



DRUID
society

Paper to be presented at
the DRUID16 20th Anniversary Conference
Copenhagen, June 13-15, 2016

Are Genetics and Environment Substitutes or Complements in affecting Entrepreneurial Choice?

Diego Zunino
Copenhagen Business School
INO
dz.ino@cbs.dk

Abstract

Recent twin and adoption studies have shown that genes matter for entrepreneurial choice. This related study addresses how a genetic predisposition to entrepreneurship interacts with the (entrepreneurship friendliness of the) environment, using a dataset of Italian twins. In particular, we study whether the genetic effect is different across genders, based on the stylized fact that barriers to entrepreneurship entry are stronger for females than for males. Using regression analysis, the study confirms earlier findings showing substantial genetic effects.

More interestingly, the study finds that the genetic effect drops when the environment is less favorable – namely when the individual is female. The result of a positive interaction between predisposition and environment implies that the environment and predisposition can be considered complements rather than substitutes, that institutions play a role, and that a favorable environment to entrepreneurship selects those with higher predisposition rather than simply increasing the rate of self-employment.

Are Genetics and Environment Substitutes or Complements in affecting Entrepreneurial Choice?

ABSTRACT

Recent twin and adoption studies have shown that genes matter for entrepreneurial choice. This related study addresses how a genetic predisposition to entrepreneurship interacts with the (entrepreneurship friendliness of the) environment, using a dataset of Italian twins. In particular, we study whether the genetic effect is different across genders, based on the stylized fact that barriers to entrepreneurship entry are stronger for females than for males. Using regression analysis, the study confirms earlier findings showing substantial genetic effects. More interestingly, the study finds that the genetic effect drops when the environment is less favorable – namely when the individual is female. The result of a positive interaction between predisposition and environment implies that the environment and predisposition can be considered complements rather than substitutes, that institutions play a role, and that a favorable environment to entrepreneurship selects those with higher predisposition rather than simply increasing the rate of self-employment.

Keywords:

Entrepreneurship; entrepreneurial environment; twins

1. Introduction

What makes an entrepreneur? Is it nature or is it nurture? The research question is not simply a philosophical one: every year governments and organizations invest resources to create an “entrepreneurial” environment. Both the benefits of entrepreneurship (King & Levine 1993; Wennekers & Thurik 1999) and the positive role of the environment (Chinitz 1961; Gleaser et al 2015) have been amply supported by research. Teaching entrepreneurship and providing resources in the form of social and financial capital are the most common programs, all of them involving a cost to society. These expenses would probably more justified the larger is the effect of nurture vis-à-vis nature. By and large, the literature has shown that both genes (Nicolaou et al 2008) and the environment (Sorenson & Audia 2000) matter as determinants of entrepreneurial choice and that the effect of nurture is larger than of nature (Lindquist et al., 2015). Therefore, the question arises whether an interaction exists and what its direction is.

Ex ante, the direction of a potential interaction can go both ways. In a first line of reasoning, individuals need to rely on their predisposition to a greater extent when the environment is not favorable (Zhang et al. 2009). Under an alternative perspective, genes allow the individuals who are genetically more entrepreneurial to exploit more opportunities (Nicolaou & Shane 2009) – which lie in entrepreneurship-favorable environments. The only adoption study we know that measures such an interaction is Lindquist et al. (2015). They find no evidence of such an interaction. In their adoption study the environment is characterized by the entrepreneurial background of adoptive parents and genes are characterized by the entrepreneurial background of biological parents. Earlier attempts to study how genes and environment interact provided ambiguous evidence and no clear answers (Zhang et al. 2009, Nicolaou & Shane 2010). These studies have acknowledged that there is a research gap and they call for studies that might result in a deeper understanding of it (Nicolaou & Shane 2009,

Nicolaou & Shane 2010, Zhang et al. 2009).

We attempt to address the research question about the interaction between genetic endowment and environment through a twin study using a unique dataset of twin pairs from the Italian Twin Registry (Stazi et al. 2002). Conscious of all the possible limitations expressed in Nicolaou & Shane (2010), we propose a regression analysis where we use half of the observations, i.e., a random twin of the pair¹. Because the assignment of being identical or fraternal twins is as good as random, the difference between identical and fraternal twins can be used as an exogenous assignment of the extent of shared genes among siblings in a regression model. In studying the existence of a genetic effect, we find evidence that is consistent with the earlier literature (Nicolaou et al 2008; Lindquist et al 2015).

In studying the interaction between the genetic effect and environment, we operationalize environment through gender (Zhang et al 2009, Nicolaou & Shane 2010). Women have to suffer the costs of stereotypes (Gupta et al. 2014), as entrepreneurship is considered a gendered institution (Buttner & Rosen 1988). The entrepreneurship literature is abundant of references about how women find it harder to attract funding from loans (Alesina et al. 2013), business angels (Becker-Blease & Sohl 2007), and venture capitalists (Gompers et al. 2014) with evidence of taste discrimination (Alesina et al. 2013).

We split the sample between men and women and we find evidence of a significant genetic effect for men and no significant effect for women, in apparent contrast with the findings of Zhang et al. (2009). We corroborate these results in three ways. First, we match the sample in order to reduce the imbalance between male and females in order to obtain a lower bound estimate, and the gender-specific effect becomes even larger. Second, in lieu of gender as proxy of the environment, we use the province of residence as an alternative. In fact, living in

¹ For robustness, and to avoid selection concerns, we ran the entire analysis also on the other random twin with no significant differences in the results.

the Milan province should approximate a favorable environment, while we approximate the opposite with residence in the Rome province. Third, we analyze our data at the pair level to be explained more precisely, in this way we observe whether being identical twins correlates with the entrepreneurial choice for both the pair members.

Our results contribute to the literature in the following way: using an alternative and more intuitive methodology for twin studies based on regression analysis, we provide evidence that a favorable entrepreneurial environment is a complement to predisposition. This means that entrepreneurship does not happen in vacuum, even when someone is naturally inclined. Among the potential explanations for the genetic antecedents of entrepreneurship, Nicolaou & Shane (2009) list the reaction of genes to environmental stimuli. While this could be the case, it can also be that genes that correlate with entrepreneurship correlate also with talent, and different environments reward talent in different ways (Baumol 1990, Murphy et al. 1991). In fact, we observed that predisposition matters only in environments that provide lower cost and better opportunities for entrepreneurship as a career.

The paper proceeds as follows: section 2 formulates hypotheses based on theory. Section 3 describes the data and the variables and section 4 presents the method. Section 5 shows the results, whereas section 6 discusses the results in light of their limitations. Section 7 concludes.

2. Theory

Since its inception as a field, the entrepreneurship literature has looked for “who is the entrepreneur” (Kirzner 1978). Especially in the earlier stage, entrepreneurship scholars searched for the roots of entrepreneurship in individuals, their characteristics, and their personality traits. However, more recent attempts tried to look at testosterone levels (Guiso and Rustichini 2011) and brain activity (Laureiro-Martínez *et al.* 2013). The type of focus is

particularly widespread among economists and psychologists and this approach is usually labeled as micro-oriented or dispositional.

Sociology takes a different stand by looking at the context (Gartner 1988). Aldrich argues that "[s]ystematic empirical research has consistently found that personality traits, taken out of context, simply do not explain very much. Indeed, research on personality traits seems to have reached a dead end." (Aldrich 1999, p. 77). Such an alternative view finds more relevant antecedents in external factors such as legitimacy (Aldrich & Fiol 1994), social networks (Sorenson & Audia 2000, Stuart & Sorenson 2003), and peers (Nanda & Sorensen 2010; Kacperczyk 2013). The labels for this approach are macro-oriented or situational– to highlight the dualism (Sørensen 2007, Thornton 1999).

The study of the genetic foundation of entrepreneurship (Nicolaou et al. 2008) represents perhaps the most radical attempt to look at micro-oriented antecedents. A genetic effect raises attention because a large share of the variance would be explained by a factor that is transmitted and over which the individual has no control. Thus, policy efforts to create an “entrepreneurial environment” would have little effectiveness and this would imply limited social mobility. However, genetic predisposition does not act in isolation, and two of the four² proposed explanations allow for a mechanism that involves the environment: environmental stimuli and selection into the environment (Nicolaou & Shane 2009). The twin studies that followed the seminal work of Nicolaou et al. (2008) tried to address how genes and environment interact. The evidence is mixed: when being male is a proxy of a better environment to do entrepreneurship, genetics matters less (Zhang et al. 2009). However, another study finds evidence of a genetic effect for both genders (Nicolaou & Shane 2010).

With a different approach, using adopted individuals in Sweden, Lindquist et al. (2015) study

² The other two mechanisms are: genes affect chemical mechanisms in brain, and genes influence individual characteristics such as extraversion and neuroticism

the role of nature versus nurture. They operationalize nature as the entrepreneurial experience of biological parents that adopted individuals have never met, while they operationalize nurture as the entrepreneurial experience of adoptive parents that did not transmit a single gene to the adopted individuals. They model the relationship as additive, and they find that the role of nurture is twice as large as the role of nature. When they added their interaction as control, they found no significance.

We formulate a set of hypotheses from the existing theory on genes and entrepreneurship. First, the necessary condition for this study is the existence of predisposition to entrepreneurship i.e., a genetic effect on entrepreneurship (Nicolaou et al. 2008, Nicolaou & Shane 2009, 2010). Therefore, we hypothesize the following:

Hypothesis 1 *An individual who shares more genes with an entrepreneur is more likely to enter entrepreneurship.*

To theorize about the interaction, we cannot have priors, given the contrasting evidence. On the one hand, we can follow Zhang et al. (2009) and posit a negative interaction between the genetic effect and the environment: when individuals are in environments that are not favorable to entrepreneurship (“low” pro-entrepreneurial environments), they have to rely more on their predisposition.

Therefore, we hypothesize:

Hypothesis 2a. *The genetic effect on entrepreneurship is stronger when an environment is not favorable to entrepreneurship.*

Similarly, we can develop an alternative theory following more directly Nicolaou & Shane (2009): predisposition makes people more sensitive to the stimuli by the environment. This would imply a positive interaction between genetic effect and environment. We explain this

theory through two arguments: epigenetic and economic. The epigenetic argument argues that some particular environmental stimuli trigger some entrepreneurship-related genes: we observe the phenotype (observable characteristic) manifestation only in the presence of certain favorable environments (Fraga et al. 2005). The economic argument comes from Baumol (1990), and more formally from Murphy et al. (1991). If entrepreneurship predisposition is just generally talent, unfavorable entrepreneurial environments provide alternative and more attractive ways to reward talent (e.g., career in the military, public administration, or in organized crime). If returns to entrepreneurship are a function of predisposition and environment, a favorable environment is a multiplier of their predisposition. Thus, it follows that:

Hypothesis 2b. *The genetic effect on entrepreneurship is stronger when an environment is favorable to entrepreneurship*

3. Data and Variables

3.1 Data

The data for this twin study come from the Italian Twin Registry (ITR), managed by the Rome-based Italian “Istituto Superiore di Sanità”. The ITR started in 2001 (Stazi et al. 2002), and its mission is to provide a scientific tool to identify the genetic, environmental, and lifestyle factors that influence the well-being of individuals (e.g., among many Greco et al. 2002; Martini et al. 2015). Because the mission of the registry is not focused on research in social sciences, it includes a limited number of socio-demographic variables. The variables we can use are: type of twin, age, level of education, employment, main type of employment during lifetime, and province of residence.

By the end of 2011, 24,800 pairs of twins were enrolled in a nonrandom way³. The registry was able to identify the type of twins through questions about their physical appearance (Fagnani et al., 2006) and DNA testing (Brescianini et al., 2013) with an accuracy rate above 90% (Fagnani et al., 2014). Among them, we select a sample of twins who were born between 1946 and 1976: the number of twins in our sample drops to 2928. The small number of twins available is due to the young age of the registry as it has been established in 2001. By 2013, the updated twin registry counted 63% of twins born after 1982 (Fanagni et al., 2014). Because the data is cross sectional, younger individuals would have entered the ITR as students, thus out of the labor force. Similarly, we excluded elderly pairs who could have entered the registry already as pensioners⁴. Such extraction leaves us with 1724 twins.

Because the data have no single code for self-employment status, we hand-coded self-employment based on the self-reported profession. We identified three categories of professions related to self-employment. In the first group, there are individuals whose self-employment status was clear (e.g. “owner of a company”). In the second group, there are individuals who are working as professionals but they clearly state that they are self-employed (e.g., “self-employed lawyer”). Finally, there are individuals who report a trivial self-employment status: they declare their job to be a category at risk of self-employment, but with some measurement error.

In order to arrive at our final sample, we add four additional restrictions. First, we exclude the pairs of fraternal twins of different sex as the difference in sex could confound the results.

³ Twins were enrolled through five channels. The first channel is the analysis of the municipalities and following contact of the pair by the registry. The second channel is the enrollment at the moment of the birth. The third channel is voluntary enrollment at the registry’s website. The fourth channel is voluntary enrollment during TwinDay, a particular day dedicated to twins. The last channel is through twin meetings organized by third parties. The relative majority of the pairs comes from Rome, where the registry is based. This highlights the selection through voluntary enrollment. The second largest group is from the province of Milan and the third largest group is from the province of Turin. Overall, at the larger regional level, Northern and Central Italy are overrepresented in the sample.

⁴ The survey asked for both the individual’s actual job and main job of their life. Because of a past Italian pension reform it may be the case that some people said they were retired both as job and main job of their life.

Second, we remove the twins for whom we do not know whether they have ever entered the labor force. Third, we omit the twins whose information is inconsistent (e.g., they declare they entered the labor force before they were born). Eventually, we omit those twins whose self-employment status is trivial. We remove also the other twin in the pair left alone from the data cleaning, if any. Table 1 shows how the final matched sample of 1724 twins (862 pairs) is selected from the total database. They are from 47 provinces⁵ and are between 35 and 65 years old in 2011.

Insert Table 1 about here

3.2 Variables

Dependent Variable

Entrepreneurship. We operationalize entrepreneurship with self-employment. While we are aware of the differences between the self-employed and the entrepreneur (Sørensen & Fassiotta 2011, Levine and Rubinstein 2013; Henrekson and Sanandaji, 2014), self-employment can be considered the simplest kind of entrepreneurship (Blanchflower & Oswald 1998). In addition, self-employment is the standard proxy in the previous literature (Nicolaou et al. 2008, Lindquist et al. 2015), and consistent with it, we will use entrepreneurship and self-employment interchangeably in the paper.

Using self-employment as an operationalization of entrepreneurship offers an opportunity for more conservative estimates. In fact, compared to the high technology industries where we would like to observe high-growth entrepreneurship, self-employment encompasses more industries where women may be more present, e.g., retail (Kalnins & Williams 2014). Therefore, we argue that the difference in the environment between men and women is smaller when we measure entrepreneurship in terms of self-employment.

⁵ There are about 110 provinces in Italy, and they are analogous to counties in the United States.

Table 2 shows the structure of the dependent variable. Entrepreneurs who are working in liberal profession explicitly as entrepreneurs are 7% of the sample, while other types of entrepreneurs are 6% of the sample. Because we selected only those for whom we are certain about their self-employment status, the overall percentage of self-employed circa 13%, lower than the Italian rate of self-employment (circa 20%).

Insert Table 2 about here

Explanatory Variables

Predisposition We exploit differences between identical and fraternal twins to operationalize predisposition. Identical twins are those who develop from one zygote that forms two embryos (they are also known as monozygotic) and they share almost 100% of their genetic endowment. Fraternal twins are those who develop from two eggs, each fertilized by different sperm cells (they are also known as dizygotic) and they share on average 50% of their genetic endowment.

In order to identify the effect of genetics, we rely on the equal environment assumption. The equal environment assumption is common among twin studies and it states that there is no difference in the way the twins are raised by their families⁶ (Scarr & Carter-Saltzman 1979, Evans & Martin 2000, Derks et al. 2006). In this way, we rule out that differences between identical and fraternal twins may be attributable to unobservable factors. In addition to this relatively common assumption, descriptive statistics show no systematic differences between identical and fraternal pairs.

We operationalize the higher share of genetic endowment with an indicator variable that

⁶ Previous literature reports that identical twins are closer to each other than fraternal twins: identical twins raised apart are more similar than fraternal twins raised together. In case of parental misunderstanding of their children's type, it is the actual rather than perceived type which predicts similarities (Scarr & Carter-Saltzman 1979).

takes the value of one if the twins are identical, and zero if they are fraternal. To arrive at our variable of interest, we interact it with a dummy variable assuming the value of one if the other twin is entrepreneur.

Environment We choose to operationalize the environment in the same way as Zhang et al. (2009), i.e., we approximate differences in the environment through gender. In fact, it is quite established in the literature that male and female individuals face different environments for being entrepreneurs. While *ex ante* there are no differences in the propensity of becoming an entrepreneur (Kourilsky & Walstad 1998), women tend to become entrepreneurs to a lower extent than men (Jennings & Brush 2013). Furthermore, women tend to be consistently treated differentially than men, at least based on two mechanisms.

First, entrepreneurship is considered a gendered institution (Greenberg & Mollick 2014). Experimental evidence shows that individuals tend to associate the characteristics of a successful entrepreneur with the characteristics of a male (Buttner & Rosen 1988). Women themselves are affected by this gendered view: controlling for business accomplishment, they tend to identify themselves less as entrepreneurs (Verheul et al. 2005) and they are harsher in their self-assessment (Thébaud 2010).

Second, research shows that women systematically receive fewer resources from different founders. While evidence for business angels (Becker-Blease & Sohl 2007) and venture capitalists (Gompers et al. 2014) can matter little if we study self-employed individuals in Italy, a study about taste discrimination in Italy (Alesina et al. 2013) supports our operationalization. In fact, they find that even if the survival rates for businesses founded by women are higher, women pay higher interest rates to banks with respect to men.

In light of the evidence from the literature, we argue that, *ceteris paribus*, the cost for women of starting a business is higher. Therefore, in order to measure the "friendliness" of the

environment, we use a gender indicator variable, taking the value of one in case of the individual being male, and zero when the individual is a female.

Control Variables We control for the socio-demographic characteristics of the individuals within our sample. To control for age, we create 5-year cohort dummies. In order to control for the level of education, we use two indicator variables: an indicator variable that takes the value of one if the individual's highest degree is high school diploma, and an indicator variable that takes the value of one if the individual's highest degree is at the college level or higher. We also control for geographical location by adding a full set of indicator variables for the province of residence.

4. Method

The ideal experiment The ideal experiment would assess the causal effect of the genetic endowment (G), of an entrepreneurship friendly environment (E), and its interaction ($G * E$) on the probability of becoming an entrepreneur by randomly and independently assigning genes and environments of all possible sorts to individuals.

One should then wait long enough to observe career outcomes. The simple difference between the average of entrepreneurs among individuals with the "entrepreneurship gene" and those without it identifies the causal effect of genes. The difference of the genetic effect between the two types of environment identifies the causal effect of the environment on the genetic effect. The difference of the differences would identify the interaction effect. Needless to say, this experiment would be impossible for at least two reasons. First, we do not know the genes responsible for entrepreneurship. Second, we cannot randomize genetic endowments and force people to stay in a particular environment. Nonetheless, we can try to approximate the ideal experiment.

First, a source of random genetic assignment is found in twins. Identical twins share 100% of

their genetic endowment, while fraternal twins share on average 50% of their genetic endowment. We can assume that the distribution of identical and fraternal twins is as good as random⁷. As a further working assumption, we assume that the populations of twins and non-twins are not systematically different. This assumption is particularly reasonable, as twin studies published across different disciplines rely on the same assumption for external validity.

Second, we exploit the different types of environment men and women experience in entrepreneurship to approximate for environmental friendliness. Gender differences can be a good proxy because conversely from, for example, the province of residence it cannot be changed and is random. We split the sample to observe whether men and women experience gender-specific genetic effects.

Model Because we take advantage of the randomness of the explanatory variables, we do not need sophisticated methodologies to deal with endogeneity. Because the dependent variable is an indicator variable, we use the Linear Probability Model (LPM). We prefer LPM to maximum likelihood methods like Probit because coefficients are easier to interpret – especially the interactions. Moreover, regression results are easier to interpret than structural equation modeling and we can compare the consistency of our results with the earlier findings in the literature⁸. To conclude, we propose the following model:

$$S_{i,c} = \alpha + \beta X_{i,c} + \gamma G_c + \delta S_{-i,c} + \mu G_c S_{-i,c} + \tau_c + \lambda_{i,c} + \varepsilon_{i,c}$$

The unit of observation is a random twin i of a pair c . The dependent variable $S_{i,c}$ indicates

⁷ Although it is common knowledge that assisted reproduction leads with a higher probability to (fraternal as well as identical) twins, we are aware that assisted reproduction increases the ratio of MZ/DZ (Schachter et al. 2001), but we assume it is a fringe phenomenon in our study. In fact, assisted reproduction births in Italy are the lowest in Europe and they account for as little as 1.4% of total births (Kocourkova et al. 2014)

⁸ As properly described in Nicolaou & Shane (2010), we are aware that by using regression instead of structural equation modeling we are testing, rather than latent heritability, genetic etiology – in lay terms, finding genes as causes of a phenomenon.

whether the twin i is an entrepreneur. X is a set of control variables such as gender and education. The indicator variable G_c tells whether the pair is of identical or fraternal twins. $S_{-i,c}$ is an indicator variable for whether the other twin is entrepreneur. The interaction term defines the coefficient of interest for the existence of a genetic effect. Given that the other twin is entrepreneur, the coefficient identifies how much more likely the individual is to become entrepreneur if they share 100% of their genetic endowment.

To study how genes and the environment interact, we perform the regression in equation 1 on two subsamples of males and females. We can compare the difference between the coefficients and observe for which group the effect is larger. We believe that split samples are a more parsimonious methodology to test our hypothesis. As an alternative strategy, we could interact gender with the genetic effect, but the genetic effect is operationalized as an interaction *per se* and triple interactions would be hard to interpret.

Balance issues. In this paragraph, we highlight some potential sources of bias from imbalance and their potential solutions. For some reason, identical twins might have observable factors that correlate with entrepreneurship. We mitigate this concern by showing that identical and fraternal couples show little or no difference in the observables through a t-test in panel A of Table 3. There is no statistically significant difference among observables between groups.

We also test the balance between male and female individuals because we would not like to have systematic differences between men and women when we test the differences in gender. We report the mean differences in panel b of table 3. We observe that there is more imbalance looking at the differences between gender rather than type of twins. Females are systematically older and they have systematically more college education. In order to mitigate the imbalance in education, we will perform coarsened exact matching (Iacus et al.

2011).

Coarsened exact matching (CEM) is a nonparametric technique that allows matching on blocks, and it has been found to perform better than propensity score matching (Iacus et al. 2011). In our case, we prefer CEM to exact matching to retain more observation when we match continuous variables such as age. In practice, we are looking within our sample for observations that can approximate the counterfactual of our male units. In our case, we are looking for individuals who have same age, they are the same type of twins, they have the same education, the same experience, and they live in the same province.

The validity of the coarsened exact matching relies on the conditional independence assumption: conditional on the observables, the gender assignment is random, which is reasonable. Because we are looking for women with men-like features and vice versa, the common support would give us a lower bound estimate of the different genetic effect. Looking at the matched sample of panel C in Table 3, the unbalance disappears for all the variables.

Insert Table 3 about here

5. Results

5.1 Descriptive statistics

Table 4 summarizes the sample's observables. In the sample, 13% of the individuals have been or were entrepreneurs at the time of inclusion in the registry. This figure is lower than 20%-25%, the average rate of self-employment in Italy, and it represents a lower bound figure. The sample includes more identical than fraternal twins (65% versus 35%) because we selected fraternal twins of the same sex only. Men are around 35.5% of the sample, which is not different from the 41% of males of comparable cohorts in the entire ITR in 2013 (Fagnani

et al., 2014); the average age is around 48 years old; 49% of the individuals studied up to high school, while 27% of them obtained a college degree or higher.

Pairwise correlation informs us that there is not any strong significant correlation of zygosity with the other observables – in line with its random distribution. In addition, male individuals are less likely to have obtained a higher education degree, while they are more likely to be entrepreneurs, as expected.

Insert Table 4 about here

5.2 Estimation of the genetic effect

To measure the genetic effect, we estimate the difference in the likelihood of being an entrepreneur yourself between individuals having an identical twin who is an entrepreneur and having a fraternal twin who is an entrepreneur. In the models, we estimate it using the main effects (being identical twins, the other twin is entrepreneur) and their interaction – the “genetic effect”. Table 5 shows the results. Model 1 estimates the genetic effect without any control, model 2 adds socio-demographic controls for age cohort, gender and education indicator variables, finally model 3 adds indicator variables for province of residence. We estimate the results selecting one of the twins of each pair randomly: twin 1 (models 1.1 to 3.1) and twin 2 (models 1.2 to 3.2).

We will first discuss the two main effects and then the interaction, our object of interest. We observe that the coefficient of being identical twins by itself is not significant and its size is close to zero. This is reassuring, as there is no reason to assume that being identical twins leads into entrepreneurship (or vice versa). The size and the significance of the coefficient are consistent across all the models. This zero coefficient lends support to the validity of our approach.

The coefficient of the “other twin is entrepreneur” is positive and significant. If the other twin is entrepreneur, it is 39.2% more likely that the individual herself is an entrepreneur too. This result can be interpreted as the effect of shared environment among twins. Shared environment could include mechanisms such as peer effects (Sørensen 2007; Nanda & Sørensen 2010; Lerner & Malmendier 2013), family background (Bjorklund *et al* 2009; Nicoletti & Rabe 2013), and neighborhood (Giannetti & Simonov 2009; Nicoletti & Rabe 2013).

It is the interaction between being identical twins and the other twin entrepreneur that allows us to estimate the genetic effect. We observe that this coefficient is positive and significant: it is 19.6% more likely for a twin to be an entrepreneur when the other twin is also entrepreneur if the pair is of identical twins – i.e., they share 100% of their genes. The coefficients are statistically significant at the 1% level across all twins and models. Interestingly, the size of the coefficients for the “genetic effect” and the “shared environment” have the same ratio of comparable studies of nature and nurture, namely 20% and 40% (Lindquist et al. 2015).

We run a series of test to look at the robustness of results. In model 2 we control for potential drivers of the entrepreneurial choice such as gender, age, and education. The size and the significance of the coefficients of interest do not substantially change. In addition, we observe that gender plays a role consistently with the literature. Education plays apparently no role: compared to middle school education, coefficients for high school and college education are small and not significant. To control for age, we added 5-years cohort fixed effects and none of the coefficient was statistically significant from zero (coefficients not reported in the table). In model 3 we added indicator variables for province of residence to control for geography. The size of the coefficients diminishes not sizably and the qualitative evidence does not change.

We also ran a test for nested models to choose the most parsimonious way to add control variables and the F-statistic recommends to exclude the socio-demographic variables ($F(8, 849)=1.51, p=0.14$ for twin 1; $F(8, 849)=1.54, p=0.14$ for twin 2) and keep the province fixed effects ($F(29, 809)=169.11, p=0.00$ for twin 1; $F(29, 808)=110.7, p=0.00$ for twin 2) for both twins.

We also addressed some other concerns related to selection. One could argue that the selection of the random twin of the pair could potentially drive the results. To this extent, we ran the three models for both groups of twins and we report the results in tables 1.2 to 3.2⁹. Descriptively, the only microscopic difference for twin 2 lies in the significance and the size of the variable for gender. The size of the coefficient is twice as large for twin 1 but the difference between coefficients is not statistically significant ($\chi^2_{(1)}=1.14; p = 0.285$). Finally, we tested the significance of the difference between the genetic effects and there is no statistically significance across the three specifications: without controls ($\chi^2_{(1)}=0.01, p=0.91$), controlling for socio-demographic variables ($\chi^2_{(1)}=0.01, p=0.93$), and controlling for province of residence ($\chi^2_{(1)}=0.01, p=0.94$).

To conclude, we seem to have found evidence of a genetic effect between 16.95 and 19.6%, which is consistent across models and twins, in support of the extant literature and Hypothesis 1.

Insert Table 5 about here

5.3 The interaction with environment

5.3.1 Gender specific genetic effect

In Table 6, we study whether the genetic effect is different across genders. In the table, we

⁹ We continue this exercise for the rest of our analysis. However, because our results do not change systematically and there are not statistically significance differences, we report the results at the end of the paper in the appendix.

report only the explanatory variables of interest: the genetic effect, the indicator variable for identical twins, and the indicator variable for other twin entrepreneur, the “shared environment”. In model 1 we include only the three key variables, in model 2 we match the sample according to gender using coarsened exact matching (Iacus et al. 2011) as described above. In model 3 and 4 we control for province indicator variables according to the results of the tests in the previous section.

For the coefficient pertaining to the control for the shared environment, we observe a difference between men and women. The size of the coefficient for men is almost half of the coefficient for women, but the difference is not statistically significant ($\chi^2_{(1)}=1.75$, $p=0.19$).

We now turn to the analysis of the genetic effect and we observe the opposite pattern: the genetic effect is large, positive, and significant for men, while for women it is negative, small, and insignificant. For men, sharing 50% more genes with respect to a sibling who is entrepreneur increases the probability to be entrepreneur of 40%. For women, it should insignificantly reduce the probability of 2%. The Wald tests confirm the significance of the gender-difference between twins.

Looking at the mean differences across gender in Table 4 we notice that the two groups are not balanced for what concerns age and college education. When we compare the two subsamples, it would be desirable to have two groups that are balanced as much as possible. Therefore, we perform coarsened exact matching (Iacus et al. 2011), a nonparametric methodology that allows for block matching. In this way, we try to have two subsamples of people who have the same socio-demographic characteristics and they differ only in terms of gender.

We believe that the specification allows us to obtain a lower-bound estimate of the interaction

with the environment as it will capture a more nuanced gender difference of “man-like women” and “woman-like men”. We match on age cohort and college education with a k2k procedure, i.e., we assign only one female to each male observation. We report the results in model 2 of Table 6.

Model 1 shows a small difference: the size of the effect of shared environment drops for males and increases for females, while the significance for men drops. The statistical significance of the difference in coefficients for shared environment is marginally significant ($\chi^2_{(1)}=2.71$, $p=0.099$). For what concerns the genetic effect, it seems that the matching procedure confirms the results of the analysis in model 1 with the same levels of significance for the coefficients. The magnitude of the coefficients increase for both men and women in opposite directions. For men, the genetic effect in the matched sample amounts to an increase in the probability of 49.7%. In addition, from the analysis of the Wald test, it seems that the significance of the difference in the genetic effect is sharper with the matched sample ($p=0.05$ in model 1 versus $p=0.004$ in model 2).

In models 3 and 4 we control for the province of residence both in the entire sample and in the matched sample. The coefficients seem not to vary sizably. Thus, it seems that genetics matters more for men than women, and that it is only when the environment is favorable to entrepreneurship that genes play a role. Moreover, these results are robust to matched samples where the gender difference is sharper. These results go in support of hypothesis 2b: in favorable environments, predisposition matters more.

Insert Table 6 about here

5.3.2 Alternative proxy for the environment

In this section, we address some of the concerns that can be related to the choice of gender as approximation of the environment. We chose gender as proxy because there is an established

literature about the gender gap and it has the favorable feature of not being possible to change it (at reasonable costs). However, one can argue that gender is only an incomplete proxy of the environment or that it encompasses something more other than the difference in the institutional environment.

As a robustness check, we look at the province of residence to try to capture the differences in environment. We chose this approach as a robustness test rather than for the main analysis because it suffers of the possibility of individuals to select into the province that fits them best. However, the literature documents that in Italy, entrepreneurs tend to stabilize in their region and they tend at higher rates than employees (Michelacci & Silva 2007). We select the largest provinces in our sample, Milan and Rome.

Rome and Milan are respectively the first and the second largest city of Italy, and they represent the two different sides of Italy. While Rome hosts the vast majority of the Italian government and state-related organizations, Milan hosts a wide array of industries like biotechnology, fashion, and it is home of more than 8 percent of all businesses in Italy, including three Fortune 500 companies. In addition, relatively to Rome, Milan is a better environment for the “creative class” as it has higher levels of talent, technology, and tolerance (Florida 2005). While with no doubt Rome could represent a better environment compared to a mountain area in the North, we look only at the differences between Milan and Rome. Thus, we operationalize Milan as a “favorable” environment compared to Rome.

Table 7 shows the results of this different specification. Model 1 is the baseline model and model 2 controls for socio-demographic characteristics (obviously we cannot control for province of residence as it is our source of variation). Qualitatively, the effects do not change: the genetic effect is positive, large, and statistically significant for the twins living in Milan, with a premium of 68%, which becomes 69% after controlling for socio-demographic

characteristics. For the twins living in Rome, the genetic is small, positive and insignificant. These results are robust even after controlling for age, gender and education and the difference in the coefficients is statistically significant according to the Wald test.

These estimates confirm that the differential genetic effect is not a phenomenon that is gender-specific, but it relates to an environment component. Thus, we argue that there is more evidence in support of hypothesis 2b.

Insert Table 7 about here

5.3.3 Alternative unit of analysis.

In this section, we provide a further robustness check for our study. In fact, given our specification, one may argue that simultaneity problems may still be present even after choosing one random twin. In order to alleviate this concern, we estimate the model at the pair level. There, our outcome variable of interest becomes an indicator variable that takes the value of one if both twins are entrepreneurs and zero otherwise. In this case, our proxy of the genetic effect becomes the simple indicator variable for identical twins. As a control, we include an indicator variable for both twins doing the same job. We report the results in Table 8. Model 1 is a pooled estimation, while models 2 and 3 are split respectively between gender groups and provinces of residence.

In the pooled sample, it is 3.5% more frequent to observe pairs of twins entrepreneurs if they share 100% of their genes. However, when we split the sample according to the environment, the results change according to our previous evidence. The effect for men is larger: it is 10.2% more likely to observe pairs of identical twins that are entrepreneurs, while for women there is no significant difference. The same results hold with the difference in provinces: it is more likely to observe identical twins being both entrepreneurs in Milan, while it is not the case in Rome. The differences in the coefficients are statistically significant, in further

support of hypothesis 2b.

Insert Table 8 about here

6. Limitations

While the data allow us to analyze the phenomenon in a precise and novel way, it exposes our analysis to two main limitations: the use of self-employment as proxy for entrepreneurship, and the cross-sectional nature of the dataset.

Self-employment as a proxy for entrepreneurship overcomes two difficulties of quantitative studies (Sørensen & Fassiotta 2011; Henrekson & Sanandaji 2014): the very small number of successful new and innovative ventures, and the impossibility to find the type of venture from archival data during early stages. Despite the fact that it may be considered “the most basic form of entrepreneurship” (Blanchflower & Oswald 1998), self-employment may not be relevant for policy purposes. In fact, we are able to measure the extensive margin of entrepreneurship but we cannot say anything about the intensive margin. With self-employment, we consider the laundry shop and the venture-backed biotech company as identical. Self-employment is a variable of interest if the policymaker wants to reduce unemployment through small businesses. It becomes inadequate to the policymaker who wants to spur growth through innovation and technology from startups.

The second data limitation is related to its cross-sectional nature.

Therefore, we cannot use panel data techniques. A panel data technique like fixed-effects would be of little use to identify the genetic effect as it does not change over time, but it would provide more effective settings for the identification of the role of environment. For example, we could observe the selection of one or both the twins in another environment or we could test the effects of policy changes in some geographical areas or for some categories

of people (e.g., young individuals or women).

In addition, longitudinal data would help us to understand not only the extensive but also the intensive margin. Our paper tested whether predisposition matters and in which environments for being an entrepreneur, and further research could explore whether genes play a role for survival, entrepreneurial income or incorporation. To conclude, a panel dataset of twins with a richer set of variables, especially those related to entrepreneurship, can help to overcome this limitation.

7. Conclusion

We study the phenomenon of the micro-foundation of entrepreneurship through a twin study. Our research addresses how the genetic effect -the predisposition - interacts with the environment. We adopt an agnostic perspective: *ex ante*, we do not know whether the interaction between predisposition and environment exists and which direction it should go into. Consistent with the earlier literature (Zhang et al. 2009, Nicolaou & Shane 2010) we approximate the environment with gender differences. Through a unique dataset of Italian twins and using regression analysis, we found evidence of a genetic effect that is consistent with the literature (Nicolaou et al. 2008; Lindquist et al. 2015).

Our core result shows that the interaction between the genetic effect and predisposition to selection into entrepreneurship is positive. The evidence is robust to a more nuanced gender difference using matching. In addition, we provide an alternative operationalization of the environment with the comparison between twins living in Milan (favorable) and Rome (unfavorable), and we perform a set of regressions using the twin pair as the unit of analysis. All this evidence goes in the same direction.

We contribute to the literature by estimating the direction of the interaction between the genetic effect and environment. Predisposition matters more when the cost of

entrepreneurship is lower. While earlier studies provided a single piece of evidence in contrasting directions (Zhang et al. 2009, Nicolaou & Shane 2010), we provide evidence of positive interaction in our dataset using two different proxies of the environment (gender and province of residence) and two different units of analysis (single twin and pair).

In this paragraph, we would like to highlight what our study does not say. For example, it does not say that only genes matter in the entrepreneurial choice. Moreover, we are not claiming that policy or institutions are useless to the development of an entrepreneurship policy. In fact, our results suggest that genetic endowment matters in entrepreneurship only in an environment that is favorable to it. In addition, we observed that without the key interaction with the environment, genes explain nothing.

Thus, we argue the contrary: genes matter only when the environmental conditions are favorable. An environment that is supportive of entrepreneurship is simply more efficient in sorting the individuals with more predisposition into what they are called for. For people with little inclination towards entrepreneurship, such an environment is desirable too. In fact, for those whose predisposition is not entrepreneurship, it is more efficient to start a career as employee elsewhere. This finding finds accordance with the earlier literature about the reward of talent: historically, there were environments where entrepreneurship was not rewarded and the talented people selected into other careers like military or political career (Baumol, 1990).

With no doubts, the interaction between predisposition and environment requires further investigation. We provided a static analysis on the extensive margin. In our dataset, the main variable that approximates the environment, gender, cannot change¹⁰. Further research could focus on programs or policy changes. With our study, we studied the interaction between

¹⁰ Even if there is evidence that when it does it leads to better outcomes (Schilt & Wiswall 2008).

predisposition and an environment looking mainly at the cost of entry into entrepreneurship. Further studies can focus on the interaction between predisposition and environments where other parameters of the entrepreneurial choice may change, for example the level of opportunities, the probability of success, or the alternative choice – a different type of interaction would not be implausible. Another possible extension would be to study the intensive margin of entrepreneurial performance.

To conclude, we call for more research on this type of studies. On the one hand, policy makers who want to promote diffusion of small businesses to contrast unemployment may be interested in abating the costs of entrepreneurship to allow people with less predisposition to enter entrepreneurship – an inclusive policy. On the other hand, policy makers may want to select the best individuals to launch and found the high-growth companies that create employment and innovation. To do so, they would be interested to provide resources to individuals who already have a predisposition to it, because predisposition and environment are complements. We believe that studies that analyze the interaction between predisposition and environment would be a useful tool for the evaluation of entrepreneurship policy.

References

- Aldrich, H. 1999. *Organizations evolving*, Sage.
- Aldrich, H. E., & Fiol, C. M. 1994. 'Fools rush in? The institutional context of industry creation.' *Academy of management review*, 19(4), 645-670.
- Alesina, A. F., Lotti, F. & Mistrulli, P. E. 2013. 'Do women pay more for credit? evidence from Italy', *Journal of the European Economic Association* 11(s1), 45-66.
- Baumol, W. J. 1990. 'Entrepreneurship: Productive, unproductive, and destructive', *The Journal of Political Economy* 98(5 Part 1), 893-921.
- Becker-Blease, J. R. & Sohl, J. E. 2007. 'Do women-owned businesses have equal access to angel capital?', *Journal of Business Venturing* 22(4), 503-521.
- Björklund, A., Jäntti, M., & 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), 671-680.
- Blanchflower, D. G. & Oswald, A. J. 1998. 'What makes an entrepreneur?', *Journal of Labor Economics* 16(1).
- Brescianini, S., Fagnani, C., Toccaceli, V., Medda, E., Nisticò, L., D'Ippolito, C., ... & Stazi, M. A. 2013. An update on the Italian Twin Register: advances in cohort recruitment, project building and network development. *Twin Research and Human Genetics*, 16(01), 190-196.
- Buttner, E. H. & Rosen, B. 1988. 'Bank loan officers' perceptions of the characteristics of men, women, and successful entrepreneurs', *Journal of Business Venturing* 3(3), 249-258.
- Chinitz, B. 1961. Contrasts in agglomeration: New york and pittsburgh. *The American Economic Review*, 279-289.
- Derks, E. M., Dolan, C. V. & Boomsma, D. I. 2006. 'A test of the equal environment assumption (eea) in multivariate twin studies', *Twin Research and Human Genetics* 9(03), 403-411.
- Evans, D. M. & Martin, N. G. 2000. 'The validity of twin studies', *GeneScreen* 1(2), 77-79.
- Fagnani, C., Brescianini, S., Cotichini, S., D'Ippolito, C., Dukic, T., Giannantonio, L., ... & Stazi, M. A. 2006. The Italian Twin Register: new cohorts and tools, current projects and future perspectives of a developing resource. *Twin research and human genetics*, 9(06), 799-805.
- Fagnani, C., Medda, E., Stazi, M. A., Caprara, G. V., & Alessandri, G. 2014. Investigation of age and gender effects on positive orientation in Italian twins. *International Journal of Psychology*, 49(6), 453-461.
- Florida, R. 2005. *Cities and the creative class*. Routledge.
- Fraga, M.F., Ballestar, E., Paz, M.F., Roperio, S., Setien, F., Ballestar, M.L., Heine-

Suñer, D., Cigudosa, J. C., Urioste, M., Benitez, J. et al. 2005. 'Epigenetic differences arise during the lifetime of monozygotic twins', *Proceedings of the National Academy of Sciences of the United States of America* 102(30), 10604–10609.

Gartner, W. B. 1988. 'Who is an entrepreneur? is the wrong question', *American Journal of Small Business* 12(4), 11–32.

Giannetti, M. & Simonov, A. 2009. 'Social interactions and entrepreneurial activity', *Journal of Economics & Management Strategy* 18(3), 665–709.

Glaeser, E. L., Kerr, S. P., & Kerr, W. R. 2015. Entrepreneurship and urban growth: An empirical assessment with historical mines. *Review of Economics and Statistics*, 97(2), 498–520.

Gompers, P. A., Mukharlyamov, V., Weisburst, E. & Xuan, Y. 2014. '*Gender effects in venture capital*', Available at SSRN 2445497 .

Greco, L., Romino, R., Coto, I., Di Cosmo, N., Percopo, S., Maglio, M., ... & Stazi, M. A. 2002. The first large population based twin study of coeliac disease. *Gut*, 50(5), 624–628.

Greenberg, J. & Mollick, E. R. 2014. 'Leaning in or leaning on? gender, homophily, and activism in crowdfunding', *Unpublished working paper* .

Guiso, L. & Rustichini, A. 2011. '*What drives women out of entrepreneurship? The joint role of testosterone and culture*'. *Unpublished working paper*.

Gupta, V. K., Goktan, A. B. & Gunay, G. 2014. 'Gender differences in evaluation of new business opportunity: A stereotype threat perspective', *Journal of Business Venturing* 29(2), 273–288.

Henrekson, M., & Sanandaji, T. 2014. Small business activity does not measure entrepreneurship. *Proceedings of the National Academy of Sciences*, 111(5), 1760–1765.

Iacus, S. M., King, G. & Porro, G. 2011. 'Causal inference without balance checking: Coarsened exact matching', *Political Analysis* p. mpr013.

Jennings, J. E., & Brush, C. G. 2013. Research on women entrepreneurs: challenges to (and from) the broader entrepreneurship literature?. *The Academy of Management Annals*, 7(1), 663–715.

Kacperczyk, A. J. 2013. Social influence and entrepreneurship: The effect of university peers on entrepreneurial entry. *Organization Science*, 24(3), 664–683.

Kalnins, A. & Williams, M. 2014. 'When do female-owned businesses out-survive male-owned businesses? a disaggregated approach by industry and geography', *Journal of Business Venturing* 29(6), 822–835.

King, R. G., & Levine, R. 1993. Finance, entrepreneurship and growth. *Journal of Monetary Economics*, 32(3), 513–542.

Kirzner, I. M. 1978. *Competition and entrepreneurship*, University of Chicago press.

- Kocourkova, J., Burcin, B. & Kucera, T. 2014. 'Demographic relevancy of increased use of assisted reproduction in european countries', *Reproductive health* 11(1), 37.
- Kourilsky, M. L. & Walstad, W. B. 1998. 'Entrepreneurship and female youth: Knowledge, attitudes, gender differences, and educational practices', *Journal of Business Venturing* 13(1), 77–88.
- Laureiro-Martínez, D., Canessa, N., Brusoni, S., Zollo, M., Hare, T., Alemanno, F. & Cappa, S. F. 2013. 'Frontopolar cortex and decision-making efficiency: comparing brain activity of experts with different professional background during an exploration-exploitation task', *Frontiers in Human Neuroscience* 7.
- Lerner, J., & Malmendier, U. 2013. With a little help from my (random) friends: Success and failure in post-business school entrepreneurship. *Review of Financial Studies*, hht024.
- Levine, R., & Rubinstein, Y. 2013. *Smart and Illicit: Who Becomes an Entrepreneur and Does it Pay?* (No. w19276). National Bureau of Economic Research.
- Lindquist, M. J., Sol, J. & Praag, M. V. 2015. 'Why do entrepreneurial parents have entrepreneurial children?', *Journal of Labor Economics* 33(2).
- Martini, M., Bufalari, I., Stazi, M. A., & Aglioti, S. M. 2015. Is That Me or My Twin? Lack of Self-Face Recognition Advantage in Identical Twins. *PLoS ONE*,10(4).
- Michelacci, C., & Silva, O. 2007. Why so many local entrepreneurs?. *The Review of Economics and Statistics*, 89(4), 615-633.
- Murphy, K. M., Shleifer, A. & Vishny, R. W. 1991. 'The allocation of talent: implications for growth', *The Quarterly Journal of Economics* 106(2), 503–530.
- Nanda, R. & Sørensen, J. B. 2010. 'Workplace peers and entrepreneurship', *Management Science* 56(7), 1116–1126.
- Nicolaou, N. & Shane, S. 2009. 'Can genetic factors influence the likelihood of engaging in entrepreneurial activity?', *Journal of Business Venturing* 24(1), 1–22.
- Nicolaou, N. & 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. & Spector, T. D. 2008. 'Is the tendency to engage in entrepreneurship genetic?', *Management Science* 54(1), 167–179.
- Nicoletti, C., & Rabe, B. 2013. Inequality in Pupils' Test Scores: How Much do Family, Sibling Type and Neighbourhood Matter?. *Economica*, 80(318), 197-218.
- Scarr, S. & Carter-Saltzman, L. 1979. 'Twin method: Defense of a critical assumption', *Behavior Genetics* 9(6), 527–542.
- Schachter, M., Raziell, A., Friedler, S., Strassburger, D., Bern, O. & Ron-El, R. 2001. 'Monozygotic twinning after assisted reproductive techniques: a phenomenon independent of micromanipulation', *Human Reproduction* 16(6), 1264–1269.

- Schilt, K. & Wiswall, M. 2008. 'Before and after: Gender transitions, human capital, and workplace experiences', *The BE Journal of Economic Analysis & Policy* 8(1).
- Sørensen, J. B. 2007. 'Bureaucracy and entrepreneurship: Workplace effects on entrepreneurial entry', *Administrative Science Quarterly* 52(3), 387–412.
- Sørensen, J. B. & Fassiotta, M. A. 2011. 'Organizations as fonts of entrepreneurship', *Organization Science* 22(5), 1322–1331.
- Sørensen, O., & Audia, P. G. 2000. The social structure of entrepreneurial activity: geographic concentration of footwear production in the United States, 1940–1989. *American Journal of Sociology*, 106(2), 424-462.
- Stazi, M. A., Cotichini, R., Patriarca, V., Brescianini, S., Fagnani, C., D'Ippolito, C., Cannoni, S., Ristori, G. & Salvetti, M. 2002. 'The Italian twin project: from the personal identification number to a national twin registry', *Twin Research* 5(5), 382–386.
- Stuart, T., & Sorenson, O. 2003. 'The geography of opportunity: spatial heterogeneity in founding rates and the performance of biotechnology firms.' *Research policy*, 32(2), 229-253.
- Thébaud, S. 2010. 'Gender and entrepreneurship as a career choice do self-assessments of ability matter?', *Social Psychology Quarterly* 73(3), 288–304.
- Thornton, P. H. 1999. 'The sociology of entrepreneurship', *Annual review of sociology* 25(1), 19–46.
- Verheul, I., Uhlaner, L. & Thurik, R. 2005. 'Business accomplishments, gender and entrepreneurial self-image', *Journal of Business Venturing* 20(4), 483–518.
- Wennekers, S., & Thurik, R. 1999. Linking entrepreneurship and economic growth. *Small business economics*, 13(1), 27-56.
- Zhang, Z., Zyphur, M. J., Narayanan, J., Arvey, R. D., Chaturvedi, S., Avolio, B. J., Lichtenstein, P. & Larsson, G. 2009. 'The genetic basis of entrepreneurship: Effects of gender and personality', *Organizational Behavior and Human Decision Processes* 110(2), 93–107.

Tables

Table 1. Twin data extraction and cleaning.

Sample	N	%
Born b/w 1946 and 1976	2928	100%
<i>Pairs of DZ twins of different sex</i>	-634	-22%
<i>Out of labor force</i>	-89	-3%
<i>Inconsistent data</i>	-16	-1%
<i>“Orphan” twins</i>	-73	-7%
<i>Trivial self-employment status</i>	-231	-8%
<i>New “Orphans”</i>	-161	-5%
Final sample	1724	59%
<i>CEM matching at age, education, zygosity and province level</i>	-820	-28%
Matched sample	904	31%

Table 2. Structure of the outcome variable.

Type	N	%
Employee	1503	87.06
Self-Employed	217	12.94
	<i>Of whom:</i>	
Liberal professionals explicitly entrepreneurs	120	6.96
Other Entrepreneurs	103	5.98

Table 3. T-test table

<i>Panel a.</i> Differences across fraternal (DZ) and identical (MZ) twins			
Variable	DZ	MZ	Difference
Entrepreneur	0.112	0.139	-0.027
Male	0.362	0.351	0.011
Age	48.493	47.823	0.670
College degree and higher	0.281	0.272	0.009
High School degree	0.467	0.504	-0.037
<i>N</i>	608	1116	
<i>Panel b.</i> Differences across male and female twins – entire sample			
Variable	Female	Male	Difference
Entrepreneur	0.091	0.199	0.109**
MZ	0.651	0.641	-0.011
Age	47.773	48.578	0.805 ⁺
College degree and higher	0.298	0.234	-0.064**
High School degree	0.484	0.503	0.020
<i>N</i>	1112	612	
<i>Panel c.</i> Differences across male and female twins – matched sample			
Variable	Female	Male	Difference
Entrepreneur	0.091	0.189	0.098**
Identical twin pair	0.670	0.670	0.000
Age	48.615	48.543	-0.071
College degree and higher	0.281	0.236	-0.045
High School degree	0.483	0.488	0.005
<i>N</i>	452	452	

Notes: Matching procedure is coarsened exact matching (Iacus et al. 2011) with 1:1 assignment based on type of twin, age, college degree, high school degree, and province of residence. The matching variable is “male”. Significance levels. ⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4. Descriptive Statistics

	Mean	S.D.	(1)	(2)	(3)	(4)	(5)	(6)
(1) Entrepreneur	0.129	0.336	1					
(2) Identical twin	0.647	0.478	0.039	1				
(3) Male	0.355	0.479	0.155**	-0.011	1			
(4) High School	0.491	0.500	0.009	0.035	0.019	1		
(5) College	0.275	0.447	-0.013	-0.010	-0.069*	-0.604**	1	
(6) Age	48.06	9.071	-0.066*	-0.035	0.043	-0.039	-0.176**	1
N	1720							

Notes: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5. Estimation of the genetic effect.

	(1.1)	(2.1)	(3.1)	(1.2)	(2.2)	(3.2)
	Baseline	Twin 1 Controls	Province	Baseline	Twin 2 Controls	Province
Genetic effect (GS)	0.196** (0.067)	0.189** (0.067)	0.180* (0.070)	0.180** (0.060)	0.177** (0.060)	0.169** (0.061)
Shared Environment (S)	0.392*** (0.055)	0.382*** (0.055)	0.390*** (0.058)	0.353*** (0.050)	0.344*** (0.050)	0.338*** (0.050)
Identical twin pair (G)	-0.007 (0.023)	-0.006 (0.023)	-0.009 (0.0230)	-0.011 (0.022)	-0.012 (0.022)	-0.018 (0.022)
Male		0.0747*** (0.021)	0.0731*** (0.022)		0.0360+ (0.021)	0.0387+ (0.021)
High school degree		-0.0263 (0.026)	-0.0263 (0.027)		0.0305 (0.025)	0.0275 (0.026)
College Degree		0.00278 (0.029)	0.00232 (0.030)		0.00355 (0.028)	0.00893 (0.029)
Cohort FE	N	Y	Y	N	Y	Y
Province FE	N	N	Y	N	N	Y
R^2	0.260	0.272	0.301	0.260	0.270	0.324
N	862	862	861	862	862	859
<i>Wald Test</i>	χ^2	<i>p-value</i>				
$GS(1.1)=GS(1.2)$	0.01	0.91				
$GS(2.1)=GS(2.2)$	0.01	0.93				
$GS(3.1)=GS(3.2)$	0.01	0.94				

Notes: Robust standard errors in parentheses. Dependent variable is: “individual i is entrepreneur” Genetic Effect is the interaction between “Other twin entrepreneur” and “Identical twin pair”. Cohort FE are a full set of 5 birth-year indicator variables. Province FE are a set of indicators variables per province of residence.

Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6. Gender-specific genetic effect.

	(1.m)	(1.f)	(2.m)	(2.f)	(3.m)	(3.f)	(4.m)	(4.f)
	Baseline				Controls			
	Entire sample		Matched Sample		Entire Sample		Matched Sample	
	Male	Female	Male	Female	Male	Female	Male	Female
Genetic effect (GS)	0.407*** (0.114)	-0.016 (0.081)	0.497*** (0.144)	-0.130 (0.137)	0.429*** (0.121)	-0.049 (0.083)	0.475** (0.149)	-0.127 (0.143)
Shared environment (S)	0.253** (0.097)	0.490*** (0.066)	0.131 (0.122)	0.571*** (0.117)	0.217* (0.101)	0.542*** (0.070)	0.149 (0.125)	0.561*** (0.119)
Identical twin pair (G)	-0.0506 (0.045)	0.0180 (0.024)	-0.056 (0.055)	0.030 (0.037)	-0.053 (0.047)	0.0155 (0.025)	-0.068 (0.057)	0.025 (0.039)
R^2	0.292	0.219	0.237	0.222	0.348	0.279	0.295	0.269
N	306	556	226	226	306	555	226	226
<i>Wald Test</i>	χ^2	<i>p-value</i>						
GS(1.m)=GS(2.f)	3.95	0.05*						
GS(2.m)=GS(2.f)	8.04	0.004**						
GS(3.m)=GS(3.f)	5.23	0.02*						
GS(4.m)=GS(4.f)	3.70	0.05 ⁺						

Notes: Robust standard errors in parentheses. Estimation results for twin 1. Dependent variable is: “individual i is entrepreneur” Genetic Effect is the interaction between “Other twin entrepreneur” and “Identical twin pair”. Controls are a set of indicators variables per province of residence. Significance levels: ⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7. Robustness check. Alternative operationalization

	(1.Mi)		(1.Rm)		(2.Mi)		(2.Rm)	
	Baseline				Controls			
	Milan	Rome			Milan	Rome		
Genetic effect (GS)	0.678 ⁺ (0.324)	0.063 (0.0997)			0.688 [*] (0.334)	0.074 (0.100)		
Shared environment (S)	-0.065 (0.311)	0.371 ^{***} (0.0797)			-0.119 (0.319)	0.355 ^{***} (0.081)		
Identical twin pair (G)	0.036 (0.0586)	-0.003 (0.0343)			0.024 (0.059)	-0.004 (0.035)		
Male					0.096 (0.063)	0.055 (0.034)		
High School degree					0.073 (0.077)	-0.025 (0.040)		
College degree					0.046 (0.085)	0.008 (0.044)		
R^2	0.294	0.209	0.348		0.231	0.294		
N	130	283	130		283	130		
<i>Wald Test</i>	χ^2	<i>p-value</i>			χ^2	<i>p-value</i>		
GS(Mi)=GS(Rm)	7.55	0.006 ^{**}			6.99	0.008 [*]		

Notes: Robust standard errors in parentheses. Estimations for twin 1. Dependent variable is: “individual i is entrepreneur”. Genetic Effect is the interaction between “Other twin entrepreneur” and “Identical twin pair”. Significance levels: ⁺ $p < 0.10$, ^{*} $p < 0.05$, ^{**} $p < 0.01$, ^{***} $p < 0.001$

Table 8. Robustness check. Alternative unit of analysis: pair level.

	(1)	(2.m)	(2.f)	(3.Mi)	(3.Rm)
	Pooled	Environment as gender		Alternative environment	
		Males Only	Females Only	Milan	Rome
Identical twins pair (G)	0.0349* (0.0172)	0.102** (0.038)	-0.001 (0.019)	0.123* (0.048)	0.002 (0.029)
Twins do the same job	0.044** (0.017)	0.072+ (0.039)	0.0362+ (0.020)	0.065 (0.050)	0.0476 (0.029)
R^2	0.010	0.034	0.006	0.102	0.018
N	862	306	556	127	278
<i>Wald Test</i>		χ^2	<i>p-value</i>	χ^2	<i>p-value</i>
$G(fav)=G(unfav)$		7.18	0.007**	7.02	0.008**

Notes. Robust Standard error in parenthesis. Unit of analysis is pair level. Outcome variable is “both the twins are entrepreneurs”. Models 2.Mi and 2.Rm control for “male pair”.

Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A1. Robustness test for the alternative twin in the pair.

	(1.m)	(1.f)	(2.m)	(2.f)	(3.m)	(3.f)	(4.m)	(4.f)	(5.MI)	(5.RM)	(6.MI)	(6.RM)
	Twin 2											
	Entire Sample				Matched Sample				Alternative EV			
	Baseline		Controls		Baseline		Controls		Baseline		Controls	
	Male	Female	Male	Female	Male	Female	Male	Female	Milan	Rome	Milan	Rome
Genetic Effect (GS)	0.407*** (0.097)	-0.091 (0.079)	0.452*** (0.104)	-0.109 (0.078)	0.383*** (0.113)	-0.358** (0.133)	0.396** (0.121)	-0.440** (0.132)	0.599** (0.182)	0.0754 (0.118)	0.572** (0.189)	0.105 (0.119)
Shared Environment (S)	0.200* (0.080)	0.518*** (0.065)	0.162+ (0.085)	0.532*** (0.064)	0.163+ (0.093)	0.741*** (0.117)	0.165+ (0.098)	0.782*** (0.117)	-0.047 (0.166)	0.421*** (0.095)	-0.037 (0.171)	0.391*** (0.097)
Identical Twins pair (G)	-0.033 (0.042)	0.001 (0.024)	-0.045 (0.045)	-0.006 (0.023)	-0.044 (0.052)	-0.021 (0.038)	-0.042 (0.054)	-0.010 (0.038)	0.012 (0.054)	-0.020 (0.037)	0.027 (0.056)	-0.021 (0.037)
R^2	0.306	0.220	0.360	0.319	0.240	0.296	0.274	0.356	0.328	0.198	0.360	0.220
N	306	556	305	554	223	223	223	223	132	287	132	287
<i>Wald Test</i>	χ^2	<i>p-value</i>	χ^2	<i>p-value</i>	χ^2	<i>p-value</i>	χ^2	<i>p-value</i>	χ^2	<i>p-value</i>	χ^2	<i>p-value</i>
GS(fav)=GS(unfav)	6.31	0.01*	7.81	0.005**	8.04	0.005**	13.09	0.000***	4.98	0.03*	3.93	0.05*
GS(2.fav)=GS(1.fav)	0.00	1.00	0.01	0.91	0.24	0.62	0.11	0.77	0.19	0.66	0.38	0.53
GS(2.unfav)=GS(1.unfav)	0.12	.72	0.08	0.78	0.46	0.50	1.04	0.31	0.00	0.96	0.01	0.91

Notes: Robust standard errors in parentheses. Estimation results for twin 1. Dependent variable is: “individual i is entrepreneur” Genetic Effect is the interaction between “Other twin entrepreneur” and “Identical twin pair”. Controls are a set of indicators variables per province of residence for models 2 and 4, and indicator variables for gender and education degrees (high school diploma, college degree) for model 6. The first Wald test confirms the hypotheses for the alternative twin, the second and the third Wald tests compare the coefficients within the pair. Significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$