ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis Strategy Economics

The effect of phenotypic traits on entrepreneurship: Turning the heritability question around

Name student: Anne de Vries

Student ID number: 481457

Supervisor: Tilbe Atav **Second assessor:** Niels Rietveld

Date final version: 02-08-2022

Table of contents

Abstract

Entrepreneurship has been a topic of interest in academic research. Since the wide availability of genetic data, scholars have tried to identify the 'entrepreneurial gene'. This research has not yielded satisfactory results. The search for the entrepreneur in this paper will therefore focus on the association between entrepreneurship and the traits ADHD, autism, height, and educational attainment. Polygenic scores are used to perform reduced form instrumental variable analysis and Mendelian randomization. The results suggest that the choice to become an entrepreneur is influenced by ADHD and autism. The results differ substantially between men and women. We discuss the limitations of Mendelian randomization and offer suggestions for the future of entrepreneurship research

1. Introduction

Entrepreneurship is one of the key drivers of today's economy. The exponential growth of the economy has historically been directly related to the emergence of entrepreneurship (Murphy et al., 2006). Entrepreneurs start businesses that range from the big tech companies like Google and Amazon to various small and medium-sized enterprises (SMEs) which are the backbone of the economy. Two out of three jobs and 50% of the GDP are created by SMEs in Europe according to the European Commission (2020). Entrepreneurship is therefore of considerable importance with respect to job creation and economic growth.

Entrepreneurship has an influence on society which cannot simply be measured through numbers as GDP and jobs created. One of the other important factors is the impact that it has on innovation (Schumpeter, 2000). Due to their ability to convert existing knowledge into innovations, entrepreneurs are able to react quickly to economic and social discontinuities and create new products which can offer utility to society (Block et al., 2013). In the wake of forthcoming challenges concerning environmental degradation and sustainability, entrepreneurship has become even more important, specifically in the form of sustainable entrepreneurship (Terán-Yépez et al. 2020).

The potential impact of entrepreneurship has led to increasing interest, not only in the field of public policy but also in the academic field. Scholars over the years have tried to identify the 'entrepreneur' and his characteristics. Understanding who the entrepreneur is an important factor in encouraging entrepreneurship. However, one of the main problems with this research was the lack of a clear definition (Shane & Venkataraman, 2000). Some even suggested that entrepreneurship is merely the creation of a business and that the question of "Who is an entrepreneur" was wrong, to begin with (Gartner, 1988).

Different streams of research have investigated the phenomenon of entrepreneurship. Entrepreneurship plays a key role in economic growth and innovation (Baumol & Strom, 2007). Understanding the entrepreneurial selection and the choices made by entrepreneurs is therefore highly relevant. A better understanding of entrepreneurs can lead to more effective policy and increase the effect entrepreneurship can have on society.

Developments in the field of molecular genetics opened up a new way of looking into the existence of the 'entrepreneur'. Genetic research could focus on the heritability of entrepreneurship. Empirical evidence from a twin study suggested that entrepreneurship was highly heritable (Nicolaou et al., 2008). Scholars tried to identify the specific entrepreneurial genes but have not been successful to date (Van der Loos et al., 2013; Rietveld et al., 2021).

The complexity of the behavioral trait and the lack of specific entrepreneurship-related data precluded the establishment of a clear genetic association.

Previous genetic research has identified specific genes of heritable traits. Although some of these traits might be related to entrepreneurial selection, genes specifically related to entrepreneurship have not yet been identified. The lack of specific genes found in combination with the high estimated heritability creates the idea of 'missing heritability'. In this thesis, an alternative method to investigate the heritability of entrepreneurship will be applied. Instead of focusing on the association between random genetic variance and entrepreneurship, the focus will be on the effect of various proven heritable traits or the genetic risk for these traits on entrepreneurship. The relationship between traits and entrepreneurship is investigated with the use of genetic data. If these heritable traits are associated with entrepreneurship, they might explain part of the 'missing heritability'.

The research question will be: Are the traits ADHD, autism, height, and educational attainment causally related to entrepreneurship. These traits are chosen because they are highly polygenic and heritable traits and could be influencing entrepreneurship for various reasons. Understanding these types of traits is important for researching the heritability and understanding the genetic background of entrepreneurship. This will provide us with a broader view of the characteristics of the entrepreneur. Furthermore, in the search for the entrepreneurial genotype, it is important to look beyond the effect of genetics only. This paper will explore the use of different methods to research the heritability of entrepreneurship.

Creating a broader view of factors related to entrepreneurship allows researchers to interpret the effect of genetic data and creates an overview of the information that needs to be considered when continuing research on the genetic background of entrepreneurship. This study will therefore add to the existing knowledge of the relationship between complex traits and entrepreneurship to create a better view of who the entrepreneur is.

To analyze the relationship between the traits and entrepreneurship, reduced form instrumental variable analysis and Mendelian randomization are used. These methods rely on the genetic risk of heritable variables. The relationship between these complex traits and entrepreneurship might be strongly confounded by different environmental factors. The instrumental variable analyses are used to eliminate bias stemming from potential confounders between the traits and entrepreneurship.

This paper will provide an introduction to the relevant concepts in genetic studies followed by a literature review of the heritability of entrepreneurship. After this, the traits of interest and the data used will be introduced. The method of Mendelian randomization will be explained extensively in the methodology section. Eventually, there are associations between ADHD and autism with entrepreneurship but no clear associations in the instrumental variable analyses for height and educational attainment.

2. Background of behavioral genetics

2.1 Heritability of behavioral traits

The effect of nature next to nurture has long been a subject of interest in the scientific field. Genetic studies have been highly important in identifying diseases and genetic disorders. However, not only the heritability of medical traits has been of interest. Genes have gained an increasingly important position in the behavioral field. Turkheimer (2000) even concluded, in his first law of behavioral genetics, that all behavioral traits are heritable. The early research in this field relied largely on family and twin studies, where comparisons of the resemblance in phenotypes between twins with and without identical genetics allowed for estimations of the heritability of complex traits. Although twin studies are important in identifying the heritability of behavioral traits, they are unable to identify the specific parts of the genome responsible for the behavior.

Following the finalization of the human genome project, it became possible to explore these specific parts of the genetic nature of humans. This development allowed the field of behavioral genetics to investigate the heritability of behavioral traits at the genetic level. Through candidate gene studies, researchers were now able to identify specific variances in the genotype related to complex traits (Kwon & Goate, 2000). When performing a candidate-gene study researchers pre-select the variation of interest. This approach, however, lacked statistical power when studying complex behavioral traits, leading to low replicability and a high number of false findings in the field of behavioral genetics (Benjamin et al., 2012). One of the problems was that complex behavioral traits depend on a variety of small genetic variations which all explained minor parts of the behavior (Chabris et al., 2015). Therefore, research relying on pre-specified genes was insufficient to answer the questions on the genetics of behavioral traits.

2.2 Genome-wide association studies (GWAS)

Genome-wide association studies offered a solution to the problems of candidate-gene studies and have been the focus of identifying genetic variance for over a decade now (Pearson & Manolio, 2008). In these studies, researchers observe the entire genome of a large sample of individuals to find associations between specific genetic variance (genotype) and certain traits (phenotype). Through GWA studies, without prior constraints of hypotheses, there have been various successful and replicated findings of risk loci for complex traits like educational attainment and body mass index (BMI) (Lee et al. 2018; Yengo et al., 2018).

The variations in the DNA that are relevant for this research are called single-nucleotide polymorphisms or SNPs. Our DNA exists of 4 nucleotides: adenine (A), thymine (T), guanine (G), and cytosine (C), which form a sequence together to entail the genetic information. SNPs are the most common form of genetic variants or polymorphisms and represent a change of a nucleotide in certain parts of the DNA sequence, thereby creating tiny variations (Pearson & Manolio, 2008; Medline, 2022). These variations can be harmless but might entail information related to various diseases and behavioral traits.

2.3 Limitations of GWAS

The significant associations found between SNPs and behavior allow for the identification of heritability. A crucial problem of GWAS is its inability to establish causal links between genetic variants and traits (Pearson & Manolio, 2008). While a GWAS identifies an association between alleles and phenotype, specific allele frequencies might be much more prevalent in certain populations. When these populations differ in behavior, a spurious correlation between differences in ancestry and specific traits will be found, instead of causal effects of genetic variants (McClellan & King, 2010). This phenomenon is known as population stratification.

A clear example of the dangers of population stratification is the ‘chopsticks-gene’. A scientist found a genetic marker strongly associated with eating with chopsticks. After reconsideration, however, it turned out that he found an unrelated gene that had different allele frequencies in Asians and Caucasians (Hamer & Sirota, 2000). To account for this, genetic studies have been performed with individuals solely of similar descent and with principal components to account for further ancestry differences (Price et al., 2006).

The idea behind principal components is that they capture as much genetic variability as possible of different groups in the dataset. In genetic data, this will be groups of different ancestral backgrounds. The principal components will adjust the variance due to common ancestry, making it possible to control for spurious correlations due to factors unrelated to genetics, like environment or culture.

While the outcomes of GWAS are significant and replicable associations are found, the effect sizes are still very small (Rietveld et al., 2013). SNPs are only able to explain small parts of the heritability of complex traits, suggesting some form of ‘missing heritability’ (Manolio et al., 2009). The SNPs found did not explain the heritability of phenotypes.

2.4 Polygenic risk scores (PGS)

Even though family and twin studies suggested high heritability, only a small effect size of the origin of this heritability could be found in the genetic architecture for both disorders (Gratten et al., 2014) and traits (Rietveld et al., 2013). Part of the reason behind this is that most SNPs that influence traits do not reach genome-wide statistical significance (Shi et al., 2016). Increasing the sample size, as is for example done in GWA studies for educational attainment, increases the strength of the genetic prediction for various traits (Okbay et al., 2016, Lee et al., 2018). Nevertheless, the lack of substantial established effect sizes, led to the proposal of a fourth law of behavioral genetics: "A typical human behavioral trait is associated with very many genetic variants, each of which accounts for a very small percentage of the behavioral variability." (Chabris et al., 2015). Due to this low explanation of variability, scientists nowadays prefer to work with polygenic risk scores (PGS). These scores are essentially weighted sums of all the SNPs found for a specific trait, disorder, or health risk (Torkamani et al., 2018). The high number of SNPs related to complex behavioral traits can be added to explain a relevant part of the inherited risk, leading to a PGS that summarizes the entire combined genetic effect found on a certain phenotype (Sugrue & Desikan, 2019).

The construction of PGS is necessary to ensure a sufficient impact of genetics in explaining phenotypical variance. This is especially important in the case of Mendelian Randomization with complex traits, where the genetic variance is used as an instrumental variable. An instrumental variable is required to have a relevant effect on the related exposure, known as the relevance assumption. To fulfill this assumption for genetic instruments, PGS are needed (Sugrue & Desikan, 2019, Davies et al., 2018). In this paper, we will therefore use PGS as an instrumental variable in our research on entrepreneurship.

3. Heritability of entrepreneurship

Entrepreneurship as a phenomenon nowadays involves a dynamic field of research over the years (Wiklund et al., 2011). The research is very diverse, with different views on entrepreneurship, including entrepreneurship as an organizational, labor market, or behavioral concept (Audretsch, 2012). One of the common organizational approaches of analyzing entrepreneurship is the examination of business ownership or self-employment (Parker, 2008; Parker, 2018). However, entrepreneurship is not only measured through taking personal risks and being self-employed. Entrepreneurship is a complex behavioral phenomenon and can for instance also be present in employees. The use of proxies harbours a risk of not identifying entrepreneurship correctly (Henrekson & Sanandaji, 2014). While we want to analyze entrepreneurship from a behavioral point of view, there are no extensive datasets on different types of entrepreneurship combined with the genetic data needed. Self-employment will therefore still be used as an indicator for entrepreneurship even though it comes with some limitations.

3.1 Heritability studies

Nicolaou et al. (2008) found that entrepreneurial behaviors were highly heritable in a twin study. The study, however, lacked representativeness due to the predominantly female sample. The latter can pose a problem due to the differences in environment females face compared to men regarding entrepreneurship (Jennings & McDougald, 2007). Newer studies also suggested a high heritability related to entrepreneurship (Van der Loos et al., 2013). While heritability has been established for entrepreneurial behavior, partly due to the heritability of traits like extraversion and neuroticism, it has remained challenging to identify specific entrepreneurial genes (Rietveld et al., 2021).

Genetics can influence entrepreneurship through different paths. Nicolaou & Sahne (2009) identified four ways through which genetics could influence entrepreneurship. From this research, we learned that, with complex behavioral traits like entrepreneurship, there are many ways through which a genetic effect could be explained. Van der Loos et al. (2013) tried to find specific SNPs that are associated with entrepreneurship but failed in finding these robust associations. The results did however suggest that SNPs explain a large part of the heritability. Likewise, Quaye et al. (2012) failed to find specific SNPs associated with entrepreneurship. The specific associations with SNPs can be found using GWAS but a large sample size will be needed (Koellinger et al., 2010). We can conclude that GWAS on the complex behavioral trait of entrepreneurship has not yielded satisfactory results yet (Rietveld et al., 2021).

3.2 Mechanisms of genetic influence

Genetics can influence entrepreneurial behavior in different ways, for instance through phenotypical traits or environmental interactions with these traits (Nicolaou & Shane, 2009). Four different mechanisms were proposed through which genetics may influence entrepreneurial selection.

The physiological effect

The physiological effects of genes are related to the chemical mechanisms in the brain and the way they react to the environment. An important genetic difference is the hormonal levels and production. Cortisol levels, for instance, are highly heritable (Bartels et al., 2003). Cortisol is the primary stress hormone of the human body and is strongly associated with behavioral inhibition in children (Fox et al., 2005). Behavioral inhibition leads to higher sensitivity in novel situations or around new people. It is furthermore related to post-error slowing: the tendency to slow down in new tasks after making an error in a previous one (Tops & Boksem, 2011). The combination of a higher sensitivity to novelty and an increased effect of past errors could become problematic for entrepreneurs since meeting people and making mistakes are daily business.

Genetic covariance

The second way that genes may influence entrepreneurship is through influencing individual differences which are associated with entrepreneurship. Several social skills such as adapting to a wide range of social situations and being persuasive will enhance the quality of social interactions and increase the network of individuals and therefore the likelihood of entrepreneurial success (Baron & Markman, 2000). The specific behavioral or personality traits that lead to entrepreneurial behavior can be found by looking into the prevailing traits in entrepreneurs (Baron, 2004). Important personality traits that are associated with entrepreneurship include self-efficacy, innovativeness, and risk-taking (Boyd & Vozikis, 1994, Frese, 2009). In addition, there is extensive literature on every Big-5 personality trait and its relation to entrepreneurship. These traits are found to be substantially heritable (Kerr et al., 2018; Loehlin et al., 1998). These Big-5 personality traits, however, are related to each other, allowing for mediation through the other behavioral traits. The latter complicates the estimation of the causal effect of specific personality traits (Awwad & Al-Aseer, 2021).

Entrepreneurship is normally thought of as a mental trait. However, even heritable physical aspects seem to be associated with entrepreneurship. There are for instance differences in entrepreneurial intentions between genders (Zhang et al., 2009). Gender can however also have a social, environmental, and behavioral effect. Another physical aspect that seems related to entrepreneurship is height. Rietveld, Hessels & Van der Zwan (2015) found that the self-employed were on average taller than employees. We can conclude that entrepreneurship or entrepreneurial intentions depend on various aspects which are thought to be heritable.

GxE interaction

Differences in attributes between individuals are decided by genetic predisposition as well as environmental factors. The extent to which a genetic predisposition influences the traits of an individual depends on the environment the individual lives in. The interaction between the economic environment and genetics in educational attainment and the way maternal stress and genes are related to ADHD symptomatology are examples of such gene-environment interactions (Thompson, 2014; Grizenko et al., 2012). These types of gene-environment interactions show how the genetic variance can be reflected through the environment and have an indirect effect on entrepreneurship. The genetic predisposition can stimulate the sensitivity to certain environmental events. Genes related to creativity or risk attitude can increase the environmental stimulus to become an entrepreneur and therefore influence the entrepreneurial selection.

Gene-environment correlation

The final way that genes can influence entrepreneurship is mediated by the environment. An individual's environment is constructed through for example education and work, which are in turn influenced by personal preferences. Genes, therefore, determine to which environment an individual is exposed (Plomin et al., 1977). There are some environments, such as the construction sector, where the tendency to become an entrepreneur is much stronger. Genetic predisposition might lead to individuals selecting into different societies with different entrepreneurial opportunities and may, through this gene-environment correlation, influence entrepreneurial selection. Because the relationship between genes and the environment is non-random, there can be a relationship between genes and traits mediated by the environment (Kendler & Eaves, 1986).

The way genes influence a trait or socioeconomic outcomes is complex due to the different pathways. The genetic basis for the traits of interest, as introduced in the next section, can influence entrepreneurship through multiple pathways. Genetic predisposition for ADHD can for instance be related to the way neurotransmitters work or to the attitude towards risk. It can furthermore be related to the attitude towards risk or certain skills that are prevalent in individuals with ADHD. In this paper, there is no specific pathway chosen through which the traits of interest influence entrepreneurship.

4. Introduction of the traits of interest

4.1 Phenotypes related to entrepreneurship

In this research, the effect of educational attainment, height, ADHD, and autism on entrepreneurship will be investigated. All these traits are found to be heritable through successful GWA studies and have an indication of being related to entrepreneurial selection.

Educational attainment

Educational attainment has been established as an important predictor of circumstances like health, income, and even life expectancy (Adams, 2002; Carlson & McChesney, 2015; Barro & Lee, 1994). Moreover, general education is related to entrepreneurial success (Dickson et al., 2008). However, when it comes to entrepreneurial selection the findings are still unclear. Van der Sluis et al. (2008), for instance, did not find any evidence for a relationship between educational attainment and the probability of entrepreneurship. The effects of schooling on entrepreneurial success were replicated but the association was influenced by differences in regions and gender.

There are several possible explanations as to why the effect on the choice for entrepreneurship is unclear. Educational attainment influences self-employment in two contradicting ways. On the one hand, there is a positive association, possibly due to increased managerial abilities. On the other hand, higher educational attainment results in an easier entry into the wage sector, inducing a negative association (Le, 1999). Another contradiction is that educated individuals gained skills useful not only for self-employment but also for wage-employment (Parker, 2008). Furthermore, higher educated individuals face higher opportunity costs of not taking a well-paid employee job (Estrin et al., 2016).

In addition, the effect of education on entrepreneurial selection differs by the personal background of individuals. For instance, individuals in developing countries prefer wage employment over entrepreneurship, likely due to less favorable policies or an increased preference for certainty (Van der Sluis et al., 2005). In research on migrant entrepreneurship in China, Cheng & Smyth (2021) found that education has a positive effect on the probability of being an employer entrepreneur as opposed to being a solo entrepreneur or employee.

Education is a determinant of the quality of entrepreneurship and also influences the number of entrepreneurs in a population. Singh and Crump (2007) found that individuals of a specific ethnicity with a lacking rate of education were less likely to become entrepreneurs. Educational attainment influences entrepreneurship and labor market outcomes depending on various aspects and is, therefore, an interesting topic of investigation in the research on entrepreneurial selection.

Height

Height and the labor market might seem unrelated at first glance; however, research has shown that height is related to earnings in some cases (Heineck, 2005). There are various explanations for this phenomenon, including increased self-esteem through height or natural dominance. Another explanation is that height is related to cognitive abilities and may therefore explain labor market differences (Case & Paxson, 2008). Schick and Steckel (2015) found that height is related to differences in cognitive as well as non-cognitive abilities. These factors combined could clarify the height premium in salaries.

Rietveld et al. (2015) found that an increase in height increased the probability of self-employment. The association may be caused by several sociological factors correlated with height. Furthermore, the socioeconomic background may influence lifestyle and subsequently an individual's height. The relationship between height and entrepreneurial selection is still unclear. The relationship will be explored in this paper through Mendelian Randomization.

ADHD

ADHD is a neurodevelopmental disorder with behavioral symptoms such as inattention, impulsivity, and hyperactivity (NIMH, 2021). These symptoms can influence work and school achievements and damage social life (Goodman, 2007). However, in the case of entrepreneurship these symptoms, especially hyperactivity, might benefit individuals (Antshel, 2018). Empirical studies have shown that the likelihood of becoming an entrepreneur increases with the diagnosis of ADHD (Dimic & Orlov, 2014). Verheul et al. (2016) found that the likelihood increases even further in the presence of stronger symptoms of ADHD.

In the labor market, ADHD is related to a lower income and lower likelihood of employment (Rietveld & Patel, 2019). These relationships are however partly mediated by educational attainment. Furthermore, part of the relationship between ADHD and earnings is explained by the increased likelihood of self-employment (Patel et al., 2021). Due to the lack of data on ADHD diagnosis, will use the reduced form of the instrumental variable regression with the PGS for ADHD as independent variable, in order to find if the genetic risk for ADHD influences later life entrepreneurial selection. In this analysis, the genetic predisposition towards ADHD will be treated as instrumental variable. Attention will therefore be paid to the various assumptions and there will be controlled for potential bias. This is where this research differs from the earlier found associations.

Autism spectrum disorder

Past research has shown that individuals with an autism spectrum disorder (ASD) have not been successful in terms of employment rates (Chen et al., 2015). Employment is important to become self-sufficient and leads to a higher quality of life. In individuals with ASD, the latter might demand some coaching or workplace modification (Hendricks, 2010). Self-employment may offer a chance for more success for individuals with ASD (Johnson, 2015). However, direct relationships between ASD and employment are hard to find since they depend on various personal and social factors (Holwerda et al., 2013). Self-employment might offer a solution for these individuals and might even connect to their specific skillset. Therefore, this study will investigate the presence of individuals with a higher genetic risk of ASD in the entrepreneurial community.

5. Data

5.1 Dataset

To illustrate the relationships between the polygenic scores and entrepreneurship, data is drawn from the US Health and Retirement Study (HRS). The HRS is a longitudinal survey of individuals over the age of 50 and their spouses in 23 000 households in the USA. The survey offers a representative panel of Americans. The survey fielded every 2 years since 1992 and provides data on income and labor market outcomes, health, cognition, and retirement. Since 2006, the HRS also contains data on genetics and biomarkers from participants that consented to genotyping. We use the latest cleaned longitudinal file from the HRS, also known as the RAND HRS longitudinal file 2018 (V1), which contains data from 1992-2018.

Multiple studies have been done to find SNPs associated with various behavioral traits and disorders. Polygenic scores were created for the genotyped participants of the HRS. We use the PGS data file from the HRS released in February 2021 (V4.3). This dataset contains polygenic scores of all the phenotypic traits mentioned in the introduction. The polygenic score for educational attainment is obtained through the GWAS of over 1 million individuals by Lee et al. (2018). The polygenic scores explain roughly 11-13% of the variance in educational attainment. The PGS for height, explaining ~24.6% of the variance, is obtained by the combined GWAS meta-analysis by Yengo et al. (2018). For the behavioral disorders, the first genome-wide risk loci for ADHD will be used (Demontis et al., 2017) and the meta-analysis of GWA studies for individuals with autism spectrum disorder (Anney et al., 2017). Due to the highly polygenic nature of these traits, only small fractions of the heritability are explained through these PGS. It is recommended, when using genetic data, to exclude all individuals that are not of similar descent. This way we can prevent some bias arising from the associations between environment and genetics such as population stratification. At last, we will also use the first 10 principal components of the covariance matrix included in the HRS. By including these we can control for the remaining population stratification which could exist due to common ancestry being related to shared culture or environmental factors as a result of non-random mating.

It is challenging to measure entrepreneurship in quantitative data due to the complexity of the trait. However, in economic research entrepreneurship is often proxied by self-employment (Parker, 2018). In the HRS data there is a survey question about self-employment that goes as follows: "Do you work for someone else, are you self-employed, or what?". The possible answers were: "for someone else", "self-employed" or "other". Only the answers: "for someone else" or "self-employed" were included and coded as a dummy variable where 1 indicates that the respondent answered: "self-employed" and 0 indicates that they answered: "for someone else".

All individuals that are not aged between 50 and 65 were excluded as well. Due to the standard retirement age and the minimum age of 50 for participating in the health and retirement study. Observations of individuals with an age lower than 50 are spouses or family members. Furthermore, all individuals that reported to be retired completely and all individuals that did not provide data on their employment status were dropped from the sample. After these individuals and all individuals that are not of recent European descent were excluded, 7915 individuals from 5297 different households were left providing 32553 person-year observations. So, every individual participated in roughly 4 periods with 2 years in between on average.

5.2 Descriptive statistics

The survey data consists of 14 waves from 7 different sample cohorts. The distribution of observations across the waves can be seen in table 2. The cohorts are based on birth year and the date of the first interview. The first 3 cohorts were obtained from separate studies. The *hrs* cohort is the initial cohort from this study. The dependent variable *Self-employed* has a mean of 0.191 indicating that in 19.1% of the observations individual *i* is self-employed at time *t*. It is furthermore interesting to mention that the average height is 170.9 centimeters. Table A of the appendix reports 0.010 variance within individuals due to people shrinking 1 cm on average in later stages of their lives. Part of this effect will be captured by controlling for the age of individuals. Educational attainment is measured through *educational years*. On average, individuals had 13.7 years of education. The distribution is shown in table B of the appendix. Most individuals had 12 years of education, which is in line with the duration of high school education of 12 years.

Table 1: Descriptive statistics of continuous variables

Variable	Obs	Mean	Std. Dev.	Min	Max
Self-employed	32553	.191	.393	0	1
PGS Height	32553	.025	.997	-3.767	4.227
PGS ADHD	32553	-.022	1.003	-3.57	3.682
PGS Autism	32553	-.06	1.01	-3.523	2.945
PGS Educational attainment	32553	.033	.993	-3.382	3.37
Height	32553	1.709	.099	1.245	2.261
Educational years	32453	13.727	2.369	0	17
Birthyear	32553	1945.767	7.991	1927	1968
Age	32553	57.286	3.819	50	64
Mother's years of education	31029	11.26	2.877	0	17
Father's years of education	29913	10.928	3.504	0	17
Childhood financial capital scale	27458	.144	.998	-3.101	3.044
wave	32553	6.733	3.816	1	14

Control variables

The sample consists of individuals born between 1927 and 1968. *Gender* is coded as a dummy variable with 1 representing males and 2 representing females. 53.33% of the sample consists of females. Individuals are married in 76.07% of the observations and are predominantly protestant (62.05%) or catholic (24.42%). The variables parental education and childhood financial capital contain a notable number of missing values. *Financial capital* is a scale-variable coded in such a way that higher numbers reflect a higher score for capital. The *self-reported health* measure is a categorical variable ranging from 1 to 5. The distribution of this variable is shown in table 2. Most observations indicate that the individual is healthy for the *self-reported health*.

Table 2: Summary statistics of categorical variables

Variable	Categories	Freq.	Percent
Gender			
	1.male	15233	46.67
	2.female	17408	53.33
Marital status			
	1.married	24819	76.07
	2.married,spouse absent	109	0.33
	3.partnered	1271	3.9
	4.separated	322	0.99
	5.divorced	3417	10.47

	6.separated/divorced	239	0.73
	7.widowed	1412	4.33
	8.never married	1036	3.18
Religion			
	1.protestant	20203	62.05
	2.catholic	7952	24.42
	3.jewish	672	2.06
	4.none/no pref	3395	10.43
	5.other	337	1.04
Cohort			
	0.hrs/ahead overlap	47	0.14
	1.ahead	60	0.18
	2.coda	19	0.06
	3.hrs	15464	47.38
	4.warbabies	5532	16.95
	5.early babyboomers	6685	20.48
	6.mid babyboomers	4834	14.81
Self-reported Health	1.excellent	6790	20.86
	2.very good	13107	40.27
	3.good	9265	28.46
	4.fair	2946	9.05
	5.poor	442	1.36

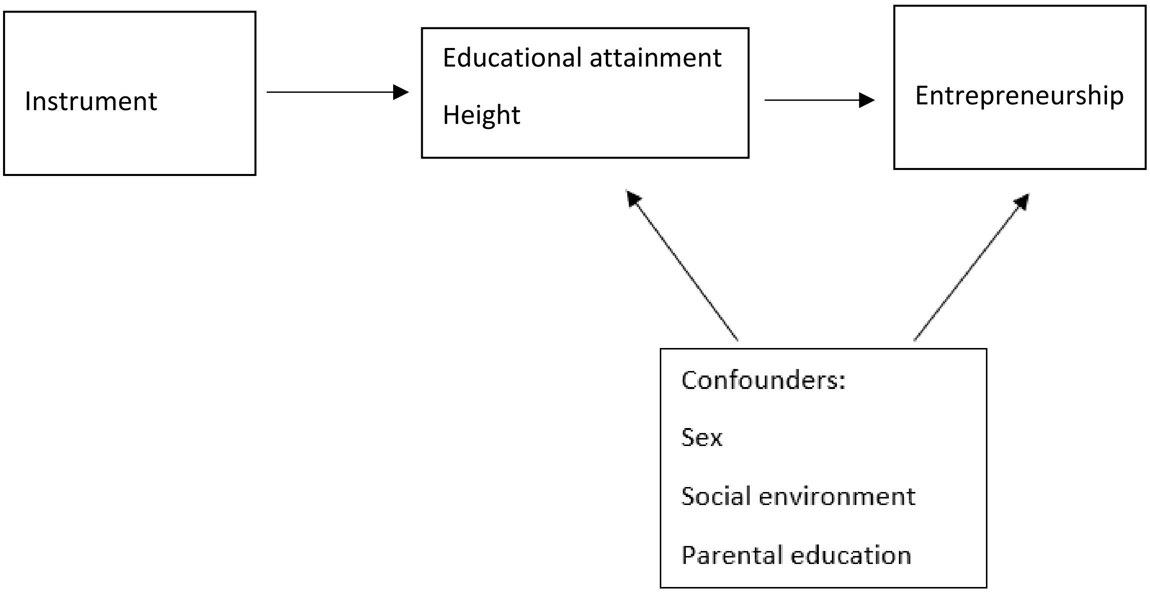
6. Methodology

6.1 Mendelian randomization

One of the main challenges in social sciences is finding causal relationships. Often only correlations or non-causal associations will be found due to confounding and omitted variables. The easiest way to conclude causality is by randomizing the treatment between groups, a study design also known as the randomized controlled trial (RCT). However, these trials are often infeasible from a financial and ethical point of view. In the case of genetics, it is even practically impossible to manually randomize treatment. In this paper, Mendelian randomization will be used as the research method. Mendelian randomization is a type of instrumental variable (IV) analysis. IV analysis was invented by Wright (1928) as a new way to infer causality. It provides an exogenous shock and therefore it is possible to isolate the effect of the variable of interest on the outcome. The idea behind Mendelian randomization is that the distribution of alleles is random and therefore suitable as instrument.

The concept of IV-analysis is best explained by using a Directed acyclic graph (DAG).

Figure 1: Directed acyclic graph (DAG) of instrumental variable (IV) analysis



A standard OLS approach would only consider the independent variables, educational attainment and height. The graph shows that the relationship between the independent variables and entrepreneurship could be biased by several confounders. We are not able to identify all factors and are unable to include all factors in the model. Since this violates the exogeneity assumption OLS would have biased results. By using an instrumental variable, we would be able to eliminate this bias.

One of the most general frameworks for IV-analysis is the two-stage least squares (2SLS) framework. The idea behind this framework is that the independent variable suffers from endogeneity, but the instrumental variable is either randomized or due to an exogenous shock. In the first stage, the independent variable is regressed on the instrumental variable. In the second stage, the fitted values of the independent variable are regressed on entrepreneurship. By using the fitted values from the regressions with the exogenous instrumental variable we eliminate potential endogeneity from the dependent variable.

6.2 assumptions for Mendelian randomization

For the IV-analysis to eliminate the endogeneity, some assumptions must hold (Lousdal, 2018). The three important assumptions are:

1. *The relevance assumption*
2. *The exchangeability or independence assumption*
3. *The exclusion restriction*

The relevance assumption indicates that the instrument should determine at least part of the independent variable. To satisfy the exchangeability or independence assumption the assignment of the instrumental variable needs to be unrelated to confounders or the dependent variable. This is often the case when there is an exogenous shock or in the case of randomization of the instrumental variable. The exclusion restriction demands that the instrumental variable only influences entrepreneurship through the independent variable. This means that the instrumental variable should not influence entrepreneurship, either directly or through pathways other than the independent variable.

With Mendelian randomization, the instruments used are the genetic variants, in our case represented by the PGS. The assumptions are represented visually in figure 2. Koellinger and De Vlaming (2019) paraphrased the assumptions for MR as follows (order is different from Koellinger and De Vlaming (2019) to match the order above):

Assumption 1: Some genes influence the exposure of interest.
Assumption 2: Genes are randomly assigned among people.
Assumption 3: The genes that influence the exposure do not influence the outcome via any other channel than the exposure.

Assumption 1: Some genes influence the exposure of interest.

The polygenic scores we use come from GWA studies that found a variety of SNPs and used a p-value threshold of 5×10^{-8}. Due to the high statistical power used in the GWA studies, we can assume that there is some association between the genes and the exposure. However, as mentioned earlier the GWAS can still find associations stemming from common ancestry. Because the GWA studies used are new and control extensively for population stratification it will be assumed that the effect found is the genetic effect.

In addition to significance, the magnitude of the effect is of importance. An instrument with a small influence on the exposure, also known as a weak instrument, can bias the results. The effect of the PGS on the exposure in the first stage regressions will therefore be examined. The results can be found in table 3. It is important to note that all controls of the second stage are also included in the first stage regressions. However, only the information relevant for testing the first assumption is reported in the table. Both coefficients are significant at a 1%-level ($p<0.01$), indicating that the instruments are relevant. The magnitude of the effect is roughly 3 centimeters per point PGS for height and an extra half-year of education for educational attainment. Given the distribution of the PGS, this would mean that on average the person with the highest PGS scores differs by 26 cm and 3.5 years of education from the individual with the lowest PGS scores ceteris paribus.

Another commonly used method to test the first assumption is by looking at the first-stage F-statistic. A general rule of thumb is that the F-statistic of the first stage should exceed 10 to be relevant (Staiger & Stock, 1997). The F-statistics reported in table 3 are of a pooled OLS regression because the other two-stage analysis does not provide an F-statistic. The values are well above the proposed benchmark. It can be concluded that there is evidence for the first assumption to hold.

Table 3: First stage regressions of PGS Height and PGS educational attainment on height and educational years respectively

	Coefficient	Robust std. err.	t	P>t	F-statistic
PGS Height	0.0327	0.0013	25.1	0.000	244.99
PGS Educational attainment	0.5213	0.0293	17.8	0.000	430.56

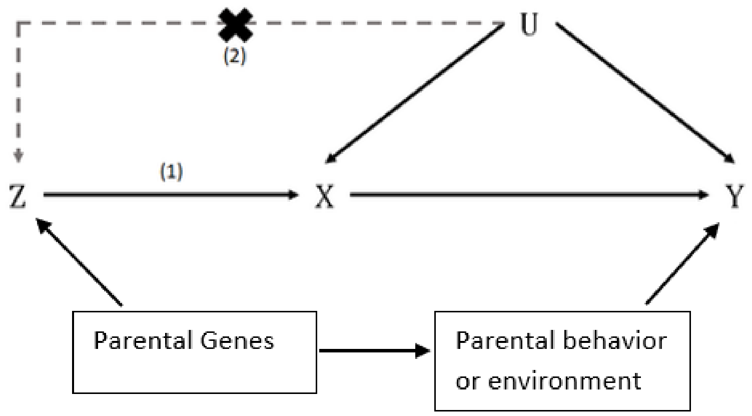
Assumption 2: Genes are randomly assigned among people.

The second assumption is that genes are randomly assigned among people. The idea behind Mendelian randomization is that every individual's genotype is assigned randomly at the moment of conception (Davey Smith & Ebrahim, 2003). This was empirically supported by the notion that genetic variants were not notably associated with socioeconomic, behavioral, and physiological factors (Davey Smith et al., 2007). However, while the genes are randomly assigned at conception, the available genes to randomize at conception are not. The randomization of the genotype is conditional on the genotype of the parents. This can lead to a 'genetic nurture' effect (Kong et al., 2018).

'Genetic nurture' occurs when the genotype of the parents influences the traits of their children mediated by environmental factors such as education and socioeconomic status. The best way to control for this is to include the parental genotype as control (Koellinger & De Vlaming, 2019). The second assumption might be violated if this is not possible. An example of this is shown in figure 2. The parental genes influence the genotype of their children directly while also influencing their behavior and environment. The genotype of the children will therefore be correlated with the environment of the parents which might correlate with our dependent variable, introducing bias to our model. The risk for this depends partly on the trait associated with the PGS. One of the instruments used in this paper is the PGS for educational attainment. Educational attainment is a complex trait that is influenced by a large variety of genetic variants. The increased number of genetic variants related to complex traits increases the risk of an assumption 2 violation (Solovieff et al., 2013).

Another possible violation of the second assumption is population stratification due to assortative mating. However, this is not expected to be a problem due to the restriction of the analyses to individuals of European descent and including the 10 first principal components.

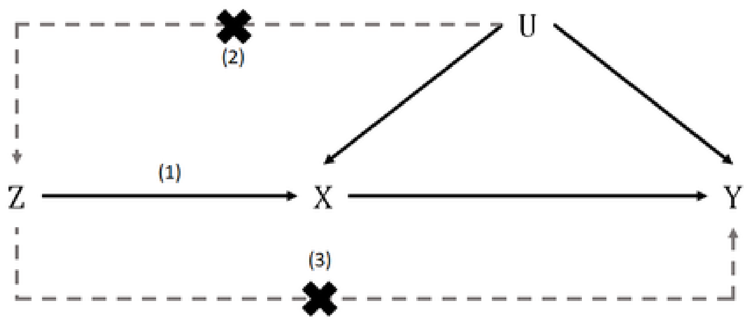
Figure 2: Visual representation of a potential assumption 2 violation



Assumption 3: The genes that influence the exposure do not influence the outcome via any other channel than the exposure.

The last assumption is the exclusion restriction. It is not possible to empirically test this assumption. The assumption is represented by line (3) in figure 3.

Figure 3: Visual representation of the assumptions for Mendelian randomization (Hazel Wade, 2021)



The most direct way that our instrument could influence the outcome is shown by line *(3)* which is known as the direct horizontal pleiotropic effect (Koellinger & De Vlaming, 2019).

Pleiotropy occurs when one genetic variant affects multiple phenotypic traits that are not directly related to each other. GWA studies identify associations between a specific trait and genetic variants and can therefore only assure that the genetic variants are associated with the trait of interest. Nevertheless, the possibility that the genetic variant also influences other traits, and therefore the possibility of bias, cannot be excluded (Paaby & Rockman, 2013).

Next to the direct effect, there is also the possibility of an indirect horizontal pleiotropic effect (Von Hinke et al., 2016; DiPrete, Burik & Koellinger, 2018). In this case, the genetic variant does not influence the dependent variable directly but through a number of other traits. The indirect pleiotropic effects create various new pathways for potential bias. Due to the complexity of genetics and the extensive ways through which bias could occur, it is not possible to ensure that the third assumption will hold.

Lastly, an additional assumption that is often not mentioned as one of the main assumptions is the monotonicity assumption. This assumption implies that every subject should have a monotonic increasing function for the increase in PGS (Small et al., 2014). It should not be possible for the genetic markers to influence the traits negatively in one person and positively in another person. It is not possible to test this assumption since there are no perfect counterfactuals for individuals. A violation of this assumption could occur in case of strong gene-environment interactions (Von Hinke et al., 2016). While gene-environment interactions are reasonable for height and educational attainment, there is no reason to expect them to be strong enough to violate this assumption.

6.3 Empirical models

To illustrate the empirical association between our phenotypes and entrepreneurial selection, we will use instrumental variable analyses. For ADHD and Autism, we cannot use the 2SLS estimator due to the lack of diagnoses for these disorders. Therefore, a reduced form IV-analysis is performed. The reduced form regresses entrepreneurship on the instruments. If all assumptions hold then the influence measured will be the influence of the exposure, in this case, ADHD or Autism, providing an indication of the influence of these behavioral disorders on the outcome.

Our dependent variable *self-employed* is collected every two years and is time-variant. If this is not accounted for, it can result in bias due to the potential serial correlation between the error terms. Multiple observations within the same individual cannot be considered as independent. Therefore, this study will make use of the panel data in the models and estimate the IV with the G2SLS-estimator by Balestra and Varadharajan-Krishnakumar (1987). This method allows for the inclusion of random effects into the model. The correlation between the error terms must be eliminated to ensure the correct standard errors. Random effects estimators use a GLS transformation to eliminate this correlation. The GLS transformation estimates the demeaning factor which can be subtracted from the initial model to estimate the model with quasi-demeaned data. This will make the observations uncorrelated with the other observations. Both the full model as the reduced form will work with random effects. The mathematical descriptions of the model are below, the demeaning factor is represented by the θ. The time-variant part of the error term is represented by μ and the time-invariant part by α.

First stage: $A_{it} = \beta_0(1 - \theta) + \beta_1(Zi - \theta Zi) + \beta_2(Zit - \theta\bar{Z}i) + \alpha_i - \theta\alpha_i + \mu_{it} - \theta\bar{\mu}_i$

Second stage: $Y_{it} = \beta_0(1 - \theta) + \beta_1\hat{A}_{it} + \beta_2(Zit - \theta\bar{Z}i) + \alpha_i - \theta\alpha_i + \mu_{it} - \theta\bar{\mu}_i$

Reduced form: $Y_{it} = \beta_0(1 - \theta) + \beta_1(Zi - \theta Zi) + \beta_2(Zit - \theta\bar{Z}i) + \alpha_i - \theta\alpha_i + \mu_{it} - \theta\bar{\mu}_i$

Y_{it} is a binary variable indicating the self-employment of individual *i* at wave *t*. $\beta_1 X_i$ is an indicator for the polygenic scores of a certain trade for individual *I* and $\beta_i Z_{it}$ represents the vector of control variables. In the first stage, the exposure (height or educational attainment) represented by A_{it} will be regressed on $\beta_1 Z_i$ and the effect of other control variables: $\beta_2 Z_{it}$. After this, we will use the fitted values of A_{it} represented by $\beta_1\hat{A}_{it}$ instead of the endogenous estimator in combination with the vector of controls $\beta_2 Z_i$ to estimate the effect on entrepreneurship Y_{it}.

The reduced form estimator regresses entrepreneurship directly on the polygenic scores without considering the exposure. If we can assume that the instruments for the non-existent exposure are valid and that all variation runs through our exposure, the effect of Z_i will still be consistently estimated. Control variables were included for some of the potential biases that may violate the assumptions.

Due to the differences between men and women, for instance, in the way, they experience behavioral disorders, general height, and motivation from the environment for education, and the differences in genetic basis and potential interactions with gender, all the analyses will be performed in subsamples divided by gender (Zhang et al., 2009).

6.4 Control variables

As mentioned previously, both the independence and exclusion assumptions might be violated in this study. A violation of the independence assumption creates a potential path for bias through the parental genotype to parental behavior and environment (figure 2). Since information on the parental genotype is lacking, it is impossible to control for this completely.

We do however have some information on the childhood circumstances. By controlling for this we can capture part of the genetic information of the parents.

Controlling for pleiotropic effects is challenging since as of today we do not have sufficient knowledge of the way the genetic variants work. It has however been demonstrated that height is related to cognitive abilities and can therefore be controlled for in the analysis (Case & Paxson, 2008; Schick & Steckel, 2015).

To diminish bias arising from the parental environment, controls are added for maternal and paternal education levels and the childhood financial capital scale made by Vable et al. (2017) containing various factors like financial status, father's occupation, and declaring bankruptcy. In this way, (part of) the correlation between the genotype of the children and the environment they grew up in is captured. Controls for marital status and religion are also included due to the association they have with the environment and entrepreneurial selection (Molina, 2020; Audretsch et al., 2007; Dodd & Gotsis, 2007).

To control for the parental role model effect of healthy lifestyle choices, self-reported health is included in the model with height as the independent variable (Hoffmann et al., 2015). A healthy lifestyle can be associated with height and the choice of becoming an entrepreneur and will therefore capture part of the genetic effect of lifestyle choices. Since genes are known to have pleiotropic effects on cognitive abilities (Schick & Steckel, 2015), the educational attainment of the individual is included as well. In addition, educational attainment is added to the reduced form analyses to prevent that the measured association is mostly mediated by the lack of good education for people with behavioral disorders in the early 20th century.

At last, some regular controls are added. The first ten principal components are included in the model as is standard practice (Price et al., 2006). Fixed effects are included for cohort and birth year to eliminate any effect stemming from the year of measurement. Age is included in the model to capture part of the influence of the decreasing height and to just control for the influence age and experience have on life decisions. Due to the complexity of the traits and the possibility of bias a 5% level ($p<0.05$) threshold is set for the results in this paper.

7. Results

7.1 Reduced form analyses

ADHD

The results for the reduced form analyses of ADHD for males and females are reported in tables 4 and 5 respectively. The results must be interpreted as average changes to the probability of becoming an entrepreneur due to the linear nature of the reduced form analysis and the binary dependent variable. When the PGS of ADHD for males increases by 1, the probability of being an entrepreneur on average increases by roughly 1.4 percentage point ceteris paribus. This result has statistical significance at a 10% level (p<0.1) but not at the 5% threshold.

Table 4: Reduced form estimation with random effects for the Polygenic scores of ADHD of males

Self-employment	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
PGS ADHD	0.013	0.008	1.65	0.099	-0.003	0.029	*
Educational years	0.003	0.003	0.94	0.347	-0.003	0.010	
Mother's years of education	0.007	0.004	1.96	0.050	0.000	0.014	**
Father's years of education	0.000	0.003	0.15	0.883	-0.005	0.006	
Childhood financial capital	0.016	0.008	1.91	0.056	-0.000	0.032	*
2.married, spouse absent	-0.036	0.021	-1.71	0.087	-0.078	0.005	*
3.partnered	0.008	0.025	0.33	0.742	-0.040	0.057	
4.separated	0.014	0.028	0.50	0.616	-0.040	0.068	
5.divorced	-0.020	0.021	-0.97	0.332	-0.061	0.021	
6.separated/divorced	-0.037	0.025	-1.47	0.140	-0.087	0.012	
7.widowed	-0.056	0.020	-2.73	0.006	-0.096	-0.016	***
8.never married	-0.047	0.037	-1.27	0.204	-0.119	0.025	
2.catholic	-0.006	0.020	-0.28	0.781	-0.046	0.034	
3.jewish	0.065	0.087	0.74	0.457	-0.106	0.236	
4.none/no pref	0.029	0.027	1.06	0.290	-0.024	0.082	
5.other	-0.044	0.090	-0.48	0.628	-0.220	0.133	
Age	0.004	0.001	4.03	0.000	0.002	0.006	***
Constant	0.187	0.360	0.52	0.604	-0.519	0.893	

Mean dependent var	0.246	SD dependent var	0.431
Overall r-squared	0.036	Number of obs	11369
Chi-square	.	Prob > chi2	.
R-squared within	0.004	R-squared between	0.047

**** p<.01, ** p<.05, * p<.1*

Base-level for marital status is married; base-level for religion is Protestant. 10 Principal components, birth year fixed effects, and cohort fixed effects are included in the analysis but not in the table

When the PGS of ADHD for females increases by 1 the probability of being an entrepreneur on average increases by 2 percentage points ceteris paribus at a 1% significance level (p<0.01). When the assumptions hold in the reduced form analysis the effect measured will only be the effect measured through the exposure. The results, therefore, suggest that females with ADHD are more likely to be entrepreneurs as compared to females without ADHD. At last, it is interesting to mention that marital status and age have significant effects on entrepreneurial selection in both analyses and maternal education and religion in one of the analyses. The base level for marital status is married and the base level for religion is protestant. Being widowed for males and being divorced for females decreases the probability of being an entrepreneur compared to being married. No religious preference increases the probability of being an entrepreneur for females compared to being protestant.

Table 5: Reduced form estimation with random effects for the Polygenic scores of ADHD of females

Self-employment	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
PGS ADHD	0.020	0.006	3.17	0.002	0.008	0.032	***
Educational years	-0.005	0.003	-1.62	0.105	-0.010	0.001	
Mother's years of education	0.001	0.003	0.35	0.726	-0.004	0.006	
Father's years of education	0.002	0.002	0.80	0.426	-0.002	0.006	
Childhood financial capital	0.004	0.007	0.62	0.533	-0.009	0.018	
2.married, spouse absent	-0.007	0.030	-0.22	0.824	-0.065	0.052	
3.partnered	-0.002	0.022	-0.08	0.940	-0.044	0.041	
4.separated	-0.049	0.025	-1.94	0.053	-0.098	0.001	*
5.divorced	-0.044	0.014	-3.23	0.001	-0.071	-0.017	***
6.separated/divorced	-0.039	0.022	-1.78	0.075	-0.081	0.004	*
7.widowed	-0.021	0.015	-1.38	0.169	-0.051	0.009	
8.never married	-0.062	0.037	-1.67	0.095	-0.135	0.011	*
2.catholic	-0.018	0.016	-1.18	0.240	-0.049	0.012	
3.jewish	-0.061	0.068	-0.89	0.372	-0.194	0.073	
4.none/no pref	0.070	0.028	2.47	0.014	0.014	0.125	**
5.other	0.135	0.124	1.09	0.276	-0.108	0.379	
Age	0.003	0.001	3.40	0.001	0.001	0.004	***
Constant	-0.039	0.130	-0.30	0.762	-0.294	0.215	

Mean dependent var	0.159	SD dependent var	0.366
Overall r-squared	0.022	Number of obs	13322
Chi-square	.	Prob > chi2	.
R-squared within	0.003	R-squared between	0.032

**** p<.01, ** p<.05, * p<.1*

Base-level for marital status is married; base-level for religion is Protestant. 10 Principal components, birthyear fixed effects, and cohort fixed effects are included in the analysis but not in the table

Autism

The results of the reduced form analyses with the PGS for autism should be interpreted in a similar manner to those of ADHD. No significant effect of the PGS on self-employed males was observed. Significant effects are found for the control variables marital status and age for males. Being widowed for males decreases the probability of being an entrepreneur compared to being married.

Table 6: Reduced form estimation with random effects of the Polygenic scores for autism of males

Self-employment	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
PGS Autism	-.01	.012	-0.85	.393	-.033	.013	
Educational years	.003	.003	0.81	.42	-.004	.009	
Mother's years of education	.007	.004	1.91	.056	0	.014	*
Father's years of education	0	.003	0.15	.879	-.005	.006	
Childhood financial capital	.016	.008	1.85	.064	-.001	.032	*
2.married, spouse	-.036	.021	-1.71	.087	-.078	.005	*

absent							
3.partnered	.008	.025	0.33	.744	-.041	.057	
4.separated	.014	.028	0.52	.606	-.04	.069	
5.divorced	-.02	.021	-0.95	.344	-.06	.021	
6.separated/divorced	-.037	.025	-1.46	.144	-.086	.013	
7.widowed	-.056	.02	-2.74	.006	-.096	-.016	***
8.never married	-.048	.037	-1.31	.19	-.12	.024	
2.catholic	-.006	.02	-0.31	.757	-.046	.034	
3.jewish	.068	.088	0.77	.44	-.105	.24	
4.none/no pref	.029	.027	1.06	.289	-.024	.082	
5.other	-.046	.089	-0.52	.604	-.222	.129	
Age	.004	.001	4.04	0	.002	.006	***
Constant	.209	.359	0.58	.561	-.494	.912	

Mean dependent var	0.246	SD dependent var	0.431
Overall r-squared	0.036	Number of obs	11369
Chi-square	.	Prob > chi2	.
R-squared within	0.004	R-squared between	0.046

**** p<.01, ** p<.05, * p<.1*

Base-level for marital status is married; base-level for religion is Protestant. 10 Principal components, birthyear fixed effects, and cohort fixed effects are included in the analysis but not in the table

When the PGS of autism for females increases by 1 the probability of being an entrepreneur on average decreases by 1.8 percentage points ceteris paribus at a statistical significance level of 5% ($p<0.05$). Given that the assumptions hold, this implies that autism in females decreases the propensity to become self-employed. As with the analysis of ADHD, there are significant effects of the control variables marital status, religion, and age indicating that being divorced for females decreases the probability of being an entrepreneur compared to being married and no religious preference increases the probability of being an entrepreneur compared to being protestant.

Table 7: Reduced form estimation with random effects of the Polygenic scores for autism of females

Self-employment	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
PGS Autism	-0.018	0.009	-2.00	0.045	-0.036	-0.000	**
Educational years	-0.005	0.003	-1.80	0.072	-0.011	0.000	*
Mother's years of education	0.001	0.003	0.27	0.791	-0.004	0.006	
Father's years of education	0.002	0.002	0.74	0.458	-0.003	0.006	
Childhood financial capital	0.003	0.007	0.46	0.649	-0.010	0.016	
2.married, spouse absent	-0.006	0.030	-0.19	0.850	-0.064	0.053	
3.partnered	-0.002	0.022	-0.08	0.937	-0.044	0.041	
4.separated	-0.048	0.025	-1.90	0.057	-0.097	0.002	*
5.divorced	-0.044	0.014	-3.21	0.001	-0.071	-0.017	***
6.separated/divorced	-0.038	0.022	-1.77	0.078	-0.081	0.004	*
7.widowed	-0.021	0.015	-1.37	0.171	-0.051	0.009	
8.never married	-0.063	0.037	-1.69	0.092	-0.136	0.010	*
2.catholic	-0.017	0.016	-1.07	0.286	-0.047	0.014	

3.jewish	-0.056	0.069	-0.80	0.422	-0.191	0.080	
4.none/no pref	0.070	0.028	2.45	0.014	0.014	0.126	**
5.other	0.135	0.124	1.09	0.277	-0.108	0.377	
Age	0.003	0.001	3.40	0.001	0.001	0.004	***
Constant	-0.032	0.129	-0.24	0.807	-0.284	0.221	

Mean dependent var	0.159	SD dependent var	0.366
Overall r-squared	0.018	Number of obs	13322
Chi-square	.	Prob > chi2	.
R-squared within	0.003	R-squared between	0.030

**** p<.01, ** p<.05, * p<.1*

Base-level for marital status is married; base-level for religion is Protestant. 10 Principal components, birthyear fixed effects, and cohort fixed effects are included in the analysis but not in the table

7.2 Two-stage analyses

Height

Instead of the effect of the PGS measured in the reduced form analysis, the second stage measures the effect of the fitted value of the exposure. The interpretation is still an average effect on the probability of self-employment.

Table 8: Second stage regression of the effect of height on self-employment for males

Self-employment	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Height	.069	.258	0.27	.788	-.436	.574	
2.very good	-.01	.008	-1.27	.206	-.025	.005	
3.good	-.022	.009	-2.40	.016	-.04	-.004	**
4.fair	-.027	.013	-2.11	.035	-.052	-.002	**
5.poor	-.026	.018	-1.47	.142	-.062	.009	
Educational years	.002	.003	0.47	.636	-.005	.008	
Mother's years of education	.006	.004	1.79	.073	-.001	.013	*
Father's years of education	0	.003	0.16	.874	-.005	.006	
Childhood financial capital	.015	.008	1.79	.074	-.001	.032	*
2.married, spouse absent	-.037	.021	-1.74	.082	-.079	.005	*
3.partnered	.008	.025	0.33	.738	-.04	.057	
4.separated	.012	.028	0.42	.672	-.043	.066	
5.divorced	-.021	.021	-1.00	.317	-.061	.02	
6.separated/divorced	-.039	.025	-1.54	.124	-.089	.011	
7.widowed	-.056	.02	-2.80	.005	-.096	-.017	***
8.never married	-.046	.037	-1.24	.214	-.119	.027	
2.catholic	-.007	.02	-0.33	.74	-.047	.033	
3.jewish	.07	.088	0.79	.429	-.103	.242	
4.none/no pref	.03	.027	1.11	.269	-.023	.083	
5.other	-.046	.091	-0.50	.615	-.223	.132	
Age	.004	.001	4.33	0	.002	.006	***
Constant	.092	.571	0.16	.872	-1.027	1.211	

Mean dependent var	0.246	SD dependent var	0.431
Overall r-squared	0.037	Number of obs	11368
Chi-square	49004.328	Prob > chi2	0.000

R-squared within	0.005	R-squared between	0.048

**** p<.01, ** p<.05, * p<.1*
Base-level for marital status is married; base-level for religion is Protestant; Base-level for self-reported health is excellent. 10 Principal components, birthyear fixed effects, and cohort fixed effects are included in the analysis but not in the table

For both males and females there are no significant results of height on self-employment found. Interesting to mention are the negative significant results of self-reported health in table 8. The negative effect is expected because the self-reported health is coded in such a way that excellent health is the base level. Compared to excellent health, good or fair health decreases the probability of being an entrepreneur for males but not for females. It is furthermore interesting to see that the significance of the control variables corresponds with the earlier analyses.

T*able 9: Second stage regression of the effect of height on self-employment for females*

Self-employment	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Height	.116	.205	0.57	.571	-.286	.517	
2.very good	-.009	.007	-1.30	.192	-.022	.004	
3.good	-.011	.008	-1.39	.164	-.027	.005	
4.fair	.005	.012	0.46	.649	-.018	.028	
5.poor	-.039	.028	-1.39	.166	-.094	.016	
Educational years	-.006	.003	-2.03	.042	-.012	0	**
Mother's years of education	.001	.003	0.31	.757	-.004	.006	
Father's years of education	.001	.002	0.67	.5	-.003	.006	
Childhood financial capital	.003	.007	0.42	.676	-.011	.016	
2.married, spouse absent	-.005	.03	-0.16	.873	-.064	.054	
3.partnered	-.002	.022	-0.08	.938	-.044	.041	
4.separated	-.048	.025	-1.92	.054	-.098	.001	*
5.divorced	-.045	.014	-3.31	.001	-.072	-.018	***
6.separated/divorced	-.04	.022	-1.85	.064	-.083	.002	*
7.widowed	-.022	.015	-1.47	.143	-.053	.008	
8.never married	-.063	.037	-1.69	.091	-.135	.01	*
2.catholic	-.017	.016	-1.11	.269	-.048	.013	
3.jewish	-.053	.068	-0.77	.441	-.187	.082	
4.none/no pref	.069	.028	2.43	.015	.013	.125	**
5.other	.133	.126	1.06	.291	-.114	.379	
Age	.003	.001	3.50	0	.001	.004	***
Constant	-.23	.365	-0.63	.53	-.946	.487	

Mean dependent var	0.159	SD dependent var	0.366
Overall r-squared	0.019	Number of obs	13320
Chi-square	768.227	Prob > chi2	0.000
R-squared within	0.004	R-squared between	0.031

**** p<.01, ** p<.05, * p<.1*
Base-level for marital status is married; base-level for religion is Protestant; Base-level for self-reported health is excellent. 10 Principal components, birthyear fixed effects, and cohort fixed effects are included in the analysis but not in the table

Educational attainment

The last variable of interest is the educational attainment represented by the years of education an individual received.

Table 10: Second stage regression of the effect of educational attainment on self-employment for males

Self-employment	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Educational years	-0.007	0.015	-0.51	0.609	-0.036	0.021	
Mother's years of education	0.008	0.004	1.99	0.047	0.000	0.016	**
Father's years of education	0.002	0.004	0.57	0.570	-0.005	0.009	
Childhood financial capital	0.018	0.009	2.00	0.045	0.000	0.035	**
2.married, spouse absent	-0.037	0.021	-1.75	0.080	-0.078	0.004	*
3.partnered	0.007	0.025	0.28	0.776	-0.042	0.056	
4.separated	0.012	0.028	0.45	0.654	-0.042	0.067	
5.divorced	-0.021	0.021	-1.02	0.308	-0.062	0.020	
6.separated/divorced	-0.039	0.025	-1.52	0.129	-0.088	0.011	
7.widowed	-0.057	0.020	-2.81	0.005	-0.097	-0.017	***
8.never married	-0.046	0.037	-1.24	0.215	-0.118	0.027	
2.catholic	-0.004	0.021	-0.19	0.847	-0.044	0.036	
3.jewish	0.078	0.089	0.87	0.382	-0.096	0.252	
4.none/no pref	0.031	0.027	1.13	0.261	-0.023	0.084	
5.other	-0.044	0.089	-0.50	0.620	-0.219	0.130	
Age	0.004	0.001	4.03	0.000	0.002	0.006	***
Constant	0.293	0.369	0.79	0.427	-0.430	1.016	

Mean dependent var	0.246	SD dependent var	0.431
Overall r-squared	0.032	Number of obs	11369
Chi-square	44131667.015	Prob > chi2	0.000
R-squared within	0.004	R-squared between	0.043

**** p<.01, ** p<.05, * p<.1*

Base-level for marital status is married; base-level for religion is Protestant. 10 Principal components, birthyear fixed effects, and cohort fixed effects are included in the analysis but not in the table

No effect of the years of education on the entrepreneurial selection is found for males. The influence of the control variables corresponds to earlier analyses with the addition of childhood financial capital and maternal education that reach the 5% significance level. In the female sample, there is no significant effect of educational years found as well. These tables lead to the conclusion that there is no effect of educational years on entrepreneurship. The significance of the control variables is again matching with the earlier analyses.

Table 11: Second stage regression of the effect of educational attainment on self-employment for females

Self-employment	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Educational years	-0.009	0.013	-0.66	0.506	-0.035	0.017	
Mother's years of	0.001	0.003	0.40	0.687	-0.005	0.008	

education							
Father's years of education	0.002	0.003	0.73	0.467	-0.003	0.007	
Childhood financial capital	0.004	0.007	0.54	0.587	-0.010	0.018	
2.married, spouse absent	-0.006	0.030	-0.20	0.842	-0.064	0.053	
3.partnered	-0.002	0.022	-0.08	0.935	-0.044	0.041	
4.separated	-0.048	0.025	-1.91	0.056	-0.097	0.001	*
5.divorced	-0.044	0.014	-3.24	0.001	-0.071	-0.018	***
6.separated/divorced	-0.039	0.022	-1.79	0.073	-0.081	0.004	*
7.widowed	-0.022	0.015	-1.41	0.159	-0.052	0.008	
8.never married	-0.059	0.038	-1.56	0.120	-0.133	0.015	
2.catholic	-0.017	0.016	-1.11	0.265	-0.048	0.013	
3.jewish	-0.053	0.069	-0.78	0.437	-0.188	0.081	
4.none/no pref	0.071	0.029	2.42	0.015	0.014	0.129	**
5.other	0.134	0.126	1.06	0.289	-0.113	0.380	
Age	0.003	0.001	3.40	0.001	0.001	0.004	***
Constant	-0.007	0.176	-0.04	0.970	-0.351	0.338	

Mean dependent var	0.159	SD dependent var	0.366
Overall r-squared	0.017	Number of obs	13322
Chi-square	765.535	Prob > chi2	0.000
R-squared within	0.003	R-squared between	0.029

**** p<.01, ** p<.05, * p<.1*

Base-level for marital status is married; base-level for religion is Protestant. 10 Principal components, birthyear fixed effects, and cohort fixed effects are included in the analysis but not in the table

In the theoretical framework, it was mentioned already that educational attainment might influence entrepreneurial selection in contradicting ways by improving skills for both regular jobs as self-employment. Moreover, some industries require less education but are more likely to exist primarily of entrepreneurs. Furthermore, the choice for self-employment might also be affected by the willingness to take risks and therefore by the financial background of an individual. For these reasons, subgroup analyses for the influence of educational attainment on self-employment were performed.

The results of the subgroup analyses divided by childhood financial capital are shown in table 12. Males and females are still separated in these subgroup analyses. For the first 4 columns, the sample is stratified between high childhood financial capital (column 1 (males) and 2 (females)) and low childhood financial capital (column 3 (males) and 4 (females)). The significant effect in table 10 indicated that childhood financial capital is a predictor of entrepreneurship. The influence of education might be strengthened by increased financing possibilities or increased risk-seeking behavior.

Table 12: Subgroup two-stage analyses divided by childhood financial capital

	(1)	(2)	(3)	(4)
Educational years	0.007	0.039	-0.013	-0.029*
	(0.028)	(0.026)	(0.017)	(0.015)

Mother's years of education	0.009	0.003	0.007	0.001
	(0.008)	(0.006)	(0.005)	(0.004)
Father's years of education	-0.006	-0.009*	0.005	0.004
	(0.006)	(0.005)	(0.004)	(0.003)
Childhood financial capital scale	0.083***	0.049*	0.005	-0.006
	(0.030)	(0.026)	(0.012)	(0.010)
2.married, spouse absent	-0.090**	-0.030	0.003	0.013
	(0.038)	(0.043)	(0.015)	(0.039)
3.partnered	-0.014	-0.003	0.025	0.001
	(0.039)	(0.026)	(0.034)	(0.032)
4.separated	0.001	-0.069	0.021	-0.034
	(0.055)	(0.059)	(0.028)	(0.022)
5.divorced	-0.030	-0.069***	-0.013	-0.031*
	(0.033)	(0.021)	(0.029)	(0.018)
6.separated/divorced	0.001	-0.070*	-0.052	-0.023
	(0.037)	(0.038)	(0.033)	(0.027)
7.widowed	-0.074**	-0.047	-0.046*	-0.008
	(0.030)	(0.030)	(0.027)	(0.017)
8.never married	-0.084**	-0.154**	-0.007	0.012
	(0.038)	(0.070)	(0.070)	(0.037)
2.catholic	-0.006	-0.025	0.003	-0.010
	(0.033)	(0.026)	(0.027)	(0.020)
3.jewish	0.006	-0.122	0.214	0.037
	(0.116)	(0.097)	(0.141)	(0.098)
4.none/no pref	0.020	0.043	0.031	0.080*
	(0.042)	(0.045)	(0.037)	(0.041)
5.other	-0.146**	0.183	0.074	0.084
	(0.071)	(0.272)	(0.135)	(0.141)
Age	0.004**	0.003**	0.004***	0.002**
	(0.002)	(0.001)	(0.001)	(0.001)
_cons	0.038	-0.143	0.320	0.000
	(0.422)	(0.309)	(0.367)	(0.179)
Observations	4674	5323	6695	7999
Birthyear FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes

Standard errors are in parentheses; base-level for marital status is married; base-level for religion is Protestant. 10 Principal components, birthyear fixed effects, and cohort fixed effects are included in the analysis but not in the table
**** p<.01, ** p<.05, * p<.1*

The subgroup analyses only provide one result close to significance (p=0.55), namely that females in the bottom half of childhood financial capital are less likely to become entrepreneurs. For every year of education, the probability of being an entrepreneur on average decreases by roughly 3 percentage points ceteris paribus. However, this result is not statistically significant at a 5% level. It is furthermore interesting to mention that even though the results are insignificant the signs of the results between the high financial capital and low financial capital differ. This could indicate that the financial background of an individual might play a role in the way educational attainment affects entrepreneurial selection.

The final subgroup analyses are shown in table 13. The regressions are stratified between individuals with degrees from higher education (i.e. AA, Lt BA, MA/MBA or Law/MD/Ph.D.) in columns 1 (males) and 2 (females) and individuals with only HS or GED in columns 3 (males) and 4 (females). Differences between these groups could be obtained due to self-employment being more regular in certain industries that are related to educational years. The results do not provide any indication of this. No significant effect of educational years was found in the subgroup analyses.

Table 13: Subgroup two-stage analyses divided by highest degree obtained

	(1)	(2)	(3)	(4)
Educational years	-0.020	0.060	-0.020	-0.039
	(0.086)	(0.061)	(0.054)	(0.042)
Mother's years of education	0.009	0.006	0.008	0.002
	(0.006)	(0.004)	(0.007)	(0.006)
Father's years of education	-0.003	-0.004	0.006	0.005
	(0.004)	(0.005)	(0.006)	(0.004)
Childhood financial capital scale	0.018	0.024**	0.021	0.002
	(0.014)	(0.012)	(0.013)	(0.010)
2.married, spouse absent	-0.070*	0.024	-0.010	-0.025***
	(0.041)	(0.056)	(0.017)	(0.007)
3.partnered	0.021	0.064	-0.006	-0.033
	(0.044)	(0.040)	(0.030)	(0.026)
4.separated	0.060	-0.052	-0.009	-0.043**
	(0.073)	(0.055)	(0.023)	(0.017)
5.divorced	0.007	-0.043*	-0.041*	-0.046***
	(0.037)	(0.025)	(0.025)	(0.017)
6.separated/divorced	0.022	-0.066**	-0.073**	-0.027
	(0.042)	(0.032)	(0.032)	(0.028)
7.widowed	-0.076**	-0.006	-0.045*	-0.025
	(0.031)	(0.026)	(0.027)	(0.019)
8.never married	-0.059	-0.108**	-0.035	-0.022
	(0.053)	(0.055)	(0.049)	(0.043)
2.catholic	0.006	-0.026	-0.019	-0.013
	(0.032)	(0.026)	(0.027)	(0.020)
3.jewish	0.093	-0.088	-0.019	-0.029
	(0.103)	(0.062)	(0.170)	(0.149)
4.none/no pref	0.008	0.037	0.052	0.082*
	(0.038)	(0.044)	(0.040)	(0.044)
5.other	-0.036	0.106	-0.051	0.185
	(0.099)	(0.171)	(0.146)	(0.178)
Age	0.007***	0.004***	0.002	0.002**
	(0.002)	(0.001)	(0.001)	(0.001)
_cons	-0.489	-0.910	0.714	0.338
	(1.229)	(0.923)	(0.677)	(0.475)
Observations	4928	4663	6441	8659
Birthyear FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes

Standard errors are in parentheses; base-level for marital status is married; base-level for religion is Protestant. 10 Principal components, birthyear fixed effects, and cohort fixed effects are included in the analysis but not in the table
**** p<.01, ** p<.05, * p<.1*

8. Robustness checks

8.1 Reduced form analyses

The reduced form estimator is linear. Due to the binary dependent variable, the estimator acts as a linear probability estimator. The main advantage of this estimation is that the mean effect can be interpreted easily. However, the relationship between the independent and dependent variables will not be linear. Furthermore, a constant marginal effect is assumed while the effect will vary most of the time. The linear form of the model might also lead to predicted probabilities outside of the (0, 1) interval. The robustness of the findings will therefore also be estimated with two non-linear models: the probit and logit model. These models use cumulative probability distribution functions (c.d.f.) and therefore only produce probabilities between 0 and 1.

To show the differences, the reduced form analysis is reported in the table's first two columns; the probit analyses are shown in columns three and four, and the last two columns show the logit analyses.

Table 14: Logit, probit, and linear probability estimation of the reduced form analysis for ADHD

	(1)	(2)	(3)	(4)	(5)	(6)
PGS ADHD	0.013*	0.020***	0.130	0.197***	0.251	0.391**
	(0.008)	(0.006)	(0.110)	(0.045)	(0.168)	(0.173)
Educational years	0.003	-0.005	0.039	-0.040**	0.083	-0.073
	(0.003)	(0.003)	(0.043)	(0.020)	(0.076)	(0.046)
Mother's years of education	0.007**	0.001	0.061	0.012	0.126	0.026
	(0.004)	(0.003)	(0.056)	(0.018)	(0.082)	(0.038)
Father's years of education	0.000	0.002	-0.004	0.011	-0.009	0.025
	(0.003)	(0.002)	(0.038)	(0.014)	(0.060)	(0.036)
Childhood financial capital scale	0.016*	0.004	0.185	0.032	0.404*	0.062
	(0.008)	(0.007)	(0.127)	(0.049)	(0.244)	(0.106)
2.married, spouse absent	-0.036*	-0.007	-1.051*	-0.134	-2.093	-0.163
	(0.021)	(0.030)	(0.635)	(0.581)	(1.385)	(0.993)
3.partnered	0.008	-0.002	0.126	0.031	0.221	0.030
	(0.025)	(0.022)	(0.289)	(0.251)	(0.503)	(0.475)
4.separated	0.014	-0.049*	0.209	-0.848**	0.320	-1.486*
	(0.028)	(0.025)	(0.334)	(0.419)	(0.566)	(0.794)
5.divorced	-0.020	-0.044***	-0.187	-0.623***	-0.481	-1.178***
	(0.021)	(0.014)	(0.284)	(0.148)	(0.508)	(0.333)
6.separated/divorced	-0.037	-0.039*	-0.501	-0.459	-1.035	-0.906
	(0.025)	(0.022)	(0.361)	(0.306)	(0.659)	(0.594)
7.widowed	-0.056***	-0.021	-0.894***	-0.311*	-1.669**	-0.540
	(0.020)	(0.015)	(0.334)	(0.177)	(0.650)	(0.336)
8.never married	-0.047	-0.062*	-0.616	-0.619**	-1.247*	-1.119
	(0.037)	(0.037)	(0.381)	(0.282)	(0.718)	(0.740)
2.catholic	-0.006	-0.018	-0.100	-0.187*	-0.206	-0.360
	(0.020)	(0.016)	(0.401)	(0.110)	(0.415)	(0.267)

3.jewish	0.065	-0.061	0.654	-1.256***	1.104	-2.192**
	(0.087)	(0.068)	(0.000)	(0.484)	(3.102)	(0.882)
4.none/no pref	0.029	0.070**	0.334	0.689***	0.643	1.309**
	(0.027)	(0.028)	(0.506)	(0.221)	(0.805)	(0.538)
5.other	-0.044	0.135	-0.498	1.182	-1.062	2.334
	(0.090)	(0.124)	(0.808)	(4.380)	(1.374)	(3.864)
Age	0.004***	0.003***	0.050***	0.040***	0.094***	0.076***
	(0.001)	(0.001)	(0.012)	(0.011)	(0.024)	(0.022)
_cons	0.187	-0.039	-4.371	-3.890*	-7.977	-7.666
	(0.360)	(0.130)	(13.813)	(2.337)	(0.000)	(5.719)
/lnsig2u			3.040	3.068	4.341	4.178***
			(0.000)	(0.000)	(0.000)	(1.048)
Observations	11369	13322	11338	13215	11338	13215
Pseudo R^2	.z	.z	.z	.z	.z	.z
Birthyear FE	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors are in parentheses; base-level for marital status is married; base-level for religion is Protestant. 10 Principal components, birthyear fixed effects, and cohort fixed effects are included in the analysis but not in the table
**** p<.01, ** p<.05, * p<.1*

The results of the different analyses in the table do not seem to differ substantially. Most significance levels are constant between the different analyses and most importantly the association between ADHD and self-employment is significant for females in all analyses.

Table 15: Logit, probit, and linear probability estimation of the reduced form analysis for autism

	(1)	(2)	(3)	(4)	(5)	(6)
PGS Autism	-0.010	-0.018**	-0.055	-0.193***	-0.095	-0.361**
	(0.012)	(0.009)	(2.820)	(0.073)	(0.215)	(0.158)
Educational years	0.003	-0.005*	0.037	-0.045**	0.066	-0.084*
	(0.003)	(0.003)	(0.191)	(0.020)	(0.056)	(0.044)
Mother's years of education	0.007*	0.001	0.061	0.011	0.117*	0.023
	(0.004)	(0.003)	(1.291)	(0.018)	(0.067)	(0.036)
Father's years of education	0.000	0.002	-0.003	0.010	-0.006	0.022
	(0.003)	(0.002)	(0.598)	(0.014)	(0.048)	(0.033)
Childhood financial capital scale	0.016*	0.003	0.179	0.021	0.365**	0.040
	(0.008)	(0.007)	(4.157)	(0.049)	(0.174)	(0.100)
2.married, spouse absent	-0.036*	-0.006	-1.033	-0.119	-1.998*	-0.140
	(0.021)	(0.030)	(3.595)	(0.581)	(1.210)	(0.992)
3.partnered	0.008	-0.002	0.117	0.032	0.225	0.032
	(0.025)	(0.022)	(1.995)	(0.252)	(0.479)	(0.469)
4.separated	0.014	-0.048*	0.202	-0.842**	0.338	-1.472*
	(0.028)	(0.025)	(1.916)	(0.423)	(0.550)	(0.798)
5.divorced	-0.020	-0.044***	-0.183	-0.620***	-0.441	-1.172***
	(0.021)	(0.014)	(4.144)	(0.149)	(0.472)	(0.316)
6.separated/divorced	-0.037	-0.038*	-0.496	-0.454	-0.969	-0.899
	(0.025)	(0.022)	(3.019)	(0.306)	(0.619)	(0.586)
7.widowed	-0.056***	-0.021	-0.889	-0.309*	-1.599***	-0.539
	(0.020)	(0.015)	(1.668)	(0.177)	(0.603)	(0.334)
8.never married	-0.048	-0.063*	-0.641	-0.632**	-1.240*	-1.134*
	(0.037)	(0.037)	(17.629)	(0.285)	(0.659)	(0.684)
2.catholic	-0.006	-0.017	-0.116	-0.171	-0.189	-0.330
	(0.020)	(0.016)	(3.587)	(0.111)	(0.342)	(0.242)

3.jewish	0.068	-0.056	0.630	-1.198**	1.204	-2.049**
	(0.088)	(0.069)	(0.000)	(0.482)	(2.462)	(0.872)
4.none/no pref	0.029	0.070**	0.325	0.685***	0.644	1.298***
	(0.027)	(0.028)	(17.100)	(0.221)	(0.552)	(0.477)
5.other	-0.046	0.135	-0.518	1.180	-1.024	2.308
	(0.089)	(0.124)	(7.210)	(4.605)	(1.242)	(3.558)
Age	0.004***	0.003***	0.050**	0.040***	0.093***	0.076***
	(0.001)	(0.001)	(0.019)	(0.011)	(0.021)	(0.022)
_cons	0.209	-0.032	-4.144	-3.761	-6.590	-7.260
	(0.359)	(0.129)	(209.097)	(2.296)	(11.381)	(5.074)
/lnsig2u			2.959	3.076	4.184	4.157***
			(0.000)	(0.000)	(0.000)	(0.814)
Observations	11369	13322	11338	13215	11338	13215
Pseudo R^2	.z	.z	.z	.z	.z	.z
Birthyear FE	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors are in parentheses; base-level for marital status is married; base-level for religion is Protestant. 10 Principal components, birthyear fixed effects, and cohort fixed effects are included in the analysis but not in the table
**** p<.01, ** p<.05, * p<.1*

The different methods produce roughly the same results for autism as well. Apart from some control variables the sign and significance are the same. The PGS for autism is negative and significantly associated with self-employment. We can conclude that the results for ADHD and autism are robust for linear and non-linear estimation methods.

8.2 Two-stage analyses

In the methodology section, two assumptions were discussed that could bias the results when violated. These assumptions were the independence and the exclusion restriction. Violation of the independence restriction could result in bias if the parental genes are related to a confounder of the exposure and the dependent variable as shown in figure 2. This assumption cannot be thoroughly tested as it is not possible to know all potential confounders however, the confounders that are measured and mentioned can be tested.

If there is a substantial effect of the PGS on the control variables included, the measured effect indicates that the PGS is not entirely random. The effect of these control variables on entrepreneurship was already measured in the results section. Based on these results the association of the PGS of height with self-reported health, maternal educational years, financial capital, and educational years will be measured. Only the possible associations with maternal educational years and financial capital will be investigated for educational attainment.

Table 16: OLS of the PGS of height on several control variables

	(1) Self-reported health	(2) Mother's years of education	(3) Childhood financial capital	(4) Educational years
PGS Height	-0.005	0.125***	-0.005	0.066***

	(0.005)	(0.016)	(0.006)	(0.013)
_cons	2.298***	11.257***	0.144***	13.725***
	(0.005)	(0.016)	(0.006)	(0.013)
Observations	32550	31029	27458	32453
R-squared	0.000	0.002	0.000	0.001

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

The results show a significant association of the PGS of height with maternal educational and educational years. The associations are stronger than chance would dictate. These two factors also showed a significant effect on self-employment in the two-stage analyses raising concerns about a potential pathway creating bias. Furthermore, the fact that a significant association is found for a PGS of height with two variables concerning education also raises concerns about potential biological pleiotropy for instance with cognitive abilities as mentioned before.

Table 17: OLS of the PGS of educational attainment on maternal education and childhood financial capital

	(1) Mother's years of education	(2) Childhood financial capital
PGS Educational attainment	0.465***	0.095***
	(0.016)	(0.006)
_cons	11.237***	0.141***
	(0.016)	(0.006)
Observations	31029	27458
R-squared	0.026	0.009

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

The results for the PGS of educational attainment show strong significant associations. The strong effect found on a mother's years of education suggests a strong relationship between the genetic potential for the educational attainment of parents and their children indicating that the independence assumption could well be violated. The effect on childhood financial capital is smaller but still relevant and significant. These results suggest that the confounders related to the parental genotype are most likely also related to the genetic potential of the children, opening a pathway for potential bias.

9. Discussion and limitations

9.1 Discussion

Academic research on the genetic background of the entrepreneur has not yet yielded clear associations. In this study genetic data was used to get a better idea of the observable characteristics of the entrepreneur. Significant associations between the genetic risk for autism and ADHD and entrepreneurship were found while educational attainment and height did not have significant effects on entrepreneurship.

The results indicated that men and women differ when it comes to research regarding genetics and entrepreneurship. The genetic risk for both ADHD and autism was only significantly associated with entrepreneurship in the female sample, suggesting that entrepreneurship in females is influenced by heritable traits. This is in line with the research by Zhang et al. (2009) where high heritability estimates were found for females but not for males. Male entrepreneurship was primarily determined by the effects of a shared environment. This indicates that becoming an entrepreneur has different causes for males and females related to genetics and environment and that therefore, the search for the 'entrepreneur' might have to be stratified into the search for the female and male entrepreneur.

This study demonstrated a relationship between ADHD and entrepreneurship, an expected result that was already pointed out before in empirical studies (Dimic & Orlov, 2014). The result was strongly significant for the females in this sample indicating that ADHD symptoms improve the female preference for self-employment. This could potentially be explained by the difference in the severity of symptoms in females with ADHD compared to males (Arnett et al., 2015). Assuming that the genetic risk for ADHD is distributed equally it is interesting to see that the ratio of individuals affected by ADHD is 3:1 at school age (Wilcutt, 2012). The way that ADHD symptoms affect individuals based on gender is beyond the scope of this study, but it is interesting to find that the PGS only influences entrepreneurial selection for females.

Autism showed a negative relationship with entrepreneurial selection for females and no significant relationship for males. The symptoms of ASD might limit individuals from starting to work for themselves. The ever-changing environment entrepreneurs must deal with is not ideal as individuals with ASD are often not comfortable with change. Individuals with autism might work better in environments with modifications and coaching available (Hendricks, 2010). Part of the difference in results for males and females for both behavioral disorders might be caused by using clinical diagnosis in the GWA studies. A higher PGS will not always result in a higher burden of the symptoms.

The two-stage analyses did not reveal significant relationships. Contrary to previous results (Rietveld et al., 2015), height was not a predictor of entrepreneurship in this study. A significant association was expected because of genetic covariance, in this case, the reaction of the environment to tall individuals. Another potential reason behind an association could be the relationship that height has with cognitive abilities (Schick and Steckel, 2015). In this study confounders related to cognitive abilities were controlled for through the type of analysis and inclusion of control variables. It is possible that studies with height as a predictor for entrepreneurship measure the covariance between genetics influencing height and cognitive abilities. Further research is needed to determine if height is a predictor or just related to cognitive abilities.

Except for one subgroup analysis, no results indicated an association between entrepreneurial selection and educational attainment. This is in line with earlier research by Van der Sluis et al. (2008). Education allows individuals to gain skills useful for employment and self-employment and increases the opportunity costs of not taking a well-paid employee job (Parker, 2008; Estrin et al., 2016). The results did suggest that preferences of becoming an entrepreneur may differ for individuals with richer and poorer backgrounds. It could be interesting to investigate the differing attitude toward the risk associated with self-employment based on the financial situation of the family.

The last results that are worth discussing are the significance of the control variables. The effect of age was positive, indicating that a higher age increases the probability of being an entrepreneur. This could be caused by an increased capital to start a business, increased work experience, or a bigger social network. Being married increased the probability of being an entrepreneur compared to being widowed for males and divorced for females suggesting that there might be an effect of more stability at home, running a household with two persons, or even a better financial situation. Furthermore, females without a religious preference were more likely to be entrepreneurs than protestants. This could be explained by the view on classical roles for women and men in religion, especially in the early 20th century.

9.2 Limitations

This study also comes with some limitations. The Mendelian randomization has problems already touched upon in the methodology section. Genotypes are random at conception but conditional on the parents. While controls are included to control for this there may be still some factors related to parental genotype that influence our dependent variable. A possible solution would be to include the parents' genome as a control. The dataset is however not providing this information. The robustness checks showed that there is a substantial risk that the violation of this assumption influences the result for the measured confounders. This also raises concern of potential confounders that were not mentioned and measured.

A bigger problem with Mendelian randomization is the potential violation of the exclusion restriction due to genetic pleiotropy. With the limited knowledge of the genome pleiotropic effects can never be ruled out completely. Examples are the cognitive abilities related to health or for instance the DRD4-gene which has been related to both ADHD and novelty-seeking/extraversion which in their turn have both been associated with entrepreneurship (Gizer et al., 2009; Gelernter et al., 1997). Although attempts of modeling for pleiotropic effects have been made, scholars have not yet succeeded completely with this (DiPrete et al., 2018).

A problem with the reduced form analyses is the lack of data indicating the diagnosis of the behavioral disorders. Behavioral disorders are not diagnosed based on some genetic linear formula but through behavior and experience. It is not possible to say something about specific behavioral traits influencing entrepreneurial selection since PGS based on clinical diagnosis is used and no information on the individuals' symptoms. Future research could focus on distinguishing the specific symptoms of ADHD that influence entrepreneurship. Furthermore, the PGS for the behavioral disorders do not explain a large part of the variance yet.

The study might also lack some generalizability due to the sample that is selected. The sample exists exclusively of individuals above the age of 50 in the United States of America born between 1927 and 1968. This was a different time period when it comes to aspects such as economic development, the labor market, and the view on self-employment. The results will therefore lack representativeness over society in our time and age. Furthermore, due to the risk of population stratification only individuals of European ancestry were used, also decreasing the representativeness of this study.

Another limitation of this study, and other studies related to entrepreneurship, is the lack of a clear definition. Many entrepreneurs are for instance self-employed because they have no other choice in their respective industries. However, the entrepreneurs that are more relevant are the ones innovating and exploiting new opportunities. The use of self-employment might therefore provide information on self-employment but not on the entrepreneurial behavior of interest.

10. Conclusions and recommendations

In this paper, associations are found between ADHD and autism with entrepreneurship. There are no indications of a relationship between height and entrepreneurship. There are some signals that the influence of educational attainment on entrepreneurship works in contradicting ways also related to the socioeconomic background.

Mendelian randomization offers a solution to overcoming confounders when trying to link exposures to certain characteristics. With the pace of genotyping in the last years, it will not be long before rich datasets with two generations of genotyped individuals will be around solving the problems associated with the independence assumption violations. The problem of pleiotropic effects will however still flaw this method. The simplest solution would be to increase the knowledge of genetics with the effect that in the future, it will be easier to estimate the risk of pleiotropic effects. However, this may be easier said than done. Controlling for pleiotropic effects might be easier for less complex traits so specific research on a smaller number of genetic variants can be done. The question is if Mendelian randomization will offer a solution in the search for the entrepreneur.

The final question that is important to consider is why we want to know who the entrepreneur is. Understanding entrepreneurial behaviors and understanding what drives innovation is important for society. Some scholars are of the opinion that the "who" question will not provide useful answers (Ramoglou et al., 2020). Due to the lack of rich data and definitions of entrepreneurship, a small shift in the research on entrepreneurship might be needed to attain useful results. For example, an idea would be to, instead of looking for the entrepreneurial gene, investigate various characteristics to get a better view of the entrepreneurial selection. This way policy can be adjusted to observable and impressionable characteristics. Furthermore, richer data on entrepreneurship is needed to split the types of entrepreneurs. At last, policy evaluations or studies on incentives for entrepreneurship will help with charting entrepreneurial behavior. In a fast-changing world that keeps on innovating and will be in desperate need of sustainable entrepreneurship, the "who" and "what" question must be combined to understand and motivate entrepreneurial behavior.

Reference list:

Adams, S. J. (2002). Educational attainment and health: Evidence from a sample of older adults. *Education Economics*, *10*(1), 97-109.

Anney, R. J., Ripke, S., Anttila, V., Grove, J., Holmans, P., ... & Neale, B. (2017). Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24. 32 and a significant overlap with schizophrenia. *Molecular Autism*, 8, 1-17.

Antshel, K. M. (2018). Attention deficit/hyperactivity disorder (ADHD) and entrepreneurship. *Academy of Management Perspectives*, *32*(2), 243-265.

Arnett, A. B., Pennington, B. F., Willcutt, E. G., DeFries, J. C., & Olson, R. K. (2015). Sex differences in ADHD symptom severity. *Journal of Child Psychology and Psychiatry*, 56(6), 632-639.

Audretsch, D. (2012). Entrepreneurship research. *Management decision*.

Audretsch, D. B., Boente, W., & Tamvada, J. P. (2007). Religion and entrepreneurship.

Awwad, M. S., & Al-Aseer, R. M. N. (2021). Big five personality traits impact on entrepreneurial intention: the mediating role of entrepreneurial alertness. *Asia Pacific Journal of Innovation and Entrepreneurship.*

Balestra, P., & Varadharajan-Krishnakumar, J. (1987). Full information estimations of a system of simultaneous equations with error component structure. *Econometric Theory*, 3(2), 223-246.

Baron, R. A., & Markman, G. D. (2000). Beyond social capital: How social skills can enhance entrepreneurs' success. *Academy of Management Perspectives*, 14(1), 106-116.

Baron, R. A. (2004). The cognitive perspective: a valuable tool for answering entrepreneurship's basic "why" questions. *Journal of business venturing*, 19(2), 221-239.

Barro, R. J., & Lee, J. W. (1994). Sources of economic growth. *In Carnegie-Rochester conference series on public policy* (Vol. 40, pp. 1-46). North-Holland.

Bartels, M., Van den Berg, M., Sluyter, F., Boomsma, D. I., & de Geus, E. J. (2003). Heritability of cortisol levels: review and simultaneous analysis of twin studies. *Psychoneuroendocrinology*, 28(2), 121-137.

Baumol, W. J., & Strom, R. J. (2007). Entrepreneurship and economic growth. *Strategic entrepreneurship journal*, 1(3-4), 233-237.

Benjamin DJ, Cesarini D, Chabris CF, Glaeser EL, Laibson DI, Gudnason V, Lichtenstein P. The promises and pitfalls of genoeconomics. *Annual Review of Economics*. 2012; 4(1):627–662.

Block, J. H., Thurik, R., & Zhou, H. (2013). What turns knowledge into innovative products? The role of entrepreneurship and knowledge spillovers. *Journal of Evolutionary Economics*, 23(4), 693-718.

Boyd, N.G. and Vozikis, G.S. (1994), The influence of self-efficacy on the development of entrepreneurial intentions and actions, *Entrepreneurship Theory and Practice*, Summer, pp.63-77.

Carlson, R., & McChesney, C. (2015). Income sustainability through educational attainment. *The Exchange*, *4*(1).

Case, A., & Paxson, C. (2008). Stature and status: Height, ability, and labor market outcomes. *Journal of political Economy*, *116*(3), 499-532.

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The fourth law of behavior genetics. *Current directions in psychological science*, 24(4), 304-312.

Chen, J.L., Leader, G., Sung, C., Leahy, M. (2015). Trends in Employment for Individuals with Autism Spectrum Disorder: A Review of the Research Literature. *Journal of Autism and Development Disorder*. (2). p. 115–127.

Cheng, Z., & Smyth, R. (2021). Education and migrant entrepreneurship in urban China. *Journal of Economic Behavior & Organization*, *188*, 506-529.

Davey Smith, G., & Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?. *International journal of epidemiology*, *32*(1), 1-22.

Smith, G. D., Lawlor, D. A., Harbord, R., Timpson, N., Day, I., & Ebrahim, S. (2007). Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS medicine*, 4(12), e352.

Davies, N. M., Holmes, M. V., & Smith, G. D. (2018). Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *Bmj*, *362*.

Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., ... Neale, B. M. (2017). Discovery Of The First Genome-Wide Significant Risk Loci For ADHD. *BioRxiv*. https://doi.org/10.1101/145581

Dickson, P. H., Solomon, G. T., & Weaver, K. M. (2008). Entrepreneurial selection and success: does education matter?. *Journal of small business and enterprise development.*

Dimic, N., & Orlov, V. (2014). Entrepreneurial tendencies among people with ADHD. *International Review of Entrepreneurship*, 13(3), 187–204.

DiPrete, T. A., Burik, C. A., & Koellinger, P. D. (2018). Genetic instrumental variable regression: Explaining socioeconomic and health outcomes in nonexperimental data. *Proceedings of the National Academy of Sciences*, 115(22), E4970-E4979.

Dodd, S. D., & Gotsis, G. (2007). The interrelationships between entrepreneurship and religion. *The international journal of entrepreneurship and innovation*, 8(2), 93-104.

Estrin, S., Mickiewicz, T., & Stephan, U. (2016). Human capital in social and commercial entrepreneurship. *Journal of business venturing, 31(4), 449-467.*

European Commision. (2020) Unleashing the full potential of European SMEs. Retrieved from: https://ec.europa.eu/commission/presscorner/detail/en/fs_20_426

Frese, M. (2009), Toward a psychology of entrepreneurship - An action theory perspective, *Foundations and Trends in Entrepreneurship*, Vol. 5, No. 6, pp.435-494. DOI: 10.1561/0300000028

Fox, N. A., Henderson, H. A., Marshall, P. J., Nichols, K. E., & Ghera, M. M. (2005). Behavioral inhibition: Linking biology and behavior within a developmental framework. *Annual Review of Psychology.*, 56, 235-262.

Gartner, W. B. (1988). "Who is an entrepreneur?" is the wrong question. *American journal of small business*, 12(4), 11-32.

Gelernter, J., Kranzler, H., Coccaro, E., Siever, L., New, A., & Mulgrew, C. L. (1997). D4 dopamine-receptor (DRD4) alleles and novelty seeking in substance-dependent, personality-disorder, and control subjects. *The American Journal of Human Genetics*, 61(5), 1144-1152.

Gizer, I. R., Ficks, C., & Waldman, I. D. (2009). Candidate gene studies of ADHD: a meta-analytic review. *Human genetics*, 126(1), 51-90.

Goodman, D. W. (2007). The consequences of attention-deficit/hyperactivity disorder in adults. *Journal of Psychiatric Practice®*, *13*(5), 318-327.

Gratten, J., Wray, N. R., Keller, M. C., & Visscher, P. M. (2014). Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nature neuroscience*, 17(6), 782-790.

Grizenko, N., Fortier, M. E., Zadorozny, C., Thakur, G., Schmitz, N., Duval, R., & Joober, R. (2012). Maternal stress during pregnancy, ADHD symptomatology in children and genotype: gene-environment interaction. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 21(1), 9.

Hamer, D., Sirota, L. (2000). Beware the chopsticks gene. *Mol Psychiatry* 5, 11–13 https://doi.org/10.1038/sj.mp.4000662

Hazel Wade, K. (2021) Directed acylic graph (DAG) for illustrating the three core assumptions for Mendelian randomization framework. Retrieved from: https://commons.wikimedia.org/wiki/File:Directed_acylic_graph_for_Mendelian_randomization_Wikipedia_page.png

Heineck, G. (2005). Up in the skies? The relationship between body height and earnings in Germany. *Labour*, *19*(3), 469-489.

Hendricks, D. (2010). Employment and adults with autism spectrum disorders: Challenges and strategies for success. *Journal of vocational rehabilitation*, *32*(2), 125-134.

Henrekson, M., & Sanandaji, T. (2014). Small business activity does not measure entrepreneurship. *Proceedings of the National Academy of Sciences*, 111(5), 1760-1

Hoffmann, A., Junge, M., & Malchow-Møller, N. (2015). Running in the family: parental role models in entrepreneurship. *Small Business Economics*, 44(1), 79-104.

Holwerda, A., van der Klink, J. J., de Boer, M. R., Groothoff, J. W., & Brouwer, S. (2013). Predictors of sustainable work participation of young adults with developmental disorders. *Research in developmental disabilities*, 34(9), 2753-2763.

Jennings, J. E., & McDougald, M. S. (2007). Work-family interface experiences and coping strategies: Implications for entrepreneurship research and practice. *Academy of Management Review*, 32, 747–760.

Johnson, G. (2015, December 1). Self-employment opens new worlds for autistic People. *The Globe and Mail*. Retrieved from http://www.theglobeandmail.com

Kendler, K. S., & Eaves, L. J. (1986). Models for the joint effect of genotype and environment on liability to psychiatric illness. *The American journal of psychiatry*.

Kerr, S. P., Kerr, W. R., & Xu, T. (2018). Personality traits of entrepreneurs: A review of recent literature. *Foundations and Trends® in Entrepreneurship, 14*(3), 279-356.

Koellinger, P. D., van der Loos, M. J., Groenen, P. J., Thurik, A. R., Rivadeneira, F., van Rooij, F. J., ... & Hofman, A. (2010). Genome-wide association studies in economics and entrepreneurship research: promises and limitations. *Small Business Economics*, 35(1), 1-18.

Koellinger, P. D., & De Vlaming, R. (2019). Mendelian randomization: the challenge of unobserved environmental confounds. *International journal of epidemiology*, 48(3), 665-671.

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., ... & Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science*, 359(6374), 424-428.

Kwon, J. M., & Goate, A. M. (2000). The candidate gene approach. *Alcohol Research & Health*, *24*(3), 164.

Le, A. T. (1999). Empirical studies of self-employment. *Journal of Economic surveys*, 13(4), 381-416.

Lee, J.J. et al. Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Forthcoming in Nat. Genet*. (2018).

Loehlin, J. C., McCrae, R. R., Costa, P. T., Jr., & John, O. P. (1998). Heritabilities of common and measure-specific components of the Big Five personality factors. *Journal of Research in Personality*, 32(4), 431–453. https://doi.org/10.1006/jrpe.1998.2225

Lousdal, M. L. (2018). An introduction to instrumental variable assumptions, validation and estimation. *Emerging themes in epidemiology*, *15*(1), 1-7.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., ... & Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), 747-753.

McClellan, J., & King, M. C. (2010). Genetic heterogeneity in human disease. *Cell*, 141(2), 210-217.

MedlinePlus [Internet]. Bethesda (MD): National Library of Medicine (US); [updated 2020 Jun 24]. SNP; [updated 2020 Jun 18; reviewed 2018 Jun 01; cited 2022 Jun 2]; [about 5 p.]. Available from: https://medlineplus.gov/genetics/understanding/genomicresearch/snp/.

Molina, J. A. (2020). Family and entrepreneurship: New empirical and theoretical results. *Journal of Family and Economic Issues*, 41(1), 1-3.

Murphy, P. J., Liao, J., & Welsch, H. P. (2006). A conceptual history of entrepreneurial thought. *Journal of management history*.

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., Cherkas, L., Hunkin, J., & Spector, T. D. (2008). Is the tendency to engage in entrepreneurship genetic?. *Management Science*, *54*(1), 167-179.

NIMH. (2021) Attention-Deficit/Hyperactivity Disorder. National Institute of Mental Heatlh, retrieved from: https://www-nimh-nih-gov.eur.idm.oclc.org/health/topics/attention-deficit-hyperactivity-disorder-adhd#part_2292.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., Rustichini, A. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, 533(7604), 539-542.

Paaby, A. B., & Rockman, M. V. (2013). The many faces of pleiotropy. *Trends in genetics*, 29(2), 66-73.

Parker, S. C. (2008). Entrepreneurship among married couples in the United States: A simultaneous probit approach. *Labour Economics*, *15*(3), 459-481.

Parker, S. C. (2018). The economics of entrepreneurship (2nd ed.). *Cambridge: Cambridge University Press*. 167–179.

Patel, P. C., Rietveld, C. A., & Verheul, I. (2021). Attention deficit hyperactivity disorder (ADHD) and earnings in later-life self-employment. *Entrepreneurship Theory and Practice*, *45*(1), 43-63.

Pearson, T. A., & Manolio, T. A. (2008). How to interpret a genome-wide association study. *Jama*, 299(11), 1335-1344.

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological bulletin*, *84*(2), 309.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8), 904–909. https://doi.org/10.1038/ng1847

Quaye, L., Nicolaou, N., Shane, S., & Mangino, M. (2012). A discovery genome-wide association study of entrepreneurship. *International Journal of Developmental Science*, 6(3-4), 127-135.

Ramoglou, S., Gartner, W. B., & Tsang, E. W. (2020). “Who is an entrepreneur?” is (still) the wrong question. *Journal of Business Venturing Insights*, 13, e00168.

Rietveld, C. A., Hessels, J., & van der Zwan, P. (2015). The stature of the self-employed and its relation with earnings and satisfaction. *Economics and Human Biology*, 17(1), 59–74.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., ... & Preisig, M. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *science*, 340(6139), 1467-1471.

Rietveld, C. A., & Patel, P. C. (2019). ADHD and later-life labor market outcomes in the United States. *The European Journal of Health Economics*, 20(7), 949-967.

Rietveld, C. A., Slob, E. A., & Thurik, A. R. (2021). A decade of research on the genetics of entrepreneurship: a review and view ahead. *Small Business Economics*, 57(3), 1303-1317.

Schick, A., & Steckel, R. H. (2015). Height, human capital, and earnings: The contributions of cognitive and noncognitive ability. *Journal of Human Capital*, *9*(1), 94-115.

Schumpeter, J. A. (2000). Entrepreneurship as innovation. *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship.*

Shane, S., & Venkataraman, S. (2000). The promise of entrepreneurship as a field of research. *Academy of Management Review*, 25(1), 217-226

Shi, H., Kichaev, G., & Pasaniuc, B. (2016). Contrasting the genetic architecture of 30 complex traits from summary association data. *The American Journal of Human Genetics*, 99(1), 139-153.

Singh, R. P., & Crump, M. E. (2007). Educational attainment: A key factor for improving the lagging rate of black entrepreneurship. *The Review of Black Political Economy*, 34(3-4), 217-229.

Small, D., Tan, Z., Lorch, S., & Brookhart, A. (2014). Instrumental variable estimation when compliance is not deterministic: the stochastic monotonicity assumption. *arXiv preprint* arXiv:1407.7308.

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., & Smoller, J. W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics*, 14(7), 483-495.

Sugrue LP, Desikan RS. (2019). What Are Polygenic Scores and Why Are They Important? *JAMA.* 2019;321(18):1820–1821. doi:10.1001/jama.2019.3893

Terán-Yépez, E., Marín-Carrillo, G. M., del Pilar Casado-Belmonte, M., & de las Mercedes Capobianco-Uriarte, M. (2020). Sustainable entrepreneurship: Review of its evolution and new trends. *Journal of Cleaner Production*, 252, 119742.

Thompson, O. (2014). Economic background and educational attainment the role of gene-environment interactions. *Journal of Human Resources*, 49(2), 263-294.

Tops, M., & Boksem, M. A. (2011). Cortisol involvement in mechanisms of behavioral inhibition. *Psychophysiology*, 48(5), 723-732.

Torkamani, A., Wineinger, N. E., & Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics*, 19(9), 581-590.

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current directions in psychological science*, 9(5), 160-164.

Vable, A. M., Gilsanz, P., Nguyen, T. T., Kawachi, I., & Glymour, M. M. (2017). Validation of a theoretically motivated approach to measuring childhood socioeconomic circumstances in the Health and Retirement Study. *PloS one*, 12(10), e0185898.

Van der Loos, M. J., Rietveld, C. A., Eklund, N., Koellinger, P. D., Rivadeneira, F., Abecasis, G. R., ... & Thurik, A. R. (2013). The molecular genetic architecture of self-employment. *PloS one*, 8(4), e60542.

Van der Sluis, J., Van Praag, M., & Vijverberg, W. (2005). Entrepreneurship selection and performance: A meta-analysis of the impact of education in developing economies. *The World Bank Economic Review*, 19(2), 225-261.

Van der Sluis, J., Van Praag, M., & Vijverberg, W. (2008). Education and entrepreneurship selection and performance: A review of the empirical literature. *Journal of economic surveys*, 22(5), 795-841.

Verheul, I., Rietdijk, W., Block, J., Franken, I., Larsson, H., & Thurik, R. (2016). The association between attention-deficit/ hyperactivity (ADHD) symptoms and self-employment. *European Journal of Epidemiology*, 31(8), 793–801.

Von Hinke, S., Smith, G. D., Lawlor, D. A., Propper, C., & Windmeijer, F. (2016). Genetic markers as instrumental variables. *Journal of Health Economics*, 45, 131-148.

Wiklund, J., Davidsson, P., Audretsch, D.B. and Karlsson, C. (2011), "The future of entrepreneurship research", *Entrepreneurship Theory and Practice,* Vol. 35 No. 1, pp. 1-9.

Willcutt, E. G. (2012). The prevalence of DSM-IV attention-deficit/hyperactivity disorder: a meta-analytic review. *Neurotherapeutics,* 9(3), 490-499.

Wright, P.G. (1928). The Tariff on Animal and Vegetable Oils. *New York: Macmillan*.

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., ... & Visscher, P. M. Consortium G (2018) Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet*, *27*(20), 3641-3649.

Zhang, Z., Zyphur, M. J., Narayanan, J., Arvey, R. D., Chaturvedi, S., Avolio, B. J., et al. (2009). The genetic basis of entrepreneurship: effects of gender and personality. *Organizational Behavior and Human Decision Processes*, 110(2), 93–107.

Appendix

Table A: Within variance of height

Variable		Mean	Std. dev.	Min	Max	Observations
Height	overall	1.709	0.099	1.245	2.261	N = 32553
	between		0.098	1.417	2.057	n = 7915
	within		0.010	1.484	2.128	T-bar = 4.11282

Table B: Tabulation of educational years

Educational years	Freq.	Percent	Cum.
0.none	18	0.06	0.06
2	11	0.03	0.09
3	25	0.08	0.17
4	22	0.07	0.23
5	23	0.07	0.31
6	45	0.14	0.44
7	110	0.34	0.78
8	397	1.22	2.01
9	420	1.29	3.30
10	907	2.79	6.09
11	906	2.79	8.89
12	10644	32.80	41.68
13	2638	8.13	49.81
14	4247	13.09	62.90
15	1512	4.66	67.56
16	4985	15.36	82.92
17.17+ yrs	5543	17.08	100.00
Total	32453	100.00	