that is well in line with those found in the literature, most of our heritability estimates are insignificant (i.e. we cannot actually reject zero genetic effects), especially for stricter definitions of entrepreneurship. Additionally, there is no clear pattern for heritability estimates by gender.²⁷ For instance, female heritability in being *Self-employed* is 0.628 (with a p -value of 0.001), but only 0.264 (with a p -value of 0.444) for *Self-employed* $\geq 4y$; while male heritability in being *Self-employed* $\geq 4y$ is 0.506 (with a p -value of 0.007), it is only 0.127 (with a p -value of 0.614) for being *Incorporated* $\geq 5y$.

We can use the MZ-DZ correlations reported in Table 9, together with our previous results that both sibling spacing and sibling peer effects were unimportant, to calculate the maximum share of the sibling correlation that could potentially be due to shared genes. We calculate this share as $100 * (\rho_{MZ} - \rho_{DZ}) / \rho_{DZ}$ and report these percentages in Table 9 for men and women separately. Shared genes contribute at most 40 percent to the sibling correlation in *Incorporated*, while they can explain up to half of the sibling correlation in *Self-employed* for males and potentially all of the sibling correlation in *Self-employed* for females. At most 66 percent of the sibling correlation in *Self-employed* $\geq 4y$ for men and 47 percent for women can be explained by shared genes. Lastly, 12 percent and 81 percent of the sibling correlation in *Incorporated* $\geq 5y$ can be explained for men and women, respectively.

We conclude, perhaps not unsurprisingly, that shared genes likely play a large role in generating sibling similarities. But, of course, this also implies that non-shared genes and variations in gene patterns also generate sibling differences. Björklund and Jäntti (2012) argue that such genetic differences should be added on top of the sibling correlation when discussing the importance of family background for determining adult outcomes.

²⁶ We use twins and non-twin pairs where the siblings are born 12-24 months apart. 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 our results 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.

²⁷ The higher heritability of self-employment compared to incorporation could perhaps be driven by an innate ‘taste for entrepreneurship’. The higher heritability of females in self-employment can be taken to suggest a stronger reliance on genes in a less favorable environment, in line with Zhang et al. (2009), but not with Zunino (2016).

6 Conclusion

In this paper, we have quantified the importance of family background and neighborhood effects as determinants of entrepreneurship by estimating sibling correlations in entrepreneurial outcomes. We also explored the extent to which families make siblings different. We then presented a series of exercises designed to help us determine the extent to which these correlations could be explained by (i) neighborhood effects, (ii) parental characteristics, (iii) sibling peer effects, and/or (iv) shared genes. The empirical results are summarized in Table 10.

Sibling correlations tell us that 18 to 26 percent of the variance in self-employment is due to family background and neighborhood effects. These same factors explain 34 to 42 percent of the variance of our variables concerning owner/managers of incorporated businesses. For males this number is almost 50 percent. These sibling correlations are similar to those obtained for incomes and earnings in Sweden (Björklund and Jääntti, 1997, 2012; Björklund et al., 2009, 2010). They are somewhat smaller (on average) than those found for height, cognitive and non-cognitive ability, grades or years of schooling, but sibling correlations in incorporation do show values in the high range of previous estimates for other outcomes, particularly for men.

While sibling correlations focus attention on sibling similarities, we have also considered that families can act in ways that make siblings different, e.g. through differential treatment of first born children or sons versus daughters. In our sample of siblings born between 1960 and 1970, we do not find any strong evidence of differential treatment (for example, towards the first born sons of entrepreneurs). Adding this kind of differential treatment effect on top of our sibling correlations would only increase our measure of the importance of family background by about 4 percent. The main sibling differences that can be traced back to the family are likely those differences due to the role of non-shared family genes.

Given the large sibling correlations in entrepreneurship we have uncovered, the key question for understanding the origins of entrepreneurship focuses on explaining the determinants of these important sibling similarities. Neighborhood effects can account for (at most) 10 to 12 percent of the sibling correlations. Thus, we do find positive effects of entrepreneurial neighborhoods on entrepreneurship, either in learning or social prestige (Giannetti and Simonov, 2009; Guiso et al., 2015), but their total impact on entrepreneurial entry and attainment is likely to be rather small.

Parental self-employment and incorporation can account for (at most) 15 to 16 percent of the sibling correlations. Interestingly, parental self-employment is a prime explanatory force in individual self-employment, but not incorporation, and parental incorporation explains best in-

dividual incorporation, but not self-employment. This suggests that individual self-employment and incorporation are different aspects of entrepreneurship, in line with [Levine and Rubinstein \(2013\)](#), [Henrekson and Sanandaji \(2014\)](#), [van Praag and Raknerud \(2014\)](#), [Åstebro and Tåg \(2015\)](#), and [Guzman and Stern \(2015\)](#), and that there are different transmission mechanisms contingent on the type of parental entrepreneurial engagement. Parental education and incomes explain a further 5 percent of the sibling correlation. Family structure and immigrant status explain almost none of the similarities in sibling outcomes.

We also investigated the possibility that sibling peer effects might be driving the sibling correlation, and found only weak (non-causal) evidence that brothers influence each other's entrepreneurial choices. Sibling peer effects are generally limited in magnitude and rarely statistically significant. 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 sibling similarities.

Since siblings share not only the family environment, but also part of their genes, our fourth exercise was designed to measure the potential importance of such genetic effects. We cannot reject the existence of genetic effects; while heritability explains up to 60 percent of total variation for some of the outcomes, it is often not significant. At the extremes, heritability could potentially explain all or none of the sibling correlation. On average, however, it appears that between 56 and 78 percent of the sibling correlations in self employment can (at most) be due to the genes that siblings share, while for incorporated these numbers are 38 to 46 percent.

Loosely speaking, summing up the results of these various accounting exercises allows us to explain nearly all of the sibling correlations in entrepreneurship. Perhaps more importantly, we are able to compare the relative importance of different factors in explaining sibling similarities. We do this within a single unified framework that allows us to put some perspective on the relative importance of different effects reported in the existing literature. We conclude that shared genes are likely the most important factor, followed by parental entrepreneurship, and then neighborhood effects. Parental income, education and immigrant status account for a surprisingly small share of the sibling correlations. The same holds true for family structure.

There may, of course, be factors other than those we address here that contribute to sibling similarities. These may include, but are not limited to, parents' managerial ability ([Lucas Jr., 1978](#)), risk and time preferences ([Dohmen et al., 2012](#); [Björklund et al., 2010](#)), a wider set of family values ([Albanese et al., 2016](#)), and even latent health ([Ahlburg, 1998](#)). Capturing such variation would be an interesting avenue for future research, although parts of these effects are arguably captured through the various observable parental characteristics we account for (e.g.

parental risk preferences may determine parental entrepreneurship) and may have a genetic component as well. In addition, a future reconciliation of heritability and sibling correlations could shed more light on the importance of genes in generating sibling similarity. To do this, researchers need to identify a set of specific entrepreneurial genes and gene patterns and then study the extent to which these specific genes and gene patterns are actually shared by siblings.

There are, of course, some limitations associated with the exercises presented in this paper. First and foremost, when ‘explaining’ the determinants of sibling similarities, we cannot claim that we have presented a set of precise causal estimates. Instead, we view our results as part of an exploratory accounting exercise that can point us towards those factors which can potentially explain the largest share of sibling similarities. Second, since we measure the degree to which siblings are similar, we cannot exclude the possibility that single-child families operate in a different manner and that lone children are influenced in different ways by family and community-wide factors. Third, our results pertain to a highly developed economy, with specific cultural and economic traits, and notably egalitarian policies. Our results may likely hold in the other Nordic countries, since we observe similar sibling correlations in other outcomes such as income and education across these countries (Solon, 1999; Björklund and Jäntti, 2011; Black and Devereux, 2011), but they may not apply for Southern Europe, the US 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 themselves (in part) the product of a long history of various policies (tax, education, business, etc.). It would be interesting to track changes both over time (Björklund et al., 2009), as well as across countries (Schnitzlein, 2014), in sibling correlations in entrepreneurship. Among other things, this would help us to decide whether the sibling correlations that we have documented are ‘high’ or ‘low’, and whether they hold across time and space.

We tend to view our findings optimistically. We do not believe that the existence of substantial, pre-determined family-wide factors means that policy is doomed to fail. A large share of the variation in entrepreneurship is, in fact, individual-specific and not solely determined by genes. Furthermore, children appear to be able to ‘learn’ about entrepreneurship through their family and community environment, which implies that it may be possible to ‘teach’ entrepreneurship to young people. Policy may even generate a social multiplier effect if the behavior of a successfully treated person also affects the behavior of her untreated family members.

References

- Acemoglu, D., Akcigit, U., Bloom, N., and Kerr, W. (2013). Innovation, Reallocation and Growth. *NBER Working Paper*, 18993.
- Acs, Z., Åstebro, T., Audretsch, D., and Robinson, D. T. (2016). Public Policy in Entrepreneurship: A Call to Arms. *Small Business Economics*, 47(1):35–51.
- 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. (2016). Identifying Sibling Influence on Teenage Substance Use. *Journal of Human Resources*.
- Andersson, L. and Hammarstedt, M. (2010). Intergenerational Transmissions in Immigrant Self-Employment: Evidence from Three Generations. *Small Business Economics*, 34(3):261–276.
- Åstebro, T. and Tåg, J. (2015). Entrepreneurship and Job Creation. *IFN Working Paper*.
- 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.
- Colombier, N. and Masclet, D. (2008). Intergenerational Correlation in Self Employment: Some Further Evidence from French ECHP Data. *Small Business Economics*, 30(4):423–437.
- 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*, 11320.
- De Wit, G. and Van Winden, F. A. A. M. (1989). An Empirical Analysis of Self-Employment in the Netherlands. *Small Business Economics*, 1(4):263–272.
- Decker, R., Haltiwanger, J., Jarmin, R., and Miranda, J. (2014). The Role of Entrepreneurship in US Job Creation and Economic Dynamism. *Journal of Economic Perspectives*, 28(3):3–24.
- 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.
- Elert, N., Andersson, F. W., and Wennberg, K. (2015). The Impact of Entrepreneurship Education in High School on Long-Term Entrepreneurial Performance. *Journal of Economic Behavior and Organization*, 111(1):209–223.
- 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 Journal: Economic Policy*, 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*, 21775.
- Guo, G. and Wang, J. M. (2002). The Mixed or Multilevel Model for Behavior Genetic Analysis. *Behavior Genetics*, 32(1):37–49.
- Guzman, J. and Stern, S. (2015). Nowcasting and Placecasting Entrepreneurial Quality and Performance. *NBER Working Paper Series*, 20954.
- 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., 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.
- Hundley, G. (2006). Family Background and the Propensity for Self-Employment. *Industrial Relations*, 45(3):377–392.
- Hurst, E. and Lusardi, A. (2004). Liquidity Constraints, Household Wealth, and Entrepreneurship. *Journal of Political Economy*, 112(2):319–347.
- 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*, 21597.
- 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*, 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.
- 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, C. 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.
- 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.
- Sulloway, F. J. (1996). *Born to Rebel*. Pantheon Books, New York, NY.
- 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.
- Tervo, H. (2006). Regional Unemployment, Self-Employment and Family Background. *Applied Economics*, 38(9):1055–1062.
- van Praag, C. M. and Raknerud, A. (2014). The Entrepreneurial Earnings Puzzle. Evidence from Matched Person-Firm Data. *Working Paper*.
- 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.
- Zunino, D. (2016). Are Genetics and Environment Substitutes or Complements in Affecting Entrepreneurial Choice? *Working Paper*.

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

N Children	No. of Families	%	No. of Individuals	%
<i>1</i>	227,519	52.40	227,519	32.26
<i>2</i>	154,051	35.48	308,102	43.69
<i>3</i>	43,026	9.91	129,078	18.30
<i>4</i>	7,881	1.82	31,524	4.47
<i>5</i>	1,401	0.32	7,005	0.99
<i>6</i>	255	0.06	1,530	0.22
<i>7</i>	56	0.01	392	0.06
<i>8</i>	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.

Table 2: **Descriptive Statistics**

	Mean	S.D.	<i>N</i>	Min	Max
A. Entrepreneurial outcomes					
Self-employed	0.146	(0.353)	705,262	0	1
Incorporated	0.083	(0.276)	705,262	0	1
Self-employed $\geq 4y$	0.076	(0.265)	705,262	0	1
Incorporated $\geq 5y$	0.042	(0.201)	705,262	0	1
High income self-employed	0.073	(0.260)	705,262	0	1
Years self-employed ^a	5.290	(5.800)	102,921	1	28
Self-employment 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 self-employed	0.152	(0.359)	705,262	0	1
Father self-employed	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 self-employed	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: Descriptive Statistics by Entrepreneurial Status

	(1)	(2)	(3)	(4)	(5)	(6)
	Employee	Self- employed	Incorporated	(2)-(1)	(3)-(1)	(3)-(2)
Years of schooling	12.311 (2.181)	11.960 (2.103)	12.366 (2.106)	-0.351 ***	0.055 ***	0.406 ***
Log income	11.829 (0.771)	11.745 (0.513)	12.222 (0.465)	-0.084 ***	0.393 ***	0.477 ***
Mother self-employed	0.134 (0.341)	0.233 (0.423)	0.195 (0.396)	0.100 ***	0.061 ***	-0.038 ***
Father self-employed	0.228 (0.419)	0.357 (0.479)	0.309 (0.462)	0.130 ***	0.082 ***	-0.048 ***
Mother incorporated	0.024 (0.153)	0.032 (0.177)	0.121 (0.327)	0.008 ***	0.098 ***	0.089 ***
Father incorporated	0.052 (0.223)	0.063 (0.244)	0.198 (0.398)	0.011 ***	0.145 ***	0.135 ***
Mother's log income	11.585 (0.847)	11.582 (0.825)	11.745 (0.671)	-0.004	0.159 ***	0.163 ***
Father's log income	12.164 (0.659)	12.124 (0.719)	12.363 (0.632)	-0.040 ***	0.199 ***	0.239 ***
Mother's years of schooling	10.000 (2.784)	10.106 (2.842)	10.404 (2.860)	0.107 ***	0.405 ***	0.298 ***
Father's years of schooling	9.946 (3.032)	10.021 (3.069)	10.376 (3.122)	0.075 ***	0.430 ***	0.356 ***

Standard errors in parentheses. All differences (with one exception for employees and the self-employed with regards to maternal income) are significant at less than 1 percent. Please note that some individuals have been both self-employed and incorporated at different points in time. They are omitted from this analysis, but the results in columns (4) and (5) are very similar if they are counted both as *Self-employed* and as *Incorporated*.

Table 4: **Sibling Correlations in Entrepreneurship**

	(1)	(2)	(3)	(4)
	All	Male	Female	Mixed
A1. Extensive margin				
<i>Self-employed</i>	0.232	0.318	0.232	0.198
	(0.004)	(0.008)	(0.011)	(0.005)
<i>Incorporated</i>	0.341	0.399	0.347	0.306
	(0.005)	(0.008)	(0.015)	(0.007)
A2. Stricter definitions				
<i>Self-employed</i> $\geq 4y$	0.262	0.365	0.245	0.219
	(0.005)	(0.010)	(0.016)	(0.007)
<i>Incorporated</i> $\geq 5y$	0.424	0.477	0.448	0.380
	(0.006)	(0.010)	(0.021)	(0.009)
<i>High income self-employed</i>	0.175	0.215	0.193	0.153
	(0.006)	(0.011)	(0.017)	(0.008)
Individuals	705,262	218,579	201,800	284,883
Families	434,203	163,583	153,211	117,409
B. Intensive margin				
<i>Years self-employed</i>	0.214	0.315	0.185	0.167
	(0.008)	(0.013)	(0.028)	(0.011)
<i>Self-employment income</i>	0.219	0.313	0.153	0.183
	(0.008)	(0.014)	(0.027)	(0.011)
Individuals	102,091	38,396	22,692	41,833
Families	91,513	34,939	21,509	35,056
<i>Years incorporated</i>	0.387	0.467	0.396	0.312
	(0.009)	(0.013)	(0.032)	(0.014)
Individuals	58,414	25,933	9,244	23,237
Families	53,160	23,754	8,902	20,504

Standard errors in parentheses. Each sibling correlation comes from a different model; all models control for gender and birth year.

Table 5: **Sibling Differences**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Self-employed	Incorporated	Self-employed $\geq 4y$	Incorporated $\geq 5y$	High income self-employed	Years self-employed	Self-employment income	Years incorporated
A. Baseline sibling correlations from Table 4								
ρ	0.232 (0.004)	0.341 (0.005)	0.262 (0.005)	0.424 (0.006)	0.175 (0.006)	0.214 (0.008)	0.219 (0.008)	0.387 (0.009)
B. Share of individual level variation, σ_b^2, explained by differential treatment ^a								
R^2	0.013	0.018	0.011	0.016	0.004	0.024	0.013	0.030
C. Revised measure of family influences ^{b,c}								
$\tilde{\rho}$	0.242 (0.004)	0.355 (0.005)	0.270 (0.005)	0.432 (0.006)	0.178 (0.006)	0.233 (0.008)	0.228 (0.008)	0.405 (0.009)
D. Percentage increase from ρ to $\tilde{\rho}$								
	4.17%	3.31%	2.96%	2.15%	1.79%	8.11%	4.48%	4.55%

Standard errors in parentheses.

^a The second stage explanatory variables are: a dummy for being first born, the two-way interactions of i) first born and parental entrepreneurship, ii) first born with gender, iii) gender and parental entrepreneurship, and the three-way interaction of first born, gender, and parental entrepreneurship.

^b The confidence intervals for the baseline and ‘revised’ sibling correlations are largely overlapping.

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

Table 6: **Neighborhood Correlations**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Self-employed	Incorporated	Self-employed $\geq 4y$	Incorporated $\geq 5y$	High income self-employed	Years self-employed	Self-employment income	Years incorporated
A. Baseline sibling correlations from Table 4								
ρ	0.232 (0.004)	0.341 (0.005)	0.262 (0.005)	0.424 (0.007)	0.175 (0.006)	0.214 (0.008)	0.219 (0.008)	0.387 (0.009)
B. Neighborhood correlations controlling for parental characteristics								
ρ_n	0.016 (0.001)	0.010 (0.001)	0.024 (0.002)	0.014 (0.001)	0.007 (0.001)	0.024 (0.002)	0.011 (0.001)	0.017 (0.002)
C. Share of sibling correlation accounted for by neighborhood effects								
	6.91%	3.00%	9.27%	3.21%	4.07%	11.39%	4.81%	4.33%

Standard errors in parentheses.

Table 7: Accounting Exercise: Family Factors Generating Sibling Similarities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Self-employed	Incorporated	Self-employed $\geq 4y$	Incorporated $\geq 5y$	High income self-employed	Years self-employed	Self-employment 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								
<i>Education, income</i>	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%
<i>Immigration</i>	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%
<i>Self-employment</i>	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%
<i>Incorporation</i>	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%
<i>Family structure</i>	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. Conditional on non-missing parental characteristics, this exercise restricts the sample to 665,662 individuals in 410,133 families in columns (1) to (5), 96,368 individuals in 85,926 families in columns (6) and (7) and 55,828 individuals in 50,804 families in column (8).

Table 8: Upper Bounds on Peer Effects Contribution to Sibling Correlations

	Effect on younger sibling		Effect on older sibling	
	(1)	(2)	(3)	(4)
A. Self-employment				
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 Significant at 5 percent. ^b Significant at 10 percent.

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 C.4 and appendix Tables C.5-C.8 (the results in this table are based on columns (3), (4), (8) and (9) in those tables).

Table 9: **Twin Correlations for DZ and MZ Twins**

	(1)	(2)	(3)	(4)	(5)
	Self-employed	Incorporated	Self-employed $\geq 4y$	Incorporated $\geq 5y$	High income self-employed
A. Males					
ρ_{MZ}	0.507 (0.068)	0.637 (0.064)	0.639 (0.067)	0.597 (0.100)	0.487 (0.091)
ρ_{DZ}	0.338 (0.035)	0.455 (0.039)	0.385 (0.048)	0.533 (0.046)	0.299 (0.057)
Heritability:	0.338 (0.173) [0.051]	0.364 (0.170) [0.033]	0.506 (0.186) [0.007]	0.127 (0.252) [0.614]	0.376 (0.251) [0.133]
$100 * (\rho_{MZ} - \rho_{DZ}) / \rho_{DZ}$	50.0%	40.0%	66.0%	12.0%	62.9%
B. Females					
ρ_{MZ}	0.546 (0.065)	0.493 (0.128)	0.416 (0.133)	0.613 (0.134)	0.423 (0.125)
ρ_{DZ}	0.232 (0.048)	0.360 (0.060)	0.284 (0.069)	0.339 (0.108)	0.219 (0.076)
Heritability:	0.628 (0.185) [0.001]	0.267 (0.318) [0.401]	0.264 (0.345) [0.444]	0.548 (0.396) [0.166]	0.408 (0.341) [0.232]
$100 * (\rho_{MZ} - \rho_{DZ}) / \rho_{DZ}$	135.3%	36.9%	46.5%	80.8%	93.2%

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 10: **Summary of Results**

	Self-employment outcomes	Incorporation outcomes
A. Main results: sibling similarities		
Sibling correlation (ρ)	0.175 - 0.262	0.341 - 0.424
B. Sibling differences		
Differential treatment ‘bias’	$\approx 4\%$	$\approx 4\%$
C. Mechanisms: share of sibling correlation explained		
Neighborhood effects	4 - 10%	3 - 12%
Parental characteristics	10 - 21%	10 - 21%
Self-employment	6 - 15%	0 - 2%
Incorporation	0 - 3%	8 - 16%
Education and income	0 - 11%	0 - 2%
Other	0 - 2%	0 - 2%
Sibling peer effects ^a	-7 - 7%	-3 - 3%
Shared genes ^b	56 - 78%	38 - 46%

^a Only significant for self-employment and male pairs, see Table 8.

^b For the significance of heritability estimates, see Table 9. We take the average of the separate values for men and women.

A Appendix: Peer Effects Model

In this appendix, we provide a more formal exposition of the correlated random effects model suggested by [Altonji et al. \(2016\)](#) that we adopt for the purpose of estimating sibling peer effects in entrepreneurship. We begin by estimating the raw association between sibling i 's self-employment (incorporation) at time t , S_t^i , and sibling i' 's self-employment (incorporation) at time $t - 1$, $S_{t-1}^{i'}$:

$$S_t^i = \beta_0 + \beta_1 S_{t-1}^{i'} + u_t^2, \quad (\text{A.1})$$

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 :

$$S_t^i = \beta_0 + \beta_1 S_{t-1}^{i'} + X^f + age_t^i + \epsilon_t^2. \quad (\text{A.2})$$

We estimate equations (A.1) and (A.2) (corresponding to columns (1), (2), (6) and (7) in Tables C.4 to C.8) 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 C.4.

Part of the effect of sibling i' 's entrepreneurial status on sibling i 's entrepreneurial status estimated in equation (A.2), however, may be due to correlated random family effects, rather than direct peer effects. [Altonji et al. \(2016\)](#) 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 directly target 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:

$$S_t^i = \beta_0 + \beta_1 (S_{t-1}^{i'} + S_{t+1}^{i'}) + \lambda_0 S_{t-1}^{i'} + X^f + age_t^i + age_t^{i'} + \epsilon_t^2, \quad (\text{A.3})$$

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

$$S_t^i = \beta_0 + \beta_1 (S_{t-1}^{i'} + S_t^{i'} + S_{t+1}^{i'}) + \lambda_0 S_{t-1}^{i'} + \lambda_1 S_t^{i'} + X^f + age_t^i + age_t^{i'} + \epsilon_t^2, \quad (\text{A.4})$$

where λ_1 is the estimate of the contemporaneous effect. This estimate should not be (over-) interpreted

²⁸ Sibling correlations for this sample closely match those reported in Table 4; see also footnote 13.

²⁹ A detailed description of the assumptions and mechanics of this model is provided in [Altonji et al. \(2016\)](#).

as a contemporaneous effect, but rather as a transitory and common shock to both siblings in the same family. 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 (A.3) and (A.4) are given in columns (3), (4), (8) and (9) of Tables C.4-C.8, while columns (5) and (10) present results from a variation of equation (A.1), where the lagged sibling effect is replaced by the contemporaneous one (this equation being necessary for calibration purposes).

As an example of how these tables should be interpreted, table C.4 shows the results of our sibling peer effects exercise on the sub-sample of sibling pairs, with panel A referring to self-employment 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 self-employment status at time $t - 1$ on the younger sibling's self-employment status at time t (Table C.4, 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.

B Appendix: Genetic Decomposition

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, following Nicoletti and Rabe (2013). The most important source of information comes from directly observable mixed sex twins, who must be DZ twins. Thus, we can directly estimate $\sigma_{DZ,FM}^2$, 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, $\sigma_{DZ,MM}^2$ and $\sigma_{DZ,FF}^2$, we make the following assumptions:

$$\sigma_{DZ,MM}^2 = \sigma_{DZ,FM}^2 + (\sigma_{NT,MM}^2 - \sigma_{NT,FM}^2), \quad (\text{B.1})$$

$$\sigma_{DZ,FF}^2 = \sigma_{DZ,FM}^2 + (\sigma_{NT,FF}^2 - \sigma_{NT,FM}^2), \quad (\text{B.2})$$

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).³⁰

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 aver-

³⁰ 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 C.1 in Appendix C 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.

age of corresponding variances for MZ and DZ twin pairs, with weights provided by their incidence in the population of same sex twins.³¹ Denoting these proportions with P (and noting that $P_{MZ,MM} + P_{DZ,MM} = P_{MZ,FF} + P_{DZ,FF} = 1$), we use:

$$\sigma_{T,MM}^2 = \sigma_{MZ,MM}^2 \times P_{MZ,MM} + \sigma_{DZ,MM}^2 \times P_{DZ,MM} \quad (\text{B.3})$$

$$\sigma_{T,FF}^2 = \sigma_{MZ,FF}^2 \times P_{MZ,FF} + \sigma_{DZ,FF}^2 \times P_{DZ,FF}. \quad (\text{B.4})$$

Now we need to approximate these proportions in order to solve the MZ-variances from equations (B.3) and (B.4). 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:

$$N_{DZ,MM} = N_{DZ,FF} = 0.5 \times N_{DZ,FM}, \quad (\text{B.5})$$

$$N_{MZ,MM} = N_{T,MM} - N_{DZ,MM} \text{ and } N_{MZ,FF} = N_{T,FF} - N_{DZ,FF}, \quad (\text{B.6})$$

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 (B.3) and (B.4). The individual-level variance is fixed at $\pi^2/3$ for the dichotomous outcomes we analyze.³² 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. We then back out measures of heritability for each gender, $h_{MM}^2 = 2 \times (\rho_{MZ,MM} - \rho_{DZ,MM})$ and $h_{FF}^2 = 2 \times (\rho_{MZ,FF} - \rho_{DZ,FF})$.

³¹ An implicit assumption is that the means of the family components for the two types of twins are identical.

³² We do not analyze the intensive margin outcomes since sample sizes are extremely small.

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, as well as the contribution of genes to sibling similarities in entrepreneurship.

C Appendix: Additional Figures and Tables

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

Table C.1 shows the results of several robustness checks: excluding singletons, re-defining the family on the basis of the father, restricting the sample to families with two children only, restricting the sample to closely spaced siblings (12-24 months) in families with two children only, restricting the sample to families we completely capture in our sample (i.e. no siblings born outside the period 1960-1970), and a placebo test, where individuals are grouped randomly, matching the original cluster structure (i.e. number and size).

Table C.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.

Table C.3 examines the contribution of parental self-employment and incorporation to sibling correlations in *Self-employed* and *Incorporated* by type of sibship (male, female, mixed).

Tables C.4-C.8 show the results of Altonji-style peer effects exercises for self-employment and incorporation, first for all siblings, and then by each type of the sibling pair (males, females, mixed gender with younger brother or younger sister).

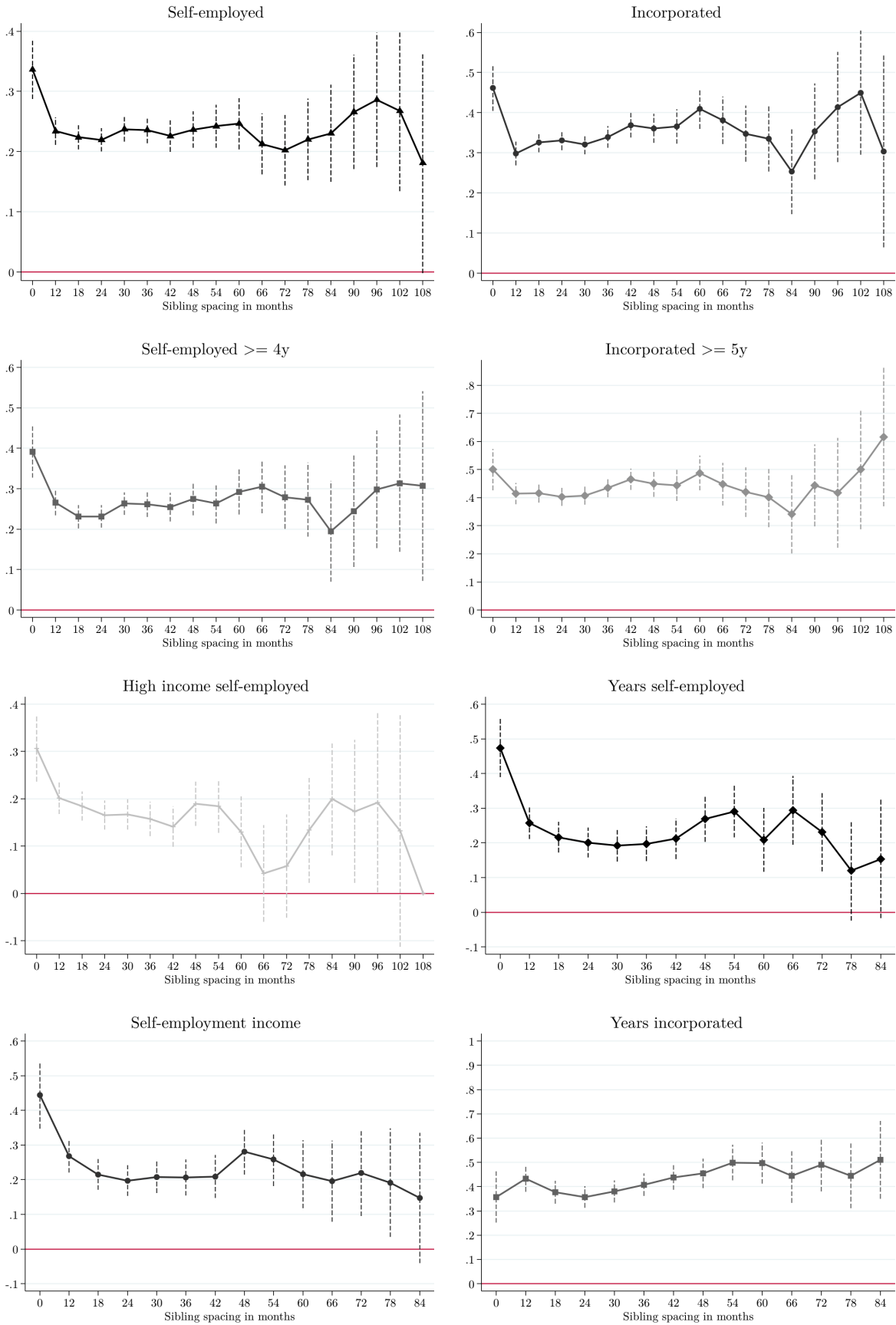


Figure C.1: Twin and sibling correlations in entrepreneurship by sibling spacing.

Table C.1: Sibling Correlations in Entrepreneurship: Robustness Checks

	(1)	(2)	(3)	(4)	(5)	(6)
	All, excl. singles	All, father	Two children	Closely spaced	Complete families	'Fake' families
A1. Extensive margin						
<i>Self-employed</i>	0.232 (0.004)	0.224 (0.004)	0.229 (0.005)	0.234 (0.012)	0.224 (0.006)	0.0001 (0.0001)
<i>Incorporated</i>	0.341 (0.005)	0.331 (0.005)	0.335 (0.007)	0.300 (0.015)	0.341 (0.007)	0.0013 (0.0033)
A2. Stricter definitions						
<i>Self-employed</i> $\geq 4y$	0.262 (0.005)	0.252 (0.005)	0.256 (0.008)	0.266 (0.016)	0.245 (0.008)	0.0002 (0.0002)
<i>Incorporated</i> $\geq 5y$	0.424 (0.006)	0.413 (0.007)	0.425 (0.009)	0.413 (0.020)	0.417 (0.009)	0.0006 (0.0006)
<i>High income Self-employed</i>	0.175 (0.006)	0.175 (0.006)	0.172 (0.008)	0.202 (0.018)	0.173 (0.009)	0.0160 (0.0074)
Individuals	477,743	694,040	308,102	64,732	277,410	705,262
Families	206,684	424,391	154,051	32,366	144,939	434,203
B. Continuous outcomes						
<i>Years Self-employed</i>	0.228 (0.009)	0.209 (0.008)	0.225 (0.012)	0.257 (0.024)	0.196 (0.011)	0.0044 (0.0116)
<i>Self-employment income</i>	0.221 (0.009)	0.214 (0.008)	0.220 (0.012)	0.268 (0.024)	0.196 (0.012)	0.0000 (0.0000)
Individuals	21,766	101,307	44,587	9,499	40,323	102,091
Families	10,358	89,954	39,271	8,337	34,810	91,513
<i>Years incorporated</i>	0.435 (0.011)	0.384 (0.009)	0.407 (0.012)	0.433 (0.028)	0.385 (0.012)	0.0000 (0.0000)
Individuals	10,096	57,823	26,585	5,482	24,120	58,414
Families	4,842	52,610	23,868	4,973	21,224	56,025

Standard errors in parentheses. Column (1) eliminates singletons, column (2) redefines the family on the basis of the father, column (3) restricts the sample to two-child families, and column (4) restricts it to closely spaced siblings (12-24 months) in two-child families. Column (5) only uses families we entirely capture in our sample. Column (6) shows an example of results from an exercise that defines random clusters that match the mother-defined families in number and size; as expected, none of the sibling correlations in this column are significantly different from zero, with the exception of the one for *High income entrepreneur*, which is still an order of magnitude below the original sibling correlation (and is an artifact of the particular randomization seed used in this example). An additional robustness check where we consider only entrepreneurship events between the ages 25 and 40 also produces results very similar to the baseline sibling correlations: 0.222 (0.005) for *Self-employed* and 0.367 (0.007) for *Incorporated*.

Table C.2: Factor Analysis Results (Factor Loadings)

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

Main factor loadings appear in bold.

Table C.3: Sibship Type and Parental Entrepreneurship

	Male		Female		Mixed ^a	
	Self-employed	Incorporated	Self-employed	Incorporated	Self-employed	Incorporated
A. Sample sibling correlations ^b						
	0.318	0.399	0.233	0.349	0.197	0.306
	(0.008)	(0.009)	(0.011)	(0.015)	(0.005)	(0.007)
B. Controlling for parental entrepreneurship						
<i>Mother self-employed</i>	0.304	0.394	0.222	0.344	0.179	0.298
	(0.008)	(0.009)	(0.011)	(0.015)	(0.005)	(0.007)
	4.64%	1.36%	4.60%	1.50%	8.90%	2.51%
<i>Father self-employed</i>	0.295	0.391	0.221	0.345	0.176	0.297
	(0.008)	(0.009)	(0.011)	(0.015)	(0.005)	(0.007)
	7.34%	2.14%	5.16%	1.26%	10.72%	2.98%
<i>Mother incorporated</i>	0.318	0.370	0.231	0.312	0.196	0.267
	(0.008)	(0.009)	(0.011)	(0.016)	(0.005)	(0.007)
	0.05%	7.21%	0.98%	10.57%	0.62%	12.71%
<i>Father incorporated</i>	0.318	0.354	0.231	0.325	0.195	0.260
	(0.008)	(0.009)	(0.011)	(0.016)	(0.005)	(0.007)
	0.14%	11.24%	0.80%	6.93%	0.68%	15.09%
Individuals	214,757	214,757	198,216	198,216	281,067	281,067
Families	160,693	160,693	150,465	150,465	116,665	116,665

Standard errors in parentheses. Numbers in bold refer to the percentage decrease in the sibling correlation.

^a The fixed part of these models controls for gender.

^b The sample is restricted to siblings for whom both parents' entrepreneurial status is available.

Table C.4: Peer Effects Exercise, All Sibling Types

A. Self-employment										
	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($S_{t-1}^{i'}$)	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($S_t^{i'}$)				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($S_{t-1}^{i'}$)	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($S_t^{i'}$)				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	
$S_{t-1}^{i'} + S_{t+1}^{i'}$			Yes					Yes		
$S_{t-1}^{i'} + S_t^{i'} + S_{t+1}^{i'}$				Yes					Yes	

^a Significant at 5 percent. ^b Significant at 10 percent. 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, self-employment, 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 (Bonnett, 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 C.5: Peer Effects Exercise, Males

A. Self-employment										
	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($S_{t-1}^{i'}$)	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($S_t^{i'}$)				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($S_{t-1}^{i'}$)	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($S_t^{i'}$)				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	
$S_{t-1}^{i'} + S_{t+1}^{i'}$			Yes					Yes		
$S_{t-1}^{i'} + S_t^{i'} + S_{t+1}^{i'}$				Yes					Yes	

See notes to Table C.4.

Table C.6: **Peer Effects Exercise, Females**

A. Self-employment										
	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($S_{t-1}^{i'}$)	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($S_t^{i'}$)				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($S_{t-1}^{i'}$)	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($S_t^{i'}$)				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	
$S_{t-1}^{i'} + S_{t+1}^{i'}$			Yes					Yes		
$S_{t-1}^{i'} + S_t^{i'} + S_{t+1}^{i'}$				Yes					Yes	

See notes to Table C.4.

Table C.7: Peer Effects Exercise, Mixed (Younger Brother)

A. Self-employment										
	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 C.4.

Table C.8: Peer Effects Exercise, Mixed (Younger Sister)

A. Self-employment										
	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)
OR($S_{t-1}^{i'}$)	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	
OR($S_t^{i'}$)				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)
OR($S_{t-1}^{i'}$)	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	
OR($S_t^{i'}$)				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	
$S_{t-1}^{i'} + S_{t+1}^{i'}$			Yes					Yes		
$S_{t-1}^{i'} + S_t^{i'} + S_{t+1}^{i'}$				Yes					Yes	

See notes to Table C.4.