CERGE - EI
Center for Economic Research and Graduate Education –
Economics Institute

Essays in Genoeconomics

Jaroslav Groero

Dissertation

Prague 2023

Dissertation Committee

Doc. Nikolas Mittag, Ph.D. (CERGE-EI; chair)

prof. Ing. Štěpán Jurajda, Ph.D. (CERGE-EI)

Doc. PhDr. Michal Bauer, Ph.D. (CERGE-EI)

Referees

Dr Selma Walther (Department of Economics, University of Sussex)

prof. Pietro Biroli (Department of Economics, University of Bologna)

Table of Contents

Abstract

This thesis explores gene-environment interaction models, which comprise a new and rapidly developing field in the empirical economics literature. I study how investments and environments complement or substitute genetic predispositions in various settings. The first chapter shows that one additional year of education moderates the role of genetic predispositions for important medical conditions and diseases. The second chapter documents that adverse macroeconomic conditions negatively affect risk tolerance for individuals with low genetic predisposition for risk tolerance. At the same time I show that these conditions have no significant effect for individuals with genetic predispositions to be risk tolerant. Finally, the third chapter discusses problems in the methodology of the current gene-environment models and proposes a new approach that addresses them.

Abstrakt

Ve své dizertaci se zabývám novou generací empirických modelů zkoumající interakci genů a prostředí, jejichž význam v ekonomické literatuře v posledních letech značně vzrostl. Se své práci analyzuji, jak investice a prostředí komplementují nebo substituují genetické predispozice a to v široké škále situací. V první kapitole ukazuji, že jeden dodatečný rok vzdělání snižuje vliv genetických predispozic pro závažné choroby. Ve druhé kapitole ukazuji, že nepříznivé makoronomické podmínky mohou negativně olivnit rizikovou toleranci jednotlivců s vysokými genetickými predispozicemi pro nízkou rizikovou toleranci. Zároveň ukazuji, že stejné makroekonomicé podmínky nemají žádný vliv na jedince s genetickými predispozicemi pro vysokou rizikovou toleranci. V poslední třetí kapitole diskutuji problémy současné metodologie modelů zkouající interakci genů a prostředí a navrhuji novou metodu, která se tyto problémy snaží vyřešit.

Acknowledgments

I hereby express my sincere and utmost gratitude to my supervisor, Nikolas Mittag, for his patience, support, and time. I appreciate his guidance and exceptional expertise, which helped me to start and finish my dissertation research.

I am also grateful to the members of my dissertation committee, Štěpán Jurajda and Michal Bauer, for their time and useful feedback, which helped me to improve the quality of this dissertation.

Thanks are also due to Pietro Biroli, who helped me with the introduction into the field of genoeconomics, and provided me with useful insights and feedback, and to the CERGE-EI academic skills center for their help with final edits of this dissertation.

Finally, I am grateful to my parents for their help and support throughout my studies.

This research has been conducted using data from the UK Biobank, a major biomedical database. This research has been conducted using the UK Biobank Resource under Application Number 48293.

Czech Republic, Prague — Jaroslav Groero
Thursday 11th May, 2023

Introduction

Social scientists have long been interested in whether the role of predetermined innate conditions can be influenced by environments, choices, or investments (Manski, 2011, Heckman, 2007). However, due to a lack of proper genetic data empirical research on this matter was not feasible in the past. Recent decline in the cost of genetic data allowed inclusion of this information in many socio-economic surveys, and opened the doors to studies that can address these important questions.

This thesis contributes to the literature by investigating how environments and investments influence the role of genetic predispositions in the formation of important social and economic outcomes in two settings. First, I investigate how one year of additional schooling in adolescence moderates the role of genetic predispositions with respect to health conditions and diseases at later stages of life. I show that one year of additional schooling in adolescence decreases the risk of experiencing a heart attack and cancer by about 40 %. In the second setting I investigate how macroeconomic conditions affect individual risk preferences and how genetic predispositions moderate this relationship. I document that adverse macroeconomic conditions negatively affect willingness to take risks for individuals with low genetic predispositions for risk tolerance. At the same time, I do not

find any evidence of the effect of macroeconomic conditions on risk tolerance for individuals with high genetic predispositions for risk tolerance.

To estimate the GxE model, I adopt a methodology from the current gene-environment (GxE) literature and estimate a GxE class of models. The thesis also discusses problems of the current GxE models and shows that under a typical scenario the results may lead to measurement error bias. To fix the problem, I propose a new two-step method, based on a split-sample approach (Wasserman and Roeder, 2009), that alleviates this issue under the assumption that the first step serves as a proper variable selection stage.

Chapter 1

Leveling Health Inequalities: Raising the School Leaving Age Reduces the Risk of Diseases and Severe Medical Conditions Related to Genetic Endowment

Abstract

Health inequality has a significant genetic component and socio-economic factors, including education, can moderate the effects of genes. However, little is known about whether more years of education can effectively moderate the relationship between genetic conditions and severe contemporary diseases and medical conditions. I use UK Biobank data to investigate the relationship between education, genetic endowment, and four health conditions: heart attack, cancer, stroke, and type-2 diabetes. To avoid the potential endogeneity of education, I focus on the long-term health consequences of a 1972 increase in the UK school-leaving age (ROSLA). As a measure of genetic endowment, I use an index of genetic predispositions for obesity. Genetic predispositions are typically summarised by a weighted average of individual genetic markers called polygenic scores (PGS), where weights are derived from analyses performed on different populations. Furthermore, the

outcomes of these analyses often differ from the outcomes the PGS is used to predict. This may skew the results of follow up studies of other outcomes, including cancer. I introduce a two-step method that adjusts the available weights to new outcomes, and show that genetic predisposition for obesity increases the risks of the four diseases I study. The results based on my new method show that the additional year of schooling driven by the ROSLA reform diminished the importance of genetic predispositions for the risks of cancer and heart attack by 40%. The results offer new evidence on how environments moderate the inequalities in health that have been tilted from birth.

JEL classification: I12,I14, I28
Keywords: health, disease, education reform, polygenic scores, UK Biobank

Introduction

Health outcomes are an important influence on both the quality and length of life. Accordingly, better understanding of the roots of health inequalities may help to diminish disparities in wellbeing that are attributable to health conditions. It has been established that later-life inequality in health outcomes has roots in the prenatal and early-life developmental stages (Almond et al., 2018, Rosales-Rueda, 2018, Almond et al., 2018). Genetic predispositions are important innate factors that have been shown to explain a large proportion of health disparities (Visscher et al., 2012, 2017), though human diseases differ in terms of the extent to which genes affect them.

While some diseases are determined mostly or solely by genetic endowment, such as Huntington's disease (Bates et al., 2015), most contemporary diseases are shaped by both genetic and environmental factors. By environmental factors I mean aspects of life that can be controlled by the individual, such as lifestyle, and those that cannot be controlled by individuals, such as macroeconomic development or pollution. For example, it has been established that genetic components play an important role in predicting susceptibility to health conditions, including prostate cancer (Conti et al., 2021),breast cancer (Michailidou et al., 2017), Alzheimer's disease (Bone et al., 2021), heart disease (Hartiala et al., 2021), and many other conditions that can be influenced by environment and lifestyle (Almond et al., 2018, Dixon, 2010, Hubert et al., 1983).

Thus, a substantial part of health inequality stems from innate conditions that cannot be changed and are an outcome of the lottery of nature. Although the genetic endowment cannot be changed, its influence on health and other outcomes can be modified throughout life. It follows that health prevention is more important for individuals who have higher genetic predispositions to develop severe medical conditions and diseases. Empirical evidence suggests that for many diseases an individual's genetic endowment represents initial disparities that can be moderated by behavior later in life (Turkheimer et al., 2003). Hence, for most contemporary diseases, genetic predispositions do not play a deterministic role, but rather their influence on individual outcomes can be shaped by lifestyle, investments, or other environmental factors.

I use UK Biobank data to investigate how education moderates the relationship between genetic endowment and severe medical conditions. My research sheds light on

how more years of schooling can help to decrease health inequalities attributed to genetic conditions. Early-life investment into education is an important environmental factor that has been shown to affect later-life health outcomes (e.g. Barcellos et al., 2019, Albouy and Lequien, 2009, Grossman, 1972, Glied and Lleras-Muney, 2003, Clark and Royer, 2013, Barcellos et al., 2018). Recent evidence suggests that higher education moderates the relationships between genetic endowment and obesity and other health indicators (Barcellos et al., 2018). However, little is known about how education moderates the relationship between genetic conditions and severe diseases and medical conditions.

Similarly toBarcellos et al. (2018), I use an index of genetic predispositions for obesity as a measure of genetic conditions. I use the index to investigate inequality in severe diseases and medical conditions, including cancer and heart disease, to uncover new patterns in the sources of health inequalities. Previous research has established that obesity can lead to health problems including heart disease, cancers, and premature death (e.g Dixon, 2010, Hubert et al., 1983) and that obesity is partly determined by genetic components (Locke et al., 2015). Importantly, as research in genetics suggests, genetic regions usually affect more than just one outcome; a notion called pleiotropy (Mills et al., 2020). I demonstrate that genetic markers that have been shown to explain variation in obesity are also related to other, more severe, health problems. This finding documents that a genetic correlation exists between obesity and severe illnesses (Zhao and Zhu, 2021, Bulik-Sullivan et al., 2015).

A potential problem of incorporating genetic data into economic research is their large dimensionality. Consequently, social science researchers usually rely on large consortia to provide estimates of individual genetic effects, which they then use to construct indexes of genetic predispositions (Purcell et al., 2009). This procedure has many drawbacks, including the fact that estimates provided by consortia are usually based on different modelling assumptions and obtained from different populations. Importantly, the numbers of outcomes a social-science researcher can effectively explore are usually constrained by the quality and availability of the results the genetic consortia provide. I develop and apply a new method designed to study gene-environment (GxE) models in situations in which there is not enough information to construct the measure of genetic predispositions for the outcome of interest. Currently researchers often use a polygenic score (PGS) index as a measure of genetic predispositions in GxE models (see e.g. Janssens

et al., 2006, Purcell et al., 2009, Belsky and Harden, 2019). The PGS is a weighted average of individual genetic markers, single-nucleotide polymorphisms (SNPs, read as SNIPs), where the weights are based on estimates from genome-wide association studies (GWAS), which are conducted by large scientific consortia. The GWAS estimates of the relationship between outcomes and SNPs are usually used as weights when constructing the PGS.

I use the results from an obesity GWAS (Locke et al., 2015) and investigate the relationship between genetic predisposition for obesity and severe medical condition outcomes other than obesity. Thus, although the individual SNPs that affect obesity may also affect other health outcomes, it is likely that the estimated GWAS weights for obesity do not carry over to empirical models with outcomes that are not a measure of body size or obesity. Consequently, if the outcomes of the GWAS population and the population of interest differ, when GxE models that rely on GWAS estimates of population genetic effects are used in economics they may result in skewed estimates of important parameters of standard GxE models.

The method I develop allows me to expand the scope of existing GWAS by not making any specific assumption about the similarity of the outcomes in GWAS samples and the survey sample used to estimate the GxE model. The method is well suited for outcomes for which there are no GWAS or for which existing GWAS do not provide high quality estimates. It is also well suited to models that aim to study the roles of cross-trait genetic predispositions in the formation of illnesses, including SNPs related to obesity. I use a two step nonlinear approach that does not rely on GWAS coefficients but rather estimates the individual SNP coefficients together with the main coefficients of interest. This approach alleviates the problem of imperfect portability of GWAS weights. Furthermore, my method solves additional issues of GWAS model mis-specification that arise if the true genetic effects are heterogeneous.

In this paper, I focus on four severe medical conditions: cancer, heart attack, stroke, and diabetes. The major advantage of studying genes over other types of initial conditions is that they are fixed at conception and cannot be altered later in life. Hence, to some extent, they can be viewed as a natural experiment that creates a random variation in initial conditions. However, the challenge is that years of schooling are most likely endogenous to the model, which makes it cumbersome to identify the causal inference of

GxE models. For this reason I investigate the long term health consequences of raising the minimum age at which a student can leave school (ROSLA), a policy introduced in the UK in 1972.

This paper presents new evidence that suggests the ROSLA policy decreased the disparities related to the genetic endowment for obesity in severe diseases and medical conditions. Specifically, I show that the policy moderated the relationship between the probabilities of developing cancer and suffering a heart attack and the genetic endowment by 40%. Although the percentage point change is rather high, it should be noted that this is a product of two point estimates. Moreover, this policy targeted the adolescent population, which is more likely to respond to an increase in length of education than older individuals (Heckman, 2007). Moreover, my results confirm the previous findings of Barcellos et al. (2018) that the ROSLA policy moderates the relationship between genetic endowment and high values of health indicators.

Together with the evidence of the attenuation of the role of genetic endowment for diseases, my findings imply that individuals start to perceive health indicators as a negative influence on their lives only after their value reaches a certain threshold. This implies that the health costs of high values of intermediary health indicators are not linear. Hence, individuals may prefer to act on their health indicator values only when the value is high enough that they clearly have a greater risk of developing a disease or a severe medical condition. I also document a positive relationship between genetic predispositions for obesity and heart attacks, strokes, cancer, and type 2 diabetes. This finding shows that genetic markers related to obesity also affect serious diseases and medical conditions, thereby shedding more light on the genetic sources of health inequalities. Finally, my results suggest that using my new method to correct for the PGS index weights matters most when the outcome of interest of the GxE study does not correspond to the GWAS outcome.

The rest of the paper is organized as follows. Section 1.1 documents the role of genetic data in economic research and identifies some potential challenges. Section 1.2 presents the problems connected with modelling the GxE using standard methods. Section 1.3 presents the analytical sample and the variables of interest. Section 1.4 proposes a new two stage procedure that I use in this paper, which is based on a nonlinear least squares estimator. Section 1.5 describes the results of my analysis of the long-term

health effects of the ROSLA policy and shows new empirical evidence about the role of education in decreasing the influence of genetic conditions on severe illnesses and medical conditions. Section 1.6 presents robustness checks. Section 1.7 concludes and summarizes the findings.

1.1 Genetic Data in the Social Sciences

For many decades, scientists have discussed the respective roles of genes and environments in the formation of human traits. Recent research suggests that the traditional dichotomy between genetic and environmental components of trait formation is outdated and imprecise, and new models have been proposed in which genes and environments interact in the formation of important human outcomes (phenotypes) (Turkheimer, 2000, Turkheimer et al., 2003, Rutter, 2006). Consequently, researchers have started to examine what parts of human DNA are correlated with specific outcomes. This type of analysis was not feasible in the past due to the scarcity and high cost of genetic data. However, contemporary researchers in large consortia now conduct large genome-wide association studies (GWAS) to establish robust correlations between genetic markers and outcomes, including education level, (Okbay and Rietveld, 2015, Lee et al., 2018), obesity (Locke et al., 2015), risk aversion (Linnér et al., 2018), height (Yengo et al., 2018), and many other important behavioral and health outcomes.

The most common genetic variant that researchers investigate are single-nucleotide polymorphisms (SNPs). Each SNP represents a position on the DNA called a nucleotide, which varies across individuals. By virtue of being diploid organisms, humans have 2 versions of each SNP (one per chromosome). Thus, SNPs are represented in genetic data by variables that can take on only three values: 0,1, or 2. The specific realization depends on how many risky alleles a person has at a given SNP[1]. The GWAS results are then used to investigate more deeply the role of genes, environments, and their interaction in the formation of important human behavioral and health outcomes.

Most current empirical studies that work with genetic data use GWAS summary statistics to construct a single index that represents the individual genetic propensity for

[1]By risky allele I mean a specific realisation of a SNP that contributes to an outcome. For more information about genetic markers, see Mills et al. (2020)

a certain trait. In a recent study, Chabris et al. (2015) argues that all behavioral traits are polygenic in their nature, which means that most human outcomes are generally affected by many genetic markers with small-effect sizes. This presents a problem for empirical studies, since there are billions of SNPs that need to be tested and often hundreds or thousands of SNPs contribute to any given outcome. Therefore, scientists use a score, often called a polygenic score (PGS), to decrease the dimensions of genetic data (see e.g. Janssens et al., 2006, Belsky and Harden, 2019). A PGS index is simple to construct and easy to implement and is used in empirical studies, which usually operate with small sample sizes (e.g. 10,000 samples). To construct the PGS, researchers use the SNP coefficients estimated in a GWAS, together with a survey's SNP data in the following way:

$$PGS_i = \sum_{j}^{J} \gamma_j^{GWAS} SNP_{j,i}$$

where γ_j are the GWAS coefficients, $SNP_{j,i}$ is a particular realization of SNP j for individual i in a survey, and J stands for the total number of SNPs in the survey. The advantage of using the single index score is that it allows estimation of the usual empirical economic models without the need to estimate the individual SNP coefficients, which in most cases would be infeasible due to the large dimensions of genetic data.

Although PGS are widely used, they have several shortcomings when applied to economic models. Recently, Mostafavi et al. (2020) show that the predictive power of a PGS depends on the specific sample it is applied to. This suggests that GWAS results are not generally applicable across samples when it comes to PGS creation. Moreover, Becker et al. (2021) shows that the PGS score weights derived from a large scale GWAS may lead to measurement bias in the PGS. Moreover, social science researchers are often limited to studying outcomes for which there are high quality GWAS. However, in some cases, it is desirable to study related outcomes that do not necessarily correspond to those for which there are GWAS results.

I study sources of heterogeneity in health outcomes that are distant from the outcomes of the obesity GWAS, which I use to create a measure of genetic endowment. It has been established that genetic markers may influence various outcomes (Mills et al., 2020). Thus, in many cases, the SNPs that affect the outcome of interest to a researcher and those that affect the outcome of a GWAS overlap. Furthermore, in many cases, the

GWAS results for an outcome of interest are missing or are not of very high quality, though there are results for a related outcome that can be used. In all these cases, the outcome of interest to a researcher differs from the outcome of the GWAS. Hence, even though the SNPs overlap, the GWAS coefficients of these SNPs are not likely to be valid for the construction of a PGS, which leads to a measurement error bias in the PGS. It is well known that measurement error can worsen the performance of standard empirical models (Meyer et al., 2020, Meyer and Mittag, 2017), which indicates that PGS measurement error caused by incorrect PGS weights may be one reason for the results the of Mostafavi et al. (2020). In section 1.2 I elaborate on this problem further. In section 1.4, I present a novel approach that solves the problems of the current method.

In economics, an important application of genetic data, and specifically the PGS, is its use in gene-environment models (GxE). It is crucial to understand how individual decisions and environmental factors interact with genes in order to understand the biological and social architecture of economic outcomes, and to inform the design of policies that can reduce disparities in these key determinants of wellbeing, thus leveling a playing field that is often tilted from birth by genetic differences. The effects of genes on social outcomes arise from an interplay between differences in initial conditions, differences in the environment the individual is exposed to, and endogenous investment choices that respond to the environment (Rosales-Rueda, 2014, Sanz-de Galdeano and Terskaya, 2019, Boneva and Rauh, 2018). The interplay between genes and the environment, or nature *and* nurture, has long been debated in the medical (Bickel et al., 1953) and genetic literature (Plomin, 1990). See e.g. Turkheimer et al. (2003), Rutter (2006), Ridley (2003), Barcellos et al. (2018), Biroli (2015a), Liu and Guo (2015), Schmitz and Conley (2016b), Domingue et al. (2015), Wedow et al. (2018) and Bierut et al. (2018) for recent discussions and empirical examples. I extend the current literature by analyzing how an increase in education may decrease health inequality that has been unequal from birth.

GxE models are also very closely related to the economic life-cycle literature. For instance, Cunha and Heckman (2007) and Cunha et al. (2010) show that initial conditions have an impact on a variety of skills. Moreover, this literature presents evidence of a dynamic complementarity of investments, which means that investments are more effective in building higher values of a skill in subsequent periods for higher values of that skill in the current periods (Cunha and Heckman, 2007). This finding is consistent with

gene-environment interactions because it implies that investment in a skill leads to higher returns for individuals with stronger genetic predispositions for that skill. Additionally, recent genetic literature shows that parental genes also affect a child's outcome indirectly through investments (Trejo and Domingue, 2019, Kong et al., 2018).

To my best knowledge, the GxE models are currently the best class of empirical models with the potential to study the dynamic complementarity between investments and initial genetic conditions, self-productivity, and cross-productivity (Cunha and Heckman, 2007). These concepts are crucial to understanding the formation of important outcomes such as skills or health. Ideally, the gene-environment interaction term in a GxE model provides a direct estimate of the dynamic complementarity, because it provides information on how effective an investment is likely to be based on initial genetic conditions.

Furthermore, due to the wide range of GWAS summary statistics, it is possible to estimate the cross-productivity of the genetic predispositions for different skills or outcomes. For instance, Barcellos et al. (2018) estimate the effect of a PGS for education on body size and blood pressure together with the interaction effect of the PGS with an additional year of schooling. I study cross-productivity by investigating how genetic predispositions towards obesity influence the probability of an individual experiencing incidences of cancer, heart attack, stroke, and diabetes later in life. Additionally, I present evidence on the negative complementarity of initial genetic predispositions towards obesity and years of schooling. Thus, this study explores how education moderates the role of genetic predispositions in health formation. Due to data limitations, I am not able to shed more light on the channels through which education can mediate the gene-health relationship. However, there are several possible paths. For instance, higher education may lead to better jobs which may lead to higher social status and income, which in turn contribute to better health (Deaton, 2008). Alternatively, more education may lead to more information about the consequences of an unhealthy lifestyle, which may promote healthier lifestyles.

1.2 The Current PGS approach in GxE models

Although studying the roles of genes and their interactions with socio-economic variables is an important strand of research, connecting biological markers including genes to

social-science outcomes raises many conceptual and practical challenges. The arguments above suggest that the true data-generating process is far more complex than the current GWAS imply, which leads to problems in constructing a PGS. Interestingly, one of the main shortcomings of the PGS is its implementation and interpretation in GxE models. PGS are often interpreted as a genetic predisposition towards a certain trait. In GxE models, the PGS coefficient and the coefficient of its interaction with an environment are often of interest to social science researchers. By construction, PGS are a weighted average of survey SNPs, in which the population of the survey generally differs from the GWAS population.

As already mentioned, this may lead to problems of performance of the PGS in terms of predictive power, but it also hinders the interpretation of the GxE model coefficients. The construction of a PGS implicitly assumes that the GWAS coefficients are valid for the sample in which the GxE is conducted. Moreover, using PGS in GxE models raises some substantive difficulties in the interpretation of the results. The problems become even more complex when the outcome in the GWAS differs from the outcome studied in an analysis for which the PGS is constructed, which is the case in my analysis. In this section, I show that using GWAS coefficients as weights in PGS construction generally leads to a measurement error bias of the GxE model coefficients. Below, I discuss the problem of applying GWAS estimates to construct a PGS when the outcome of the GWAS differs from that of the GxE study[2].

A) GWAS step

I first consider the GWAS step. N, J, and K denote the number of observations, the number of SNPs, and the number of environments. Next, denote the SNP matrix as $G_{N\times J}$, the environment matrix as $E_{N\times K}$ and the interaction matrix as $(E \times G\Gamma)_{N\times K}$, where $G\Gamma$ represents the PGS with $\Gamma_{J\times 1}$ being the matrix of J SNP coefficients (genetic effects) γ_j. Then, general true and estimated GWAS models can be described as follows:

$$GWAS \quad Stage:$$

$$Y = G\Gamma^{survey}\beta + E\theta + E \times G\Gamma^{survey}\rho + \epsilon \tag{1.1}$$

[2]In section A of the appendix I show that the PGS is generally measured with error if the true model is the GxE model and the weights used to create it are based on GWAS estimates.

$$Y = G\Gamma^{gwas} + v \tag{1.2}$$

$$\mathbb{E}[W\epsilon] = \mathbf{0} \tag{1.3}$$

$$\mathbb{E}[Gv] = \mathbf{0} \tag{1.4}$$

$$W = [G \quad E \quad G \times E]$$

where equation (1.1) represents the standard GxE model, which has been estimated in many gene-environment applications and is here assumed to be the true data generating process, and equation (1.2) represents the GWAS model. Equation (1.2) represents a multivariate GWAS that generates SNP coefficients, which are then used to construct the PGS [3].

B) Difference between GWAS and Survey Outcomes

In the setting of this paper, differences between Γ^{gwas} and Γ^{survey} are likely to arise due to differences in outcomes between the GWAS step and the estimation step. In this application, I use the BMI GWAS (Locke et al., 2015) to create an index of genetic endowment, which I use in a GxE model that explores diseases. However, in general, the difference in outcomes may arise for several reasons. For instance, it can be the case that, in both stages, the two outcomes aim to provide information about the same or similar substantive topics (e.g. education). However, the methodology and the nature of the question may differ in the two samples, which may cause differences between their outcomes. An interesting aspect of genetic data is that, with current technology, it is possible to investigate the role of genetic predispositions for an outcome Y_{gwas} on outcome Y_2. This is conceptually close to the cross-productivity idea, and more insight into this type of analyses will shed more light on important human trait formation. I investigate the role of genetic predispositions for obesity on diseases and medical conditions including cancer and heart attack. In this case it is even more obvious that the weights from the GWAS are not applicable to the PGS. To make the argument clear, consider a simple

[3]Note that I abstract from the additional problem that GWAS coefficients from univariate regressions may be biased. Additional or different biases may arise if the true model does not conform to the assumption of a simple linear interaction. These issues are likely to make the problem even worse. Moreover, I acknowledge that the GWAS usually also includes other variables such as sex, age, and principal components of the genetic relationship matrix. In this example, I abstract from this as it does not affect the results in any significant way.

setup where the GWAS stage and the estimation stage look as follows:
GWAS (First) Stage:

$$Y_{gwas} = \gamma_0^{gwas} + \gamma_1^{gwas} SNP + \zeta \tag{1.5}$$

Estimation (Second) Stage

$$Y_2 = \beta_0 + \beta_1 PGS + \epsilon \tag{1.6}$$

For simplicity, I consider a case where only 1 SNP is related to an outcome Y_{gwas}. In the typical scenario, the GWAS stage produces SNP weight γ_1^{gwas} by estimating the model (1.5). In the second stage, a researcher uses survey data to construct the PGS and uses it in the estimation stage. Note that the PGS is an estimate of a conditional average of Y_{gwas}, $E[Y_{gwas}|SNP]$. It follows that equation (1.6) amounts to $Y_2 = \beta_0 + \beta_1 \mathbb{E}[Y_{gwas}|SNP] + \epsilon$. Also, there is an implicit relationship between Y_2 and SNP such that $Y_2 = \gamma_0^{survey} + \gamma_1^{survey} SNP + v$. Abstracting from additional problems that may lead to differences in the $gamma_1^{gwas}$ and γ_1^{survey}, the important implication of this procedure is that if a researcher wants to estimate the effects of genetic predispositions for trait Y_{gwas} on trait Y_2, then, if he or she uses the standard $PGS = \gamma_1^{gwas} SNP$, it will be measured with error. Consequently, the standard method will not identify the relationship between the true $PGS^* = \gamma_1^{survey} SNP$ and the outcome because of measurement error in the PGS. The model, as described by equation (1.6) then becomes:

$$Y_2 = \beta_0 + \beta_1 \frac{\gamma_1^{gwas}}{\gamma_1^{survey}} PGS^* + \epsilon \tag{1.7}$$

Instead of the coefficient of interest β_1, the model identifies $plim \quad \widehat{\beta_1} = \beta_1 \frac{\gamma_1^{gwas}}{\gamma_1^{survey}}$. If the GWAS estimates were portable across outcomes and samples then model 1.7 will identify β_1 but in general a case when $\gamma_1^{survey} \neq \gamma_1^{gwas}$, the estimate of β_1 will be biased.

To overcome the problems described in this section, I estimate the effects of the GxE on adult health using a new approach, which I present in section 1.4.

1.3 Data

To study the long-term relationship between initial genetic conditions, education, and health, I use the UK Biobank data. Similarly to Barcellos et al. (2018), I apply this data

and investigate the consequences of the ROSLA policy, which raised the school leaving age. Specifically, I study the heterogeneous effects of the policy on health outcomes by genotype. The UK Biobank is a prospective cohort study of more than 500,000 people who were aged between 37 and 73 at the time of recruitment between 2006 and 2010 (Sudlow et al., 2015). It provides a unique combination of large sample size and rich information about individual health and socio-economic outcomes. Survey participants were asked to complete a self-reported touchscreen questionnaire and to participate in a computer-assisted interview. Additionally, trained nurses collected their blood, saliva, and urine samples together with additional physical and medical measures. Finally, all participants were genotyped. I focus on health outcomes that were collected and standardized across the recruitment centers. All medical variables were collected by trained nurses.

To make the results of my analysis comparable to the benchmark model of Barcellos et al. (2018), I first focus on the same health outcomes (body size, lung function, blood pressure, and health summary index). I provide additional evidence on the heterogeneous effects of the policy on severe diseases and health conditions: stroke, heart attack, cancer, and type 2 diabetes. The final analytical sample consists of almost 260,000 white individuals with European ancestry born in England, Scotland, or Wales between September 1, 1947 and September 1, 1967. I focus on European ancestry only because GWAS summary statistics are available for this ancestry group. Similarly to the original study, I focus on birth cohorts born within a ten year window from September 1957. Children born in September 1957 were the first who directly experienced the policy change.

1.3.1 Outcomes

The first four outcomes analyzed in this paper are created using the same procedure as in Barcellos et al. (2018). The first outcome measure is body size index. This is generated using 3 measures from the data: body mass index (BMI), body-fat percentage, and waist-hip ratio. The UK Biobank provides two measures of BMI. First is the standard weight-to-height squared ratio. The second combines height with mass quantified using electrical impedance. I create the final BMI measure by taking the average of these two measures. The three measures are then combined into a single body-size index that

represents a weighted average of the three measures. Before taking the weighted average, I standardize the three measures for men and women by taking those born within the 1-year window before September 1957 as a reference. Then, similarly to Barcellos et al. (2018), I use Anderson's (2008) procedure, which uses input weights according to the variance-covariance matrix of the input variables. Higher values of the index denote poorer health status.

Second, I create a measure of lung function using spirometry-test data. A spirometer is a machine that measures the speed and volume of air after a forced exhale. The data provides information about three values of the spirometry test: (i) forced expiratory volume in one second represents the volume in litres of air exhaled during the first second; (ii) forced vital capacity represents the volume in litres of air exhaled during the forced breath; and (iii) peak expiratory flow represents the fastest rate of exhalation measured in litres per minute. Next, I standardize the three measures for men and women using the same procedure as in the previous paragraph. The final measure of lung function is a weighted average of the three spirometry measures constructed in the same way as the body size index. Finally, the signs of the respective input measures are reverted so that higher values of the index denote worse health status.

The third outcome is a measure of blood pressure. The data provides information on two measures of both systolic and diastolic blood pressure. To create one measure for each type of blood pressure, I first standardize the outcomes as previously. Then I take the simple average of each pair of measures. Finally, I use the same weighted average procedure as before and create a blood pressure index. Similarly to the previous two indexes, higher values of the blood pressure index denote poorer health status. The fourth outcome is a summary index of the three measures presented above, and represents a weighted average of the three main outcomes.

In addition to the outcomes investigated by Barcellos et al. (2018), I present new evidence of the effects of genetic endowment connected to BMI and of the heterogeneous effect of the ROSLA policy by the BMI genetic endowment. Specifically, I investigate the role of GxE on diseases and medical conditions that are linked to lifestyle and other environments. Following the first part of the analysis, I choose diseases related to obesity. The first outcome I study is the probability of having a heart attack. Second, I study the probability of having a stroke. The third outcome is the probability of receiving a

diagnosis of any kind of cancer and the fourth is a diagnosis of type 2 diabetes. All the disease outcomes are provided in the UK Biobank medical history of the patients. The UK Biobank provides two sources of information on occurrences of heart attack and stroke. The first source is self-reported occurrence of the event in the presence of an interviewer (a trained nurse). The second source is a touchscreen question about whether the respondent was ever diagnosed by a doctor as having had a heart attack or stroke.

I construct the final measure using both sources such that the final outcome is a binary variable with a value equal to one if either of the two sources provides evidence that a health event occurred and zero otherwise. For the cancer outcome, I use only the touchscreen question about whether a doctor ever diagnosed the respondent with any form of cancer.

Finally, I identify those who have type 2 diabetes based on self-reported information provided to an interviewer (a trained nurse). I study the origins of just type 2 diabetes because this type can be affected by lifestyle, unlike type 1, which often emerges early in life.

In Table 1.1 I show that the selected health problems correlate substantially with body size. Table 1.1 shows that a one standard deviation increase in the body size index is associated with an average increase of heart attack incidence of 0.8 percentage points. A correlation coefficient that corresponds to approximately 60 % of the total baseline heart attack incidence rate in the sample, which is 1.3 %. Furthermore, table 1.1 shows that one standard deviation increase in the body size index is associated with a 0.4 percentage point increase in stroke and cancer incidence in the data. These associations correspond to 40 % and 68 % of the total baseline stroke and cancer incidence rates in the sample. Finally, table 1.1 shows that one standard deviation increase in the body size index is associated with a 0.7 percentage points increase in the type 2 diabetes incidence rate, an association that is even larger than the baseline type 2 diabetes incidence rate.

Table 1.2 shows how higher values of health indicators are correlated with severe health conditions, which suggests that the relationship between indicators and health problems is non linear. Specifically, for most of the health indicators, the health problems are positively correlated with the third quartile and higher values of the standardized health indicator distribution. Table 1.3 presents the sample using some simple descriptive statistics and baseline health problem probabilities.

Table 1.1: Health Outcomes and Body Size

Outcome:	Body Size Coefficient
Heart Attack	0.008***
	(0.000)
Stroke	0.004***
	(0.000)
Cancer	0.004***
	(0.000)
Type 2 Diabetes	0.007***
	(0.000)
Number of observations: 259 380	
Heteroskedasticity robust standard errors in parenthesis	

Table 1.2: Health Outcomes and Health Indicators

Outcome:	Heart Attack	Stroke	Cancer	Type 2 Diabetes
Body Size ($1^{st}Qartile$)	-0.010***	-0.005***	-0.0058***	-0.006***
	(0.000)	(0.000)	(0.001)	(0.000)
Body Size ($3^{rd}Qartile$)	0.016***	0.008***	0.008***	0.013***
	(0.001)	(0.001)	(0.001)	(0.001)
Lung Function ($1^{st}Qartile$)	-0.008***	-0.005***	-0.012***	-0.004***
	(0.000)	(0.000)	(0.001)	(0.000)
Lung Function ($3^{rd}Qartile$)	0.011***	0.009***	0.014***	0.005***
	(0.001)	(0.001)	(0.001)	(0.000)
Blood Pressure ($1^{st}Qartile$)	0.011***	-0.000	-0.004***	-0.001*
	(0.001)	(0.000)	(0.001)	(0.000)
Blood Pressure ($3^{rd}Qartile$)	-0.007***	0.000	0.005***	-0.002***
	(0.000)	(0.000)	(0.001)	(0.000)
Number of observations: 259 380				
Heteroskedasticity robust standard errors in parenthesis				

Table 1.4: Descriptive Statistics: By Sex

	Men		Women	
Variable	Mean	Standard deviation	Mean	Standard deviation
Body Size	0.000	1.023	-0.005	0.965
Lung Function	0.022	1.035	0.069	1.055
Blood Pressure	0.001	0.994	-0.007	1.006
Summary Index	0.011	1.025	0.026	1.027
Heart Attack	0.024	0.154	0.004	0.064
Stroke	0.012	0.110	0.008	0.086
Cancer	0.041	0.198	0.074	0.261
Type 2 Diabetes	0.009	0.093	0.005	0.068
Age	52.786	5.897	52.848	5.797
Number of observations:	116210		143170	

Table 1.3: Descriptive Statistics

Variable	Mean	Standard deviation
Body Size	-0.003	0.992
Lung Function	0.048	1.047
Blood Pressure	-0.004	1.001
Summary Index	0.019	1.026
Heart Attack	0.013	0.114
Stroke	0.010	0.098
Cancer	0.059	0.236
Type 2 Diabetes	0.006	0.080
Male	0.448	0.497
Age	52.82	5.842

Number of observations: 259 380

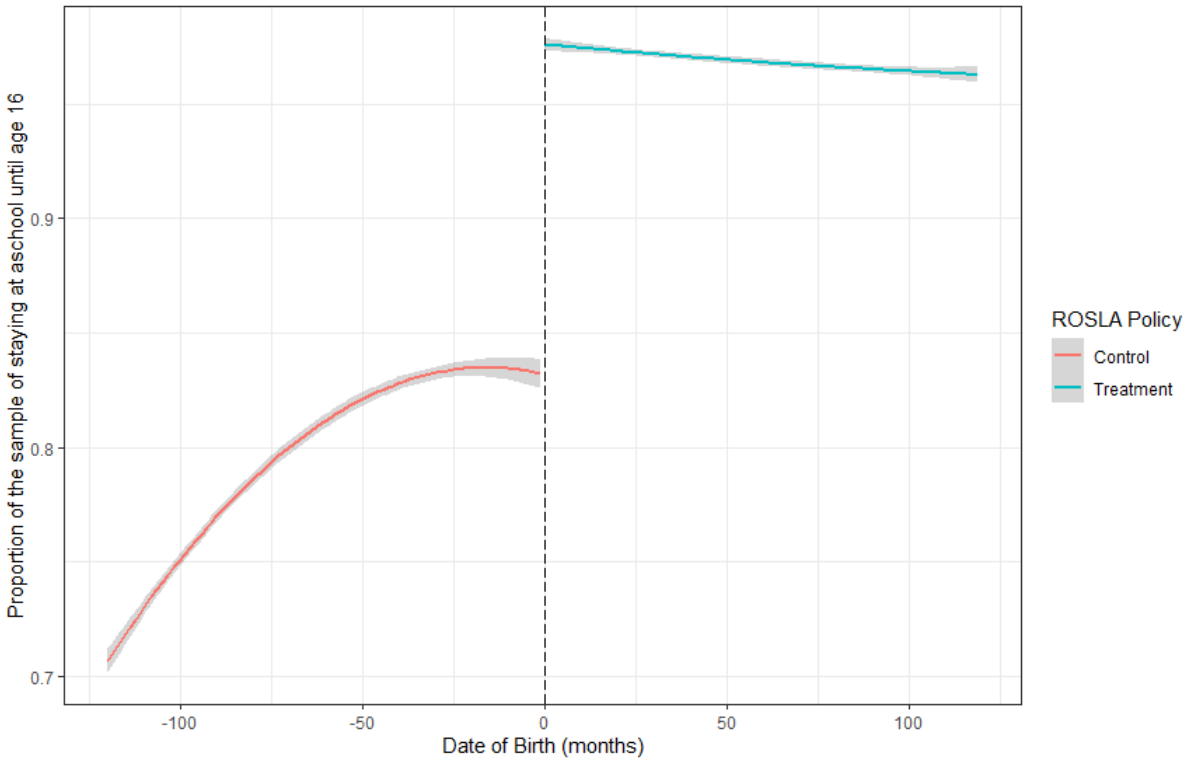
1.3.2 The ROSLA Policy

On September 1, 1972, the government of Great Britain raised the minimum school-leaving age from 15 to 16 years old in England, Scotland, and Wales. The policy implied a "cutoff" date of birth on September the first of 1957. Children born before this date were not affected by the ROSLA policy and could leave school at the age of 15, whereas those born after the cutoff date had to remain in school at least until they were 16. Figure 1.1 demonstrates the relationship between the year of birth and the propensity to stay in school until the age of 16. The figure also shows the discontinuous jump in this propensity in the birth year affected by the policy. The design of the policy presents an ideal environment to study the effects of an additional year of schooling on important economic and health outcomes. This paper builds on the results of Barcellos et al. (2018) and presents new evidence on the heterogeneous effects of the ROSLA policy on health by different genetic risk levels for poor health, using a new two step nonlinear method presented in section 1.4. Contrary to Barcellos et al. (2018), I do not have data on the day of birth but only on the birth month. Thus, in my analysis I construct trends based on the year and month of birth of the individuals.

1.3.3 Genetic Data

To investigate the heterogeneous effect of the ROSLA policy by genetic endowment, I use individual SNP data provided by the UK Biobank. In the main analysis, I compare results from the standard linear model with the newly proposed nonlinear GxE model.

Figure 1.1: Discontinuity around the date of birth on the propensity to stay in school until the age of 16.



As suggested above, I use two different measures of the genetic propensity to high BMI. First, I use the standard PGS, which I create based on Locke et al.'s (2015) summary statistics, and which I adjust for linkage disequilibrium (LD) [4] using LDpred. Finally, I standardize the final PGS for BMI using the procedure from Barcellos et al. (2018)

In the nonlinear model, I do not use the GWAS weights to construct the PGS, but rather estimate the individual SNP coefficients together with all other coefficients from the GxE model. Finally, I standardize the selected SNPs the same way I standardize the PGS[5].

[4]LD is a problem of standard GWAS because the analyses test each SNP at a time. Thus, if there is a non-zero covariance between different SNPs, the raw GWAS coefficient suffers from omitted variable bias due to the omitted SNPs

[5]I describe the selection of the relevant SNPs in 1.5 section

1.4 The Gene-Environment Model

Here I introduce the GxE model that I apply to the data to study how an additional year of education impacts the relationship between initial genetic conditions and health. To overcome the challenges presented in section 1.2, I propose a two-step nonlinear approach that aims to alleviate measurement error in the PGS and provide a more direct way to answer the important question on how policy interventions can help to level health differences that stem from different genetic predispositions.

My proposed two-step method is similar to current practice, except that it does not construct the PGS using GWAS weights, but rather builds an index function that also represents, an individual's propensity towards an outcome. The proposed index is as follows:

$$PGS_{new} = \mathbb{G}[Y|SNP] = \sum_{j}^{K} w_j^Y SNP_j \tag{1.8}$$

Y is the outcome of interest and w_j^Y are individual SNP weights estimated by the GxE model, together with the main parameters of interest. The main challenge is to estimate the individual weights. To make the problem feasible, I use GWAS summary statistics to select only K SNPs in the analysis, such that $K = \{SNP|p - value <= p - value^*\}$, where the p-value stands for the p-value from the GWAS stage summary statistics. This index does not depend on the GWAS weights and therefore represents a raw form of initial genetic conditions, which makes interpretation of the GxE model coefficients more straightforward and comparable to important economic concepts such as dynamic complementarity, self-productivity, and cross-productivity (Heckman, 2007).

The new approach I use consists of two steps. In the first step, I use GWAS summary statistics to select genome-wide significant SNPs for a given outcome. Standard GWAS summary statistics are based on univariate regressions that have a high propensity for false positive-discoveries. To address the problem of the SNP selection based univariate regressions, I adjust the p-values using the COJO method developed in Yang et al. (2011). COJO analysis aims to correct for the omitted variable bias of the GWAS estimates that arises due to the linkage disequilibrium (LD) structure of the genetic data[6] by considering the survey data's genetic correlation matrix.

[6]LD essentially means that the SNP variance-covariance matrix is not diagonal

To overcome the potential problem of overfitting, I use the part of the analytical sample that is not used in the main analysis to conduct the COJO analysis. One of the major challenges of the new approach is to select the right amount of SNPs that are related to the outcome of interest and make the estimation of the GxE model feasible. While the standard method typically, uses all available SNPs to construct the PGS, my new method is constrained by the amount of SNPs it can include for the GxE model to be identified. This constraint may potentially lead to problems because most traits are polygenic (Chabris et al., 2015), which means that many SNPs contribute to an outcome. Hence, the potential downside of my new method is that the proposed index function 1.8 may not include all the relevant SNPs, which may lead to skewed results. However, unless the total number of variables in the GxE model is equal to or larger than the number of observations, my new method identifies the coefficients of interest. Hence, with large enough samples and with outcomes that are determined with relatively small amount of SNPs, my new method delivers consistent results. However a proper selection device is needed in order to select the correct amount of relevant SNPs that is large enough to account for the polygenic nature of an outcome and at the same time estimation of the model is still feasible. To accomplish this, I first run the standard approach with the PGS created using GWAS summary statistics adjusted for LD using LDpred (Vilhjálmsson et al., 2015). Next, I calculate the mean squared error (MSE) of the standard GxE model and then choose the amount of SNPs for the new nonlinear approach such that the MSE of the standard approach approximately equals the MSE of the new nonlinear GxE model.

In the second step, after selecting the variables, I estimate the GxE model to provide causal evidence about how education moderates the effects of adverse genetic predispositions on health. In this setup, the causality requires both G and E to be exogenous. The standard approach in the genetic literature is to consider the genetic endowment as being random, conditional on parental genotype. In most applications, this is still not feasible, so to diminish the confounding of the G variables, researchers often include principal components (PCs) of the genetic relatedness matrix into their models. The PCs serve as controls for population stratification, which is a known confounding factor in the genetic models. Next, in most economic applications of the GxE model, the environment E is endogenous, which leads to inconsistent estimates of the causal parameters (θ_0, ρ_0). Re-

cently, researchers have aimed to identify the causal effects by employing identification strategies and focusing on the exogenous variation in E (Schmitz and Conley, 2017b, 2016a,c, Barcellos et al., 2018).

I analyse my GxE model specification and the performance of my new two step nonlinear method, exploiting the exogenous shift in years of schooling resulting from the ROSLA policy described above. To make the differences between the standard and my new methods clear, I first analyze the effect of the interaction of the ROSLA policy and the PGS on health outcomes using the standard method, which relies on GWAS weights to construct the PGS. Next, I estimate the same model using my new method and compare the results. For the purpose of this paper, I only consider the reduced form equation, in which I use the date of birth to discriminate between those who were and were not affected by the policy.

The major advantage of the two step nonlinear approach is that the interaction coefficient provides more direct evidence of the heterogeneous effect of the policy by genotype. Finally, the method overcomes the measurement error problem presented in section 1.2 that arises when the outcomes in GWAS and in the GxE analysis do not align. Finally, as equation (1.9) documents, my new method allows a researcher to study the cross-effects of genetic predispositions, because the individual SNP weights are estimated in the main specification. The β_1 coefficient captures the cross-effect of the index (1.8) on an outcome. Finally, the flexibility of my new method allows the individual SNP (genetic) effects to differ for different outcomes, which overcomes most of the problems mentioned in sections 1.2 and 1.4.

In the following, I apply a regression discontinuity design together with nonlinear least squares to estimate the coefficients of interest.

$$\begin{aligned} Y = c + \beta_1 PGS_{new} + \theta ROSLA + \rho ROSLA \times PGS_{new} + \\ t_1 DoB + t_2 DoB^2 + t_3 DoB \times ROSLA + t_4 DoB^2 \times ROSLA + \alpha X + \epsilon \end{aligned} \tag{1.9}$$

where PGS_{new} is the genetic risk score; its weights are estimated by the model. ROSLA represents the cutoff birth date of September 1, 1957, DoB stands for the date of birth trend, and X is a matrix of other covariates. The X matrix includes 15 principal

components of the genetic relationship matrix and their interaction with the ROSLA dummy variable, male dummy, dummies for place of birth, and a second order polynomial of age. I run model specification (1.9) for all outcomes presented in section 1.3. In the analysis, I use the same 10-year bandwidth of the running variable as in previous studies (Barcellos et al., 2018, 2019).

To estimate model (1.9), I first specify a sample loss function. For the purpose of this paper, I minimize a standard sample mean square error, i.e., $\widehat{Loss_S} = \sum_i^N [y_i - \hat{y}_i]$ by combining the gradient descent algorithm with the Adadelta learning rate [7]. One potential issue of the nonlinear least squares estimator is that, depending on the starting values of the parameters, it may converge to a local optimum, which is not necessarily the global optimum. Note that model (1.1) to a large extent resembles a standard linear model, which makes it attractive because the MSE loss function resembles the standard convex loss function of the linear models, which alleviates the problem of the local optima not being global. However, to ensure that I converge to a global optimum, in the analysis I try several initial values of the parameters in the analysis. For more complex functions, that may not have a convex loss function, the nonlinear nature of my new method may lead to a set of parameters that do not represent the global optimum. In that case, it is necessary to try various starting values and observe whether the coefficients converge to the same values.

1.5 Results

Here, I present and discuss the results of the GxE model that appear in section 1.4. I also compare the results from my new method to those of the benchmark GxE model that uses a standard PGS as a measure of genetic endowment. Following the procedure described in section 1.4, the estimation of the GxE model is accomplished in 2 steps. The first step of the nonlinear two step approach involves selecting the significant SNPs. I use Locke et al.'s (2015) GWAS summary statistic for BMI and run a COJO analysis

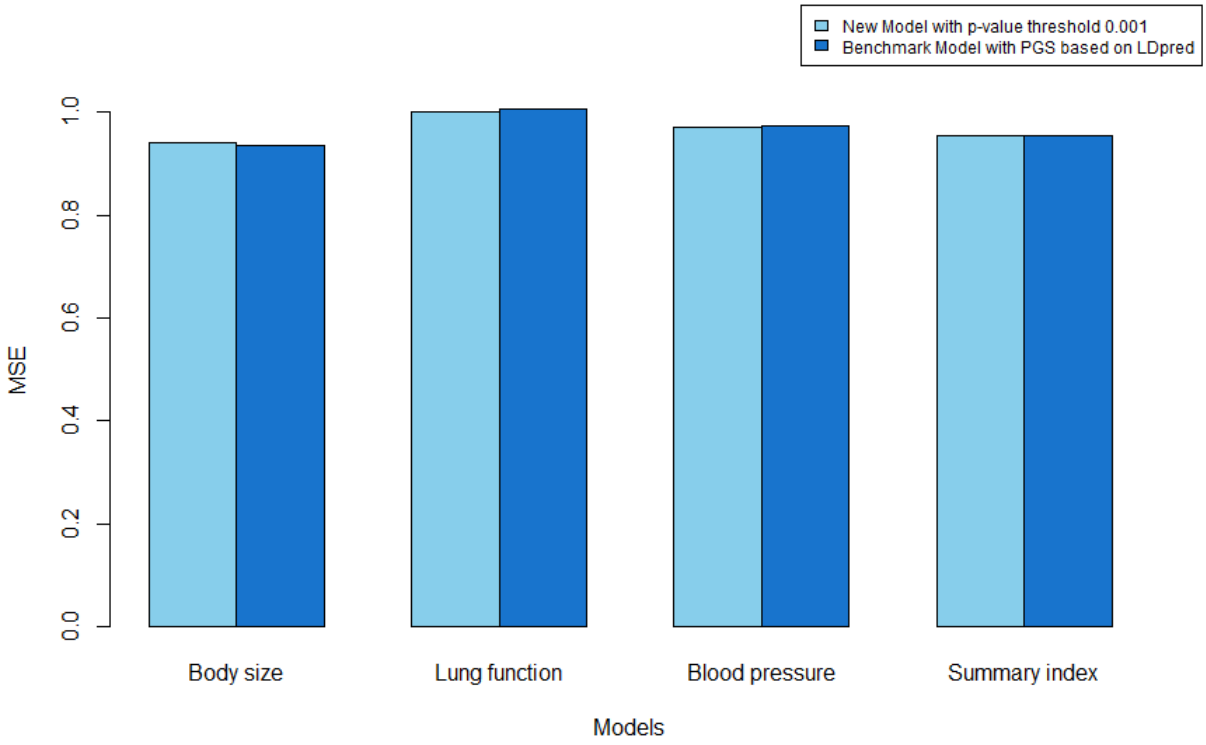
[7] The gradient descent algorithm is easy to implement, because it does not rely on a Hessian matrix to calculate the step size, which makes it attractive for large models. Instead of the Hessian matrix, I use the Adadelta method to calculate the step size of each iteration. Adadelta adapts the step size of each coefficient according to the average gradient and previous step sizes. Consequently, the algorithm is less likely to overshoot the optimum.

(Yang et al., 2012). The COJO analysis corrects for the omitted variable bias of the univariate GWAS regressions caused by omitted SNPs. A COJO analysis needs two inputs, the GWAS summary statistics and the survey data, in order to estimate the variance-covariance SNP matrix. To avoid overfitting, I use the UK Biobank sample, which I do not use in the main analysis, to estimate the GxE model. This sample consists of individuals of European ancestry who were born outside the time window used in the main analysis.

I select the SNPs that I use in the main empirical model based on the COJO p-value. The challenge is to select the optimal number of SNPs. Because the PGS usually includes all the SNPs, it is likely that it will have greater predictive power. To select the optimal p-value threshold for my analysis, I first run the standard GxE model using a BMI PGS based on LDpred SNP weights (Vilhjálmsson et al., 2015). Then I calculate the mean squared error (MSE) of this GxE model and select the number of SNPs such that the MSE of the nonlinear model is close to the GxE linear model. Figure 1.2 shows the comparison of the MSEs between the linear GxE model and the final nonlinear GxE model, which uses a SNP p-value threshold of 0.001. The total number of SNPs selected for the analysis is 337. After I select the optimal number of SNPs, I proceed with the second step, in which I estimate the GxE nonlinear model (1.9).

Table 1.5 presents the first set of results of the new nonlinear model (1.9) and compares them to the benchmark linear GxE RDD model, similar to the one estimated in Barcellos et al. (2018), with a polygenic score based on LDpred weights. The results in Table 1.5 show that the BMI PGS has a statistically significant positive impact on later-life body size, lung function index, and blood pressure. Before the ROSLA reform, the effects of the PGS was significant in all three cases at the 1% significance level. As expected, the BMI PGS affects body size the most, out of all the outcomes investigated. According to the new nonlinear model, the results suggest that before the ROSLA reform an increase by one standard deviation in the BMI PGS led, on average, to a 0.141 standard deviation increase in body size. The results also suggest a cross-effect or pleiotropy of the BMI genetic predispositions. The genetic predispositions towards obesity do affect lung function and blood pressure later in life. The analysis shows that before the reform, a one standard deviation increase in the BMI PGS led to worse lung function by an average of 0.078 standard deviations and increased blood pressure by 0.072 standard

Figure 1.2: Mean Squared Error Comparison Between the Benchmark Method and my new method with Selected SNPs: Indicators



deviations. Table 1.5 also suggests that the benchmark model underestimates the impact of these cross-effects.

Furthermore, the results suggest that the ROSLA reform intensified the role of genetic endowment on intermediary health indicators. Specifically, the results show that after the reform the effect of the genetic endowment increased by 0.014 standard deviations for body size and 0.008 standard deviations for blood pressure. This effectively means that for individuals who were affected by the reform a one standard deviation increase in the BMI PGS led to, on average, a 0.155 increase in body size. Similarly, one increase in the BMI PGS for this sub-population means an average 0.08 standard deviation increase in blood pressure [8].

[8]In this paper, blood pressure is an index that accounts for both systolic and diastolic blood pressure. For more information see section 1.3

Table 1.5: Results A: Analysis of Health Indicators

	Body Size		Lung Function		Blood Pressure		Summary Index	
	New Model	Benchmark Model	New Model	Benchmark Model	New Model	Benchmark Model	New Model	Benchmark Model
BMI PGS × ROSLA	0.014***	0.009**	0.003	-0.011**	0.008**	0.012***	0.010**	0.004
	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
BMI PGS	0.141***	0.160***	0.078***	0.027***	0.072***	0.034***	0.101***	0.102***
	(0.000)	(0.002)	(0.000)	(0.003)	(0.000)	(0.003)	(0.000)	(0.003)
ROSLA	-0.018	-0.018	-0.023	-0.022*	0.013	0.015	-0.022*	-0.020
	(0.012)	(0.012)	(0.014)	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)
N	255395	255395	212287	212287	259151	259151	209519	209519

Significance levels: ***0.01 **0.05 *0.1
Standard errors: (i) OLS heteroskedasticity robust (ii) NLS bootstrapped with 1000 resamples according to MacKinnon (2006).

Interestingly, the comparison of the results from the benchmark linear model, which includes a PGS based on LDpred weights, and the new two-step nonlinear model introduced in this paper, suggests significant differences in the two approaches. In Table 1.8 I provide a statistical test of the equality of the coefficients from the two models, based on bootstrapped T-statistics. As described in sections 1.2 and 1.4, the benchmark model is likely to suffer from PGS measurement error bias induced by using incorrect weights in the PGS construction. I apply my new approach, which does not use the standard weights in the PGS construction. Consequently, the discrepancies between the two approaches likely point to the measurement error bias of the benchmark method.

The largest discrepancies between the two methods appear in the blood pressure and lung function models. The BMI PGS coefficient of my new method in the case of the lung function model is almost three times higher than the corresponding coefficient of the benchmark method for the lung function model. Similarly, in my new method, the PGS BMI coefficient for blood pressure is twice as large as it is in the benchmark method coefficient. This evidence points to the problem presented in section 1.2 that when the outcomes in the GWAS and those from the main analysis differ, the PGS weights from the GWAS generally lead to skewed results. The analysis performed in this paper is better suited to estimate the crosseffect of genetic predispositions for BMI on outcomes that do not measure obesity or body size in general. At the same time, Table 1.5 shows only minor differences between the BMI PGS and BMI PGS × ROSLA coefficients for body size. Note that this is an outcome that is conceptually closer to the one from the GWAS that I use to construct the PGS index.

Overall, the evidence suggests that the benchmark model underestimates the effect

that the genetic endowment for BMI has on outcomes that are conceptually more distant from BMI. Finally, the results suggest that, if the outcome at hand is similar to that of the GWAS, then the PGS delivers similar results to those of the my new method, which does not use GWAS estimates to construct the index of genetic predisposition.

The results from the new two GxE method in Table 1.5 suggest that the genetic predisposition for BMI leads to poorer results of intermediary health indicators. The results also suggest that the ROSLA policy, which increased the compulsory years of schooling, also increased the role of genetic predispositions in the formation of the health indicators. However, it is necessary to acknowledge that health indicators are not the actual diseases or severe medical conditions which lead to worse quality of life and sometimes even death. Thus, poorer outcomes of the health indicators may not necessarily signal a poorer quality of life due to health problems.

To investigate the ways in which the ROSLA policy and genetic predisposition towards obesity contribute to adverse health conditions, I first analyze how these two inputs influence the higher quantiles of the health indicators. Table 1.6 presents the results of a version of model (1.9). The outcomes are binary variables, equal to one if the value of the given health indicator is higher, or equal to the third quartile of the respective health indicator distribution. This analysis is similar to Barcellos et al. (2019) and Barcellos et al. (2018), and is another necessary step in the analysis of the links between health indicators and actual health problems. In the analysis, I focus on the third quartile because it provides a better insight into how the gene-education interplay contributes to the higher, and therefore worse, values of health indicators. Table 1.6 provides the first indicative evidence of the role of education as a mediating factor for unfavorable genetic endowment in regards to health. The evidence in Table 1.6 is in accordance with previous literature (Barcellos et al., 2018) and suggests that an increase in years of schooling decreased the inequality in the higher values of the health indicators that stems from genetic predispositions. However, this result is still only indicative, as it does not provide concrete evidence of the mediating role of education on actual diseases and dangerous health conditions. The top quartiles of distributions of health indicators are indicative of a potential effect on worse health. However, one can argue that the exact cutoff when the worse values of health indicators including BMI, blood pressure, or lung function severely affect health is not clear. For that reason, below I analyse and discuss

the mediating role of education on actual health conditions and diseases.

To continue with the discussion above and to shed more light on the mediating role of education on health conditions, I connect the policy change and genetic predispositions to illnesses and life threatening conditions. While the health indicators are indicative of life expectancy and overall quality of life, it is important to study the consequences of lifestyle and other environmental factors and genetic predispositions on the probability of developing a severe life-threatening disease. This analysis sheds more light on the sources of heterogeneity in the quality of life at later stages of life. Hence, I investigate how the results on intermediary health indicators translate into the probability of developing cancer, severe cardiac diseases, stroke, and type 2 diabetes, which are all outcomes that severely disrupt the quality of life. Table 1.7 explores the relationship between the ROSLA policy, BMI PGS, and obesity-related diseases.

Table 1.6: Results B: Outcome $\geq 3^{rd}$ quartile

	Body Size		Lung Function		Blood Pressure		Summary Index	
	New Model	Benchmark Model	New Model	Benchmark Model	New Model	Benchmark Model	New Model	Benchmark Model
BMI PGS × ROSLA	-0.004**	-0.008***	-0.012***	-0.006***	0.008*	0.002	-0.011**	-0.010***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)
BMI PGS	0.054***	0.059***	0.032***	0.013***	0.072***	0.011***	0.041***	0.040***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
ROSLA	-0.007	-0.007*	-0.009	-0.009	0.013***	0.002	-0.010*	-0.010*
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
N	255395	255395	212287	212287	259151	259151	209519	209519

Significance levels: ***0.01 **0.05 *0.1
Standard errors: (i) OLS heteroskedasticity robust (ii) NLS bootstrapped with 1000 resamples according to MacKinnon (2006).

Table 1.7: Results C: Probability of Medical Conditions

	Heart Attack		Stroke		Cancer		Type 2 Diabetes	
	New Model	Benchmark Model	New Model	Benchmark Model	New Model	Benchmark Model	New Model	Benchmark Model
BMI PGS × ROSLA	-0.002***	-0.001***	-0.000	-0.001*	-0.004**	0.001	-0.001	-0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.002)	(0.001)	(0.000)	(0.000)
BMI PGS	0.005***	0.002***	0.002***	0.001***	0.010***	-0.000	0.002***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
ROSLA	-0.001	-0.001	-0.001	-0.001	-0.002	-0.002	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.001)	(0.001)
N	259380	259380	259380	259380	258644	258644	182144	182144

Significance levels: ***0.01 **0.05 *0.1
Standard errors: (i) OLS heteroskedasticity robust (ii) NLS bootstrapped with 1000 resamples according to MacKinnon (2006).

The results from Table 1.7 indicate that genetic predispositions for BMI do positively affect the probability of developing cancer or of having a heart attack. Specifically, a one standard deviation increase in the BMI PGS increases the risk of heart attack by, on average, 0.5 percentage points, and the risk of cancer by 1 percentage point. Interestingly, Table 1.7 demonstrates that the ROSLA policy decreased the effect of genetic predisposition on the probability of developing cancer by 40% (from 1 percentage point to 0.06 percentage points) and the probability of having a heart attack also by 40% (from 0.5 percentage points to 0.3 percentage points). Hence, the policy helped to level the health disparities tilted from birth, by 40% on average for incidences of heart attack and cancer.

Furthermore, note that Tables 1.7 and 1.8 present more evidence on the difference between the novel and benchmark methods, which is consistent with previous results. Specifically, when applied to an outcome that does not align with the GWAS step outcome, the PGS coefficient from the benchmark method is attenuated compared to the coefficient from my new method, which does not incorporate the GWAS estimates as weights but rather estimates them together in the method. Moreover, figure 1.3 shows the correlation between individual SNP coefficients from LDpred and my new method. In the figure, the chosen SNPs are subset of all LDpred SNPs that are included in the nonlinear model. Figure 1.3 suggests that the individual SNP coefficients from the two methods are not correlated. This presents even further evidence in support of the hypothesis that the coefficients from GWAS are generally not portable across outcomes.

In terms of magnitude the coefficients are similar to the effects of previous literature, which finds no or a very small effect of education on health outcomes. For instance, Clark and Royer (2013) finds that one additional year of education in early adolescence decreases the probability of fair or poor health by 0.08%. Similarly, Barcellos et al. (2019) finds no signifficant effect of one additional year of education on health indicators, although they do find substantive effect on higher quantiles of the health indicator distributions. Moroever, Davies et al. (2018) finds that a year of additional education dicreases the risk of cancer, heart attack, and stroke by approximately 1%.

It is interesting to link the results from Tables 1.5, 1.6, and 1.7, as they suggest a broader picture of health development. While the results of the first set suggest that education may intensify the impact of genetic endowment on the average values of the health

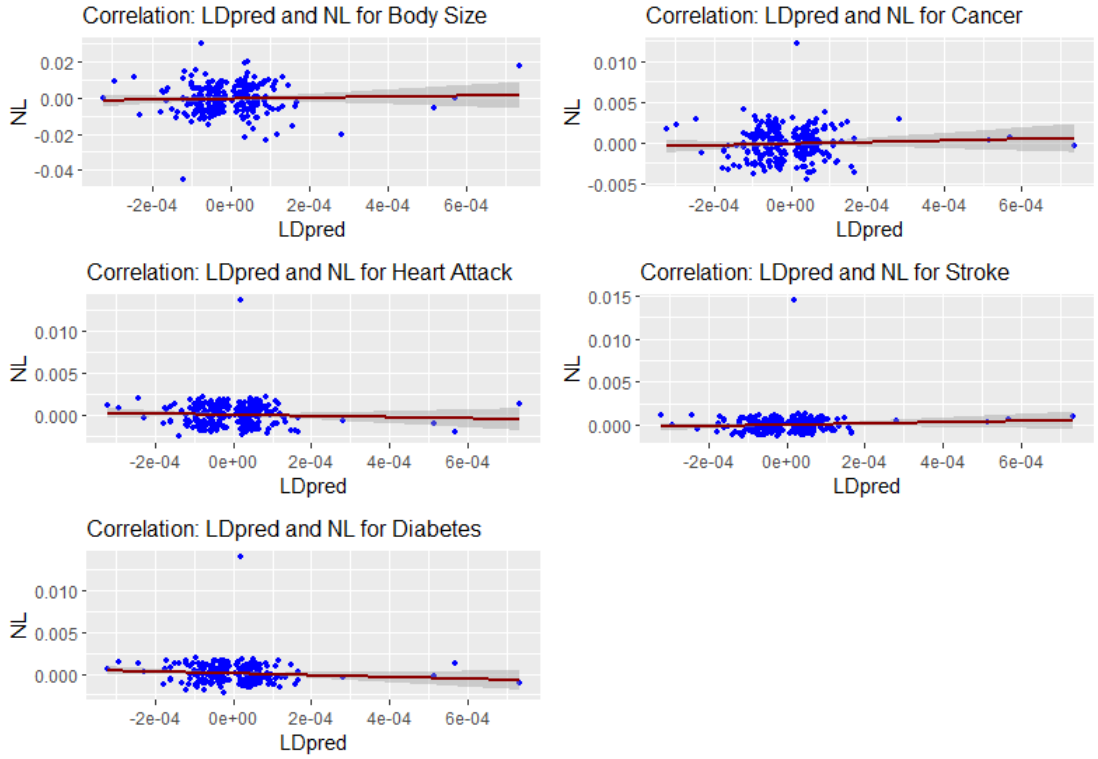
Table 1.8: P-values of the Tests of Equality of Coefficients

	Pval: PGS	GxE
Heart Attack	0.000	0.000
Stroke	0.000	0.006
Cancer	0.000	0.000
Type 2 Diabetes	0.000	0.014
Body Size	0.000	0.998
Lung Function	0.000	0.801
Blood Pressure	0.000	0.884
Summary Index	0.000	0.967
High Body Size	0.000	0.021
High Lung Function	0.000	0.000
High Blood Pressure	0.000	0.004
High Summary Index	0.000	0.000
P-values are based on bootstrapped T test using 1000 bootstrap samples		

indicators, the story reverses for more extreme values of these indicators. More importantly, the analysis in this paper shows that the ROSLA intervention, which increased the years of compulsory schooling, helped to diminish the health inequality that stems from unequal initial genetic predispositions by decreasing the role of genetic endowment on the probability of developing cancer or experiencing a heart attack. These results add to the debate that genetic endowment is not primarily deterministic and, even for outcomes such as cancer, it is possible to adjust the environment such that the probability of actually developing a severe medical condition is lower than genetic predispositions suggest. The results also show that a policy starts to moderate the relationship between genetic predispositions and health indicators only when the values of the indicator are rather high. Taken together with the moderation effect of the policy on the diseases, the evidence suggests that individuals start to adopt preventive actions when the values of the indicators are high enough that their health may clearly be in danger. Consequently, the results suggest that the cost of an unhealthy lifestyle captured by the effect on health indicators is nonlinear.

Previous research suggests that individuals derive utility from an unhealthy lifestyle, but this also carries a cost of potential health problems in the future (Biroli, 2015b). This cost, as shown in this paper, is heterogeneous by genetic endowment. If the cost

Figure 1.3: Correlation between SNP Coefficients from my new method and LDpred



of an unhealthy lifestyle is indeed nonlinear, one would expect individuals to engage in preventive behavior when there is a higher risk of developing a severe disease. This claim is supported by the evidence, as it is a direct implication of the moderation effect being more important for higher values of the health indicators and for diseases.

1.6 Robustness Checks

This section presents robustness checks and investigates the validity of the regression discontinuity design. A potential problem of the design is the strategic behavior of individuals close to the cutoff. They may behave in such a way to fall in the desired part of the threshold value of the assignment variable. In this case, the policy had been in preparation since 1964 and was introduced in 1972. Thus, it is unlikely that there

are substantial differences within the 1957 birth cohort as a consequence of the strategic behavior of the parents. However, it is possible that, because the investigated outcomes are related to death, the results might be distorted by survival bias. In this specific scenario, individuals with higher genetic predispositions for obesity and with lower years of education may be more prone to die early and hence not respond to the survey. To investigate the possibility of survival bias, Table 1.9 shows the results of the b_2 coefficient of the balance test regressions $Y = b_0 + b_1 Cutoff + b_2 date - of - birth + b_3 date - of - birth^2 + b_4 date - of - birth \times cutoff + b_5 date - of - birth^2 \times cutoff + u$. Table 1.9 suggests that, in the treatment group, there are approximately one percent more males than in the control group. Moreover, some differences can be seen in the principal components that stand for population stratification.

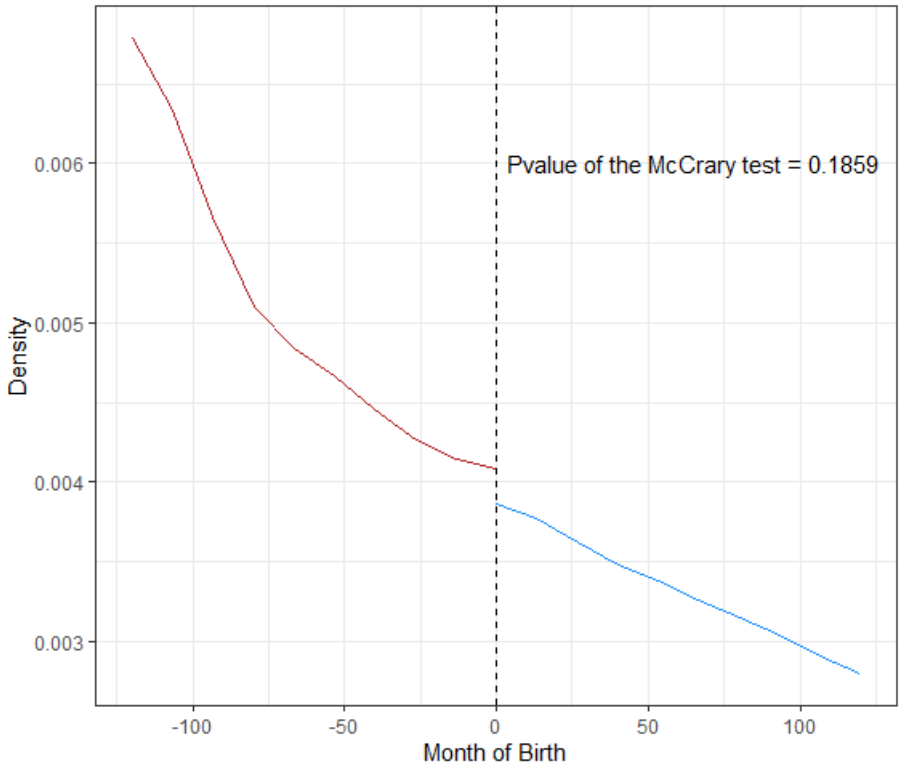
Table 1.9: Balance Test

Outcome	Cutoff Coefficient	Robust Standard Errors
BMI PGS	0.004	0.012
Fraction of Males	0.012**	0.006
Fraction of Wales	0.001	0.003
Fraction of Scotland	-0.001	0.003
PC1	0.019	0.013
PC2	-0.009	0.012
PC3	0.017	0.012
PC4	-0.018	0.012
PC5	-0.030***	0.012
PC6	0.029***	0.012
PC7	-0.024*	0.012
PC8	0.029**	0.013
PC9	-0.011	0.012
PC10	-0.020	0.012
PC11	-0.007	0.012
PC12	0.012	0.012
PC13	0.004	0.012
PC14	-0.009	0.012
PC15	0.006	0.012
Significance levels: ***0.01, **0.5 ,*0.1		

The main analysis includes all variables from the balance check to ensure that they

do not affect the results. Importantly, if individuals with less education are more likely to develop a mortal disease, it is possible that the control group would be missing those who died early, which would be reflected in the McCrary test of the distribution of the running variable (date of birth) in the form of discontinuity around the cutoff. Figure 1.4 shows that there is not enough evidence to support this hypothesis.

Figure 1.4: McCrary Test



The results are based on the method developed by Cattaneo et al. (2020)

Next, I test for the difference in the PGS distributions between treated and the controls. Differences in genetic predispositions may also hint at survival bias, as those with a higher PGS for obesity may be more likely to die early and will not be included in the sample. Figure 1.5 presents the results of the Kolmogorov-Smirnov test of equality of the treatment and control distributions of PGS based on LD pred weights. The formal test rejects the null hypothesis about the equality of the distributions. However, as figure 1.5 suggests, the actual difference is substantively negligible. To further document the

substantively small difference in the two distributions, in Table 1.10, I show a formal test of the equality of means of the two PGS distributions. The test suggests a substantially negligible but statistically significant difference in means. However, due to the small substantive difference, the bias that may potentially arise from this is in the order of 10^{-8}, which does not affect the results in any meaningful way.

Figure 1.5

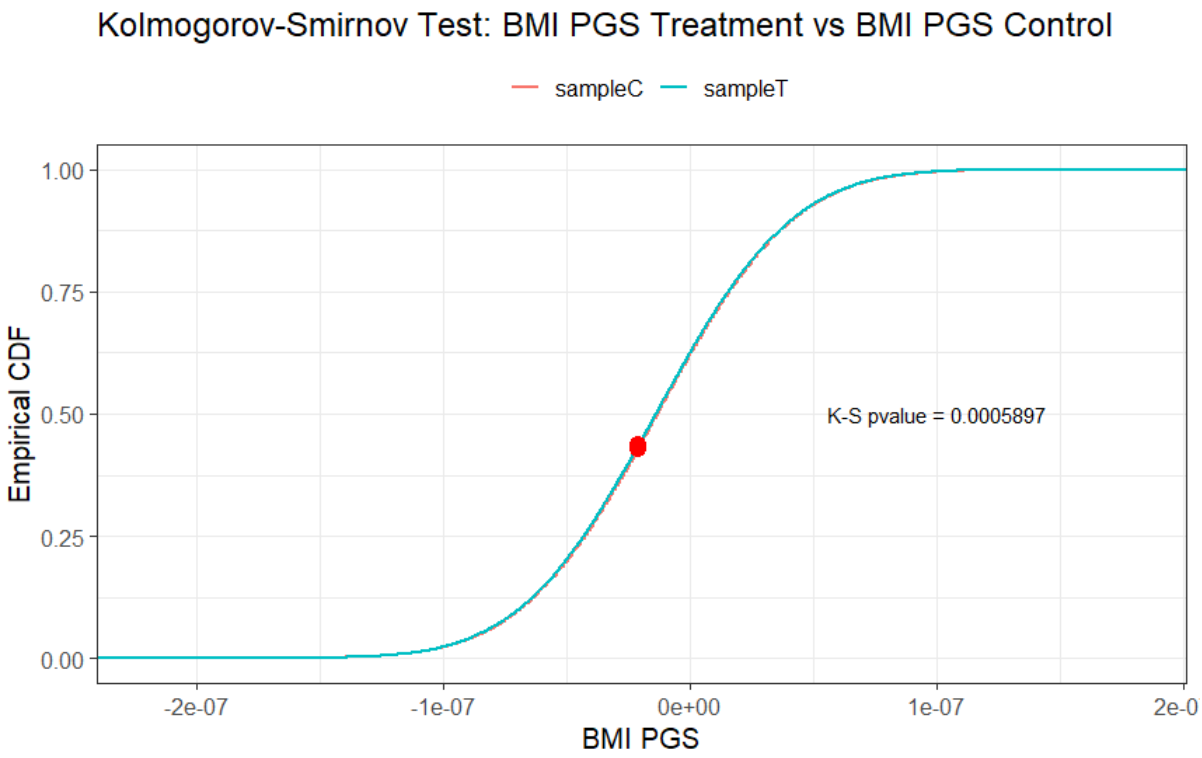


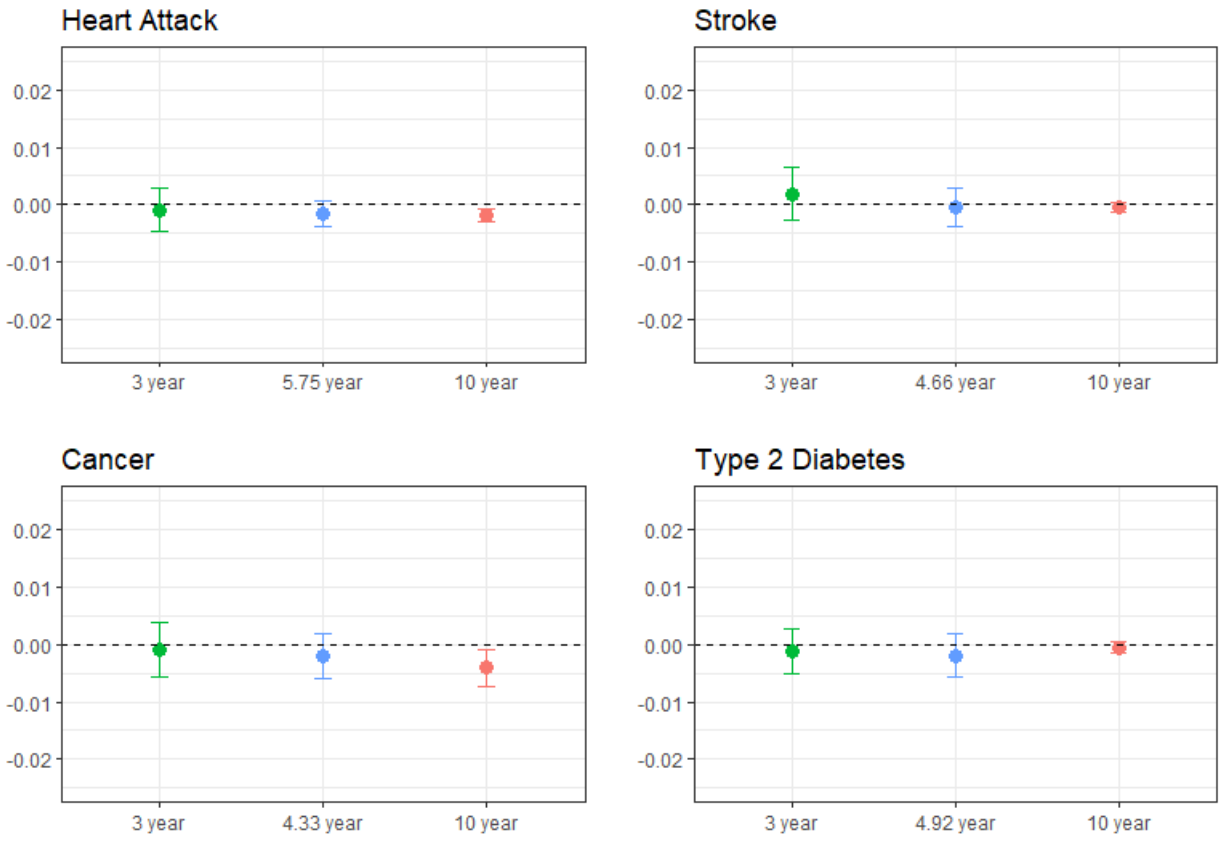
Table 1.10: Difference between the Control and Treatment Means of BMI PGS

Parameter	Value	T statistic
Mean Treatment BMI PGS	-1.400×10^{-8}	-
Mean Control BMI PGS	-1.336×10^{-8}	-
Difference	$-6.362 \times 10^{-10***}$	-3.655

Next, I check how different definitions of bandwidth alter the results of the interaction

term. Figure 1.6 shows the results of this test for the disease outcomes studied in Table 1.6. Similar graphs for the other outcomes can be found in the appendix section C. As figure 1.6 suggests, the results are robust to smaller bandwidths. The main difference between the 10 year, 3 year, and the data-driven Calonico et al. (2015) bandwidths stems from the lower confidence intervals of the 10-year bandwidth's RDD design, which is a natural outcome since this specification includes more observations.

Figure 1.6: 3 year, 10 year, and Calonico et al.'s (2015) Bandwidth of the RDD for GxE Estimates of the Model for Medical Conditions



The bars represent 95% confidence intervals.

The next set of robustness checks investigates whether the effect of GxE persists when the cutoff changes. Table 1.11 shows the results. Finally, Table 1.12 presents the results of the new two step method for individuals who did not attend college. As mentioned,

the ROSLA policy raised the school leaving age from 15 to 16. This essentially means that the affected group, or the compliant group, is composed of individuals who would have left school had the policy not been implemented. Therefore, the group most affected by the policy should be the population that did not attend college.

Table 1.11: Robustness Check: Different Cutoffs

PGS × Cutoff	Heart Attack
September 1955	-0.002***
	(0.001)
September 1959	-0.002*
	(0.001)
September 1961	-0.001
	(0.001)
September 1965	-0.001
	(0.001)
PGS × Cutoff	Cancer
September 1955	-0.001
	(0.001)
September 1959	-0.005***
	(0.001)
September 1961	-0.004***
	(0.001)
September 1965	-0.005***
	(0.002)

Significance levels: ***0.01 **0.05 *0.1

Standard errors: Bootstrapped with 1000 re-samples according to MacKinnon (2006).

Table 1.12: Robustness Check: Probability of Negative Health Outcomes on the Non-College Sub-population

	Heart Attack	Stroke	Cancer	Type 2 Diabetes
BMI PGS × ROSLA	-0.004***	-0.001	-0.005***	-0.000
	(0.001)	(0.001)	(0.002)	(0.001)
BMI PGS	0.007***	0.003***	0.014***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
ROSLA	-0.002	-0.002	0.003	-0.001
	(0.002)	(0.001)	(0.003)	(0.002)
N	167593	167593	167593	167593

Significance levels: ***0.01 **0.05 *0.1

Standard errors: Bootstrapped with 1000 resamples according to MacKinnon (2006).

1.7 Conclusion

This paper investigates the heterogeneous effect of the ROSLA policy by genotype to show how it decreased health inequality that stems from initial genetic endowment. I exploit the exogenous introduction of the ROSLA policy, which increased the school leaving age by one year and effected an increase in the years of education of individuals who would have left school earlier if allowed. The results suggest that the ROSLA intervention decreased the role of genetic endowment in the risk of developing cancer or experiencing a heart attack by 40%. Moreover, I show that initial genetic predispositions for obesity are positively related to heart attack, strokes, cancer, and diabetes. Hence, the paper shows that health inequality stemming from innate genetic endowment is malleable by intervention in the environment that increases years of schooling. This knowledge sheds light on how education policies and investments into education in general can help to level a playing field that has been tilted since birth by innate genetic factors. The results support an important claim in the literature on nature via nurture by showing that, even though genetic factors affect severe diseases such as cancer or heart attack, these effects may be moderated by behavior.

I also confirm findings from previous literature that the ROSLA policy moderates the effect of genotype for higher values of health indicators, and I show that a similar effect

applies to severe diseases. These two findings together suggest that the health indicators represent a measure of poorer health. However, the results suggest that individuals start to perceive the health indicators as a negative influence on their lives only after a certain threshold. This in turn suggests that the costs associated with high values of the intermediary health indicators is not linear. Additionally, given that individuals do derive utility from an unhealthy lifestyle, and that its costs are not linear, individuals may choose to work to counteract the high value of a health indicator only when there is a high risk of developing a disease.

To measure the genetic endowment, I use results from the obesity GWAS conducted by Locke et al. (2015). The standard approach is to construct a measure of genetic endowment called the PGS, which is a weighted average of individual SNP data where the weights are usually based on results from a GWAS conducted on different populations.

I demonstrate the problems with the current methods of constructing a PGS when the outcome of the GWAS does not correspond to the outcome of a GxE study. To overcome the problem, I present a novel method for studying GxE models that does not rely on individual SNP estimates from a GWAS. The method consists of two stages.In the first stage, I use the BMI GWAS results to select the relevant SNPs. In the second stage I re-estimate the individual SNP coefficients in the GxE model.

More generally, the paper expands the scope of current GWAS to situations in which the GWAS and survey outcomes of the GxE study differ. Unlike the standard approach, the novel method presented in this paper does not make any implicit assumptions about the portability of individual SNP weights across samples and outcomes. Instead, it uses a GWAS as a selection device and then estimates the weights in the GxE model. The results support the claim that my new method delivers estimates that have clear interpretations and alleviate the measurement error bias of PGS constructed based on GWAS weights.

Apendix

A Extensions of section 1.2

Consider the GWAS stage presented in section 1.2 by equations 1.1 - 3.4

$$
\begin{aligned}
GWAS \quad Stage: \\
&Y = G\Gamma^{survey}\beta + E\theta + E \times G\Gamma^{survey}\rho + \epsilon \\
&Y = G\Gamma^{gwas} + v \\
&\mathbb{E}[W\epsilon] = \mathbf{0} \\
&\mathbb{E}[Gv] = \mathbf{0} \\
&W = [G \quad E \quad G \times E]
\end{aligned}
$$

Consequently, the probability limit of the estimated GWAS coefficients approach is as follows

$$
\begin{aligned}
plim \quad \widehat{\Gamma^{gwas}} &= plim \left(\sum_i G_i^T G_i \right)^{-1} \left(\sum_i G_i^T Y_i \right) \\
&= plim \left(\sum_i G_i^T G_i \right)^{-1} \left(\sum_i G_i^T (G_i\Gamma^{gwas} + v) \right) \\
&= plim \left(\sum_i G_i^T G_i \right)^{-1} \left(\sum_i G_i^T (G_i\Gamma\beta + E_i\theta + (E_i \times G_i\Gamma)\rho + \epsilon_i) \right) \\
&= \Gamma\beta \\
&+ plim \left(\sum_i G_i^T G_i \right)^{-1} \left(\sum_i G_i^T E_i\theta \right) \\
&+ plim \left(\sum_i G_i^T G_i \right)^{-1} \left(\sum_i G_i^T (E_i \times G_i\Gamma)\rho \right) \\
&+ plim \left(\sum_i G_i^T G_i \right)^{-1} \left(\sum_i G_i^T \epsilon_i \right)
\end{aligned}
$$

Under assumption $\mathbb{E}[W\epsilon] = \mathbf{0}$

$$plim \quad \widehat{\Gamma^{gwas}} = \Gamma\beta + plim\left(\sum_i G_i^T G_i\right)^{-1} \sum_i G_i^T E_i\theta + plim\left(\sum_i G_i^T G_i\right)^{-1} \sum_i G_i^T \left(E_i \times G_i\Gamma\right) \rho$$

$$plim \quad \widehat{\Gamma^{gwas}} = \Gamma\beta + plim\left(\sum_i G_i^T G_i\right)^{-1} \sum_i G_i^T E_i\theta + plim\left(\sum_i G_i^T G_i\right)^{-1} \sum_i G_i^T \left(E_i \times G_i\Gamma\right) \rho \tag{1.10}$$

Equation (1.10) shows that not including the environment variables from the GxE studies may lead to omitted variable bias of the GWAS coefficients. The bias depends on two terms. First, if the environment is correlated with genetic endowment G, then the bias in $\widehat{\Gamma^{gwas}}$ depends on the relationship between E with G, E with Y, E with G, and $E \times G$ with Y. Thus, the bias in the estimated SNP coefficients $\widehat{\Gamma^{gwas}}$ is a complex function of environments, including parental investments, individual life experiences, and initial genetic conditions. Under the typical omitted variable scenario, it would be enough to either include the omitted variable in the regression or to exploit the variation in G that is exogenous. GWAS studies aim to partially solve the issue by including principal components of the genetic relatedness matrix that controls for population stratification (Price et al., 2006), which is one of the factors that leads to gene-environment correlation. More recently, GWAS research focuses on family data samples to better control for confounding factors such as population stratification, assortative mating, or omitted parental genotype (Kong et al., 2018, Young et al., 2019).

However, it is important to note that the bias in equation (1.10) also depends on the interaction term. Analogous to the argument presented in Solon et al. (2015), Deaton (1997), if the true genetic effect is heterogeneous, then GWAS identifies some sample weighted average of the heterogeneous genetic effects that is generally not the true average genetic effect.

Consider a case where the environment is uncorrelated with the genetic endowment.

$$Assumption \quad A.1: \quad \mathbb{E}[EG] = 0$$

Moreover, assume that the true data generating process (DGP) can be described by

equation (1.1). Then the population's average genetic effect is equal to:

$$\mathbb{E}[\delta_i] = \Gamma\beta + \Gamma\mathbb{E}[E]\rho \tag{1.11}$$

Then equation (1.10) is written as:

$$plim \quad \widehat{\Gamma^{gwas}} = \Gamma\beta + plim\left(\sum_i G_i^T G_i\right)^{-1} \sum_i [G_i^T G_i \Gamma E_i]\rho \tag{1.12}$$

Equation (1.12) implies that, under the GxE model (1.1), the SNP coefficient estimates are a weighted average of the heterogeneous genetic effects $\Gamma_i = \Gamma\beta + \Gamma E_i\rho$. Importantly, the average genetic effects identified by a GWAS depend on the distribution of the environment in the respective sample and on the conditional SNP variance-covariance matrix $G^T G$ [9].

Equation (1.12) suggests that, under GxE heterogeneity described by model (1.1), the estimated average genetic effects in a GWAS sample may not identify the population's average genetic effects. To better illustrate the problem and to analyze the sources of the potential bias, consider a simpler case where E is just one discrete variable. For the purpose of this paper, consider a policy that increases the years of schooling as an E measure. In this paper, I analyze how such an increase in the years of schooling modifies the health inequality that arises due to genetic endowment. However, if genetic effects on health outcomes depend on years of schooling during childhood and adolescence, then the estimated average genetic effects identified by equation (1.2) depend on the distribution of years of education in the GWAS sample. To formalise the argument, suppose that E is discrete and can take on a limited amount of possible values E_l such that $l = \{1, 2, ..., L\}$, representing whether a person was treated by the policy and consequently was exposed to more years of education. Then $\Gamma_i^{gwas} = \Gamma_l^{gwas} = \Gamma\beta + \Gamma E_l\rho$ [10]. Next, denote the GWAS sample size by N^{gwas} and the sample size of each l group by N_l^{gwas}. Finally,

[9] In section A of the appendix, consider a case of independence of E and G, which shows that, even under this strong assumption, the estimated coefficients may still lead to biased average genetic effects

[10] Note that in this case ρ is just a scalar. Furthermore, for simplicity, I consider a case where the heterogeneous effect is linear in E. The argument presented here even applies to a more realistic example of a fully saturated model where $\Gamma_l^{gwas} = \Gamma\beta + \sum_{l=1}^{L} \Gamma E_l\rho_l$

suppose that as the sample size N^{gwas} grows, the proportions of each l groups remains the same. Then, following the result of Deaton (1997) equation (1.12) can be rewritten as a weighted sum of genetic effects in a GWAS sample

$$plim \quad \widehat{\Gamma^{gwas}} = \Gamma\beta + plim \left(\sum_{l=1}^{L} \frac{N_l^{gwas}}{N^{gwas}} \Omega_l^{gwas} \right)^{-1} \left(\sum_{l=1}^{L} \frac{N_l^{gwas}}{N^{gwas}} \Omega_l^{gwas} \Gamma E_l \rho \right) \tag{1.13}$$

Where I denote the variance-covariance matrix of SNPs as $\frac{1}{N_l^{gwas}} G_l^T G_l$ as Ω_l^{gwas} [11].

Formula (3.8) illustrates several important points. First, the GWAS estimates of the genetic effects are not consistent estimates of the population's average genetic effect (1.11) unless $\rho = 0$, which is equivalent to saying that the genetic effects are homogeneous and equal to $\mathbb{E}[\delta_l] = \Gamma\beta$. Next, if $\rho > 0$, then the environment and the genetic endowment complement and reinforce each other, which will lead to an overestimation of the average genetic treatment effect. If $\rho < 0$, then the environment and genetic endowment mitigate each other, which will in turn lead to underestimation of the average genetic effect. This suggests that the GWAS estimates and PGS weights bias depend on the a priori complementarity between genetic endowment and the environment.

On a more technical note, the bias in the estimated average genetic effects also depends on the different genetic variance-covariance structure Ω_l^{gwas} among different groups in the environment. Interestingly, the bias does not disappear even if the genetic correlation structure is similar among different values of the environment. To see this, consider a case where $\Omega_l^{gwas} = \Omega^{gwas} \; \forall \, \text{l}$. Then $plim \quad \widehat{\Gamma^{gwas}} = \Gamma\beta + \frac{N_2^{gwas}}{N^{gwas}} \Gamma\rho$. Therefore, even if the variance-covariance structure of the SNPs is the same across environments, GWAS may still yield inconsistent estimates of the average genetic effects because the proportions of the environmental groups may differ from the population proportions. However, if G and E are independent terms, then proper weighting of the inverse population shares would lead to a consistent estimates of the population's average genetic effect. Even though in this case, weighted least squares would yield consistent estimates, it is unlikely that the independence of E and G holds. As economic research shows, individuals make strategic choices and select themselves into different environments and make important decisions

[11] In section A of the appendix I consider a special case where E takes on only 2 values.

about their human capital investments and about investments into their children. This selection most likely appears in the very early stages of life or even in the prenatal period, as parents also make choices about their offspring in the early stages of their development (e.g. Rosales-Rueda, 2014, Sanz-de Galdeano and Terskaya, 2019, Attanasio et al., 2018, Boneva and Rauh, 2018). Consequently, even the current advances in GWAS (Young et al., 2019, Kong et al., 2018) that aim to deal with the omitted variable bias from equation (1.10) will most likely not identify the population average genetic affects. In the best case scenario, they will alleviate the bias caused by the omitted environment, but they will not solve the issue of the omitted genetic effect heterogeneity.

Finally, equation (3.8) implies that (unless $\rho = 0$) GWAS estimates of the genetic effects will in general identify a different weighted average of the individual genetic effects than would a GWAS performed on survey samples used by researchers to estimate GxE models. Consequently, the GxE applications test for $\rho \neq 0$ but at the same time use PGS built using SNP coefficients that would be correct for a given survey sample only if $\rho = 0$. I illustrate the problem in more detail in section B of the appendix.

Extension of the omitted interaction effect formula (1.12). Assume a special case where E and G are independent.

$$Assumption \quad A.2: \quad E \perp G = 0$$

The independence assumption allows me to simplify equation (1.12) further to[12]:

$$\begin{aligned} plim \quad \widehat{\Gamma^{gwas}} &= \Gamma\beta + plim\left(\sum_i G_i^T G_i\right)^{-1} \sum_i G_i^T G_i \Gamma \sum_i E_i \rho \\ &= \Gamma(\beta + plim \sum_i E_i \rho) \end{aligned} \tag{1.14}$$

Hence, in this case, each of the estimated J SNP coefficients $\hat{\gamma}_j^{gwas}$ from a GWAS

[12]The independence is necessary because the bias term in equation (1.12) is a matrix of higher order moments. Therefore, the standard mean independence assumption is not enough to simplify the equation

converges in probability to

$$plim \quad \gamma_j^{gwas} = \gamma_j \left(\beta + plim \sum_i E_i \rho \right)$$

Equation (3.13) shows that the independence assumption alleviates the problem presented in section 1.2 because the estimated average genetic effects do not depend on the conditional variance-covariance genetic matrix $G^T G$, Nevertheless, the estimated average genetic effects still depend on the distribution of the environment in the sample. Therefore, even under the independence assumption, the GWAS estimates do not generally identify the population's average genetic effect and there is no reason to believe that the weighted average of the genetic effects identified by a GWAS is the correct weighted average of these effects that a researcher should use in GxE analyses performed in survey samples that differ from the GWAS samples.

Formula (3.8) represents a case where the heterogeneous effect is linear in E. To illustrate that the problem also applies to the saturated models, consider a case when E is categorical and can take only 2 values (i.e. 0 or 1) for $l = \{1, 2\}$ The formula above then rewrites as:

$$\begin{aligned} plim \quad \widehat{\Gamma^{gwas}} &= \left(\frac{N_1^{gwas}}{N^{gwas}} \Omega_1^{gwas} + \frac{N_2^{gwas}}{N^{gwas}} \Omega_2^{gwas} \right)^{-1} \left(\frac{N_1^{gwas}}{N^{gwas}} \Omega_1^{gwas} \Gamma\beta + \frac{N_2^{gwas}}{N^{gwas}} \Omega_2^{gwas} (\Gamma\beta + \rho\Gamma) \right) \\ &= \Gamma\beta + \left(\frac{N_1^{gwas}}{N^{gwas}} \Omega_1^{gwas} + \frac{N_2^{gwas}}{N^{gwas}} \Omega_2^{gwas} \right)^{-1} \frac{N_2^{gwas}}{N^{gwas}} \Omega_2^{gwas} \Gamma\rho \end{aligned} \tag{1.15}$$

B Measurement Error Bias in the GxE Model

The problems presented above imply that the PGS weights a researcher should use for a GxE model in a survey, $\widehat{\Gamma^{survey}}$, likely often differ from those that are actually used, $\widehat{\Gamma^{gwas}}$. In consequence, the researcher estimates a miss-specified model of the following

form:

$$Y_i = \alpha + [\text{PGS}_i^* + G(\Gamma^{gwas} - \Gamma^{survey})]\,\beta + E_i \times [\text{PGS}_i^* + G(\Gamma^{gwas} - \Gamma^{survey})]\,\rho + E_i\theta + \varepsilon_i \tag{1.16}$$

where $\text{PGS}^* = G\Gamma^{survey}$ is the true PGS among the population of interest. This formalization shows that the PGS as currently constructed can be seen as a version of the correct PGS that is affected by systematic measurement error. In line with this observation, some recent studies have pointed to the low predictive power of PGS, which is a common consequence of measurement error. Its predictive power varies with the specification of the outcome model and the population to which it is applied, which is likely to occur if the measurement error arises from differences in the model specification, or if the population of interest differs from the GWAS population (Mostafavi et al., 2020, Tropf et al., 2017). Importantly, if the PGS is mis-measured as described by equations (3.14) and (1.16) the estimated coefficients of the GxE model estimated in survey data will generally suffer from measurement error bias that depends on the relationship between genetic endowment and environments. Therefore, it is not classical measurement bias that would generally lead to attenuation. Instead, the direction of the measurement error bias will depend on the complementarity and covariance structure of the genetic endowment and environment. To analyze the nature of the bias, consider a case of only one environment interaction. First, introduce some notation. Denote the PGS constructed from the GWAS weights as follows:

$$\widetilde{PGS_i} = PGS_i^* + G_i(\Gamma^{gwas} - \Gamma^{survey})$$

Then the measurement error in the interaction term is:

$$E \times \widetilde{PGS_i} = E_i \times PGS_i^* + E_i \times G_i(\Gamma^{gwas} - \Gamma^{survey})$$

Denote the PGS measurement error as $G_i(\Gamma^{gwas} - \Gamma^{survey}) = G_i\Gamma^{\Delta}$ and the matrix of covariates as $\widetilde{X_i} = \left[\widetilde{PGS_i} \quad E \quad E \times \widetilde{PGS_i}\right]$.

Then the asymptotic measurement error bias of the GxE model (1.16) $\mathbb{B} = \mathbb{E}[\left[\widehat{\beta} \quad \widehat{\rho} \quad \widehat{\theta}\right]]^T - \left[\beta \quad \rho \quad \theta\right]^T$ depends on the variance-covariance structure of the genetic matrix, the en-

vironment and the difference between the true SNP coefficients and the GWAS SNP coefficients. Given that I consider only one environment, the probability limit of the model estimates is written as:

$$
\begin{aligned}
plim \begin{bmatrix} \widehat{\beta} \\ \widehat{\rho} \\ \widehat{\theta} \end{bmatrix} &= \begin{bmatrix} \beta \\ \rho \\ \theta \end{bmatrix} + plim \left(\sum_i \widetilde{X_i}^T \widetilde{X_i} \right)^{-1} \sum_i \widetilde{X_i}^T (-G_i \Gamma^{\Delta}) \begin{bmatrix} \beta \\ \rho \\ \theta \end{bmatrix} \\
&= \begin{bmatrix} \beta \\ \rho \\ \theta \end{bmatrix} - plim \left\{ \left(\sum_i \widetilde{X_i}^T \widetilde{X_i} \right)^{-1} \begin{bmatrix} \sum_i \widetilde{PGS_i} G_i \Gamma^{\Delta} & 0 & \sum_i \widetilde{PGS_i} E_i G_i \Gamma^{\Delta} \\ \sum_i E_i G_i \Gamma^{\Delta} & 0 & \sum_i E_i^2 G_i \Gamma^{\Delta} \\ \sum_i \widetilde{PGS_i} E_i G_i \Gamma^{\Delta} & 0 & \sum_i \widetilde{PGS_i} E_i^2 G_i \Gamma^{\Delta} \end{bmatrix} \begin{bmatrix} \beta \\ \rho \\ \theta \end{bmatrix} \right\}
\end{aligned}
$$

Therefore, the $\mathbb{B}$ bias amounts to:

$$
\mathbb{B} = -plim \left\{ (\textstyle\sum_i \widetilde{X_i}^T \widetilde{X_i})^{-1} \begin{bmatrix} \sum_{i=1}^{N} \sum_{k=1}^{J} \sum_{j=1}^{J} G_{i,k} G_{i,j} \gamma_k^{gwas} \gamma_j^{\Delta} & 0 & \sum_{i=1}^{N} \sum_{k=1}^{J} \sum_{j=1}^{J} G_{i,k} G_{i,j} \gamma_k^{gwas} \gamma_j^{\Delta} E_i \\ \sum_{i=1}^{N} E_i G_i \Gamma^{\Delta} & 0 & \sum_{i=1}^{N} E_i^2 G_i \Gamma^{\Delta} \\ \sum_{i=1}^{N} \sum_{k=1}^{J} \sum_{j=1}^{J} G_{i,k} G_{i,j} \gamma_k^{gwas} \gamma_j^{\Delta} E_i & 0 & \sum_{i=1}^{N} \sum_{k=1}^{J} \sum_{j=1}^{J} G_{i,k} G_{i,j} \gamma_k^{gwas} \gamma_j^{\Delta} E_i^2 \end{bmatrix} \begin{bmatrix} \beta \\ \rho \\ \theta \end{bmatrix} \right\} \tag{1.17}
$$

Equation (1.17) has several important implications for the estimated coefficients of the GxE model. Note that if $\Gamma^{gwas} = \Gamma^{\Delta}$ then the bias $\mathbb{B}$ goes to 0 because all the terms inside the middle matrix of $\mathbb{B}$ will be 0. Importantly, using results from equations (3.8), (3.14), and (1.17) it is easy to see that this condition holds if

$$
\begin{aligned}
&plim \left(\sum_{l=1}^{L} \frac{N_l^{gwas}}{N^{gwas}} \Omega_l^{gwas} \right)^{-1} \left(\sum_{l=1}^{L} \frac{N_l^{gwas}}{N_l^{gwas}} \Omega_l^{gwas} \Gamma E_l \rho \right) = \\
&plim \left(\sum_{l=1}^{L} \frac{N_l^{survey}}{N^{survey}} \Omega_l^{survey} \right)^{-1} \left(\sum_{l=1}^{L} \frac{N_l^{survey}}{N_l^{survey}} \Omega_l^{survey} \Gamma E_l \rho \right)
\end{aligned} \tag{1.18}
$$

Therefore, the measurement error bias will disappear if there is no heterogeneity in the genetic effects ($\rho = 0$) or if the structure of the genetic relatedness matrix $\sum_{l=1}^{L} \frac{N_l^{gwas}}{N^{gwas}} \Omega_l^{gwas}$

resembles the structure of the survey genetic relatedness matrix $\sum_{l=1}^{L} \frac{N_l^{survey}}{N_l^{survey}} \Omega_l^{survey}$. An important implication of the above thought experiment is that, even if the environment is exogenous in both samples, i.e., the GWAS sample and the survey sample, the measurement error will not disappear. To see this, note that if the environment is orthogonal to the genetic structure of the two populations (or samples) then $\Omega_l^{gwas} = \Omega_l^{survey} = \Omega$ for all l, which is not enough to satisfy the equality in equation (1.18), which in turn does not guarantee the measurement error bias $\mathbb{B}$ to be 0. In the previous section, I discussed that it is unlikely that, in the GWAS step, the environment will probably not be orthogonal to SNPs in the genetic matrix G. However, in the survey, researchers often employ identification strategies from econometrics that are built to identify causal effects. Note that if assumption A.2 is satisfied (i.e. $E \perp G_j, \forall j$), then the middle matrix of equation (1.17) simplifies and the measurement bias in $\widehat{\rho}$ and $\widehat{\beta}$ asymptotically approaches the following:

$$\mathbb{B}(\widehat{\beta}) = \beta \frac{1}{\sigma_G^2} \left(\sum_{j=1}^{J} \gamma_j^{gwas} \gamma_j^{\Delta} \sigma_{g,j}^2 + (J-1) \sum_{k=1}^{J} \sum_{j=1}^{J} \gamma_k^{gwas} \gamma_j^{\Delta} \sigma_{g,k,j} \right) \tag{1.19}$$

$$\mathbb{B}(\widehat{\rho}) = \rho \frac{1}{\sigma_E^2} \left(\sum_{j=1}^{J} \gamma_j^{gwas} \gamma_j^{\Delta} \sigma_{g,j}^2 + (J-1) \sum_{k=1}^{J} \sum_{j=1}^{J} \gamma_k^{gwas} \gamma_j^{\Delta} \sigma_{g,k,j} \right) \tag{1.20}$$

Where $\sigma_{g,k,j}$ denotes the covariance between SNPs k and j and $\sigma_{g,j}^2$ denotes the variance of SNP j. Note that, in order to obtain the results in (1.19) an(1.20) it is not enough to assume no correlation between E and G. The result in (1.19) an(1.20) is most likely to hold in an experimental setting where treatment and control groups are chosen completely randomly. Although equations (1.19) an(1.20) imply that, even under independence, the estimates of ρ and β in a GxE study would yield inconsistent estimates, in this special case it is still possible to test for $\rho = 0$ even if the PGS is measured with error as described above. Even though ρ and β are generally biased, their fraction will identify the true fraction up to a scale that is equal to the E ang G respective variances.

$$\frac{\widehat{\rho}}{\widehat{\beta}} = \frac{\rho}{\beta} \frac{\sigma_G^2}{\sigma_E^2}$$

Therefore, if G and E are independent at least in the survey sample, a researcher may

conduct a statistical test for $\frac{\rho}{\beta} = 0$ which would essentially test for $\rho = 0$[13]. It is important to acknowledge that this is a very special case that relies on a strong assumption that is unlikely to hold outside an experimental setting. Hence, a researcher should present strong evidence that assumption 2 is likely to hold in his or her setting using proper tests such as the test of conditional independence introduced in Mittag (2018) . In the general case, the measurement will lead to biased estimates of $\rho, \beta, and \theta$ in the GxE studies.

[13]Note that, in this test, it is important to assume that $\beta \neq 0$ which is a condition that is likely to hold in most GxE applications.

C Additional Figures

Figure 1.7: Mean Squared Error Comparison Between the Benchmark Method and my new method with Selected SNPs: Medical Conditions

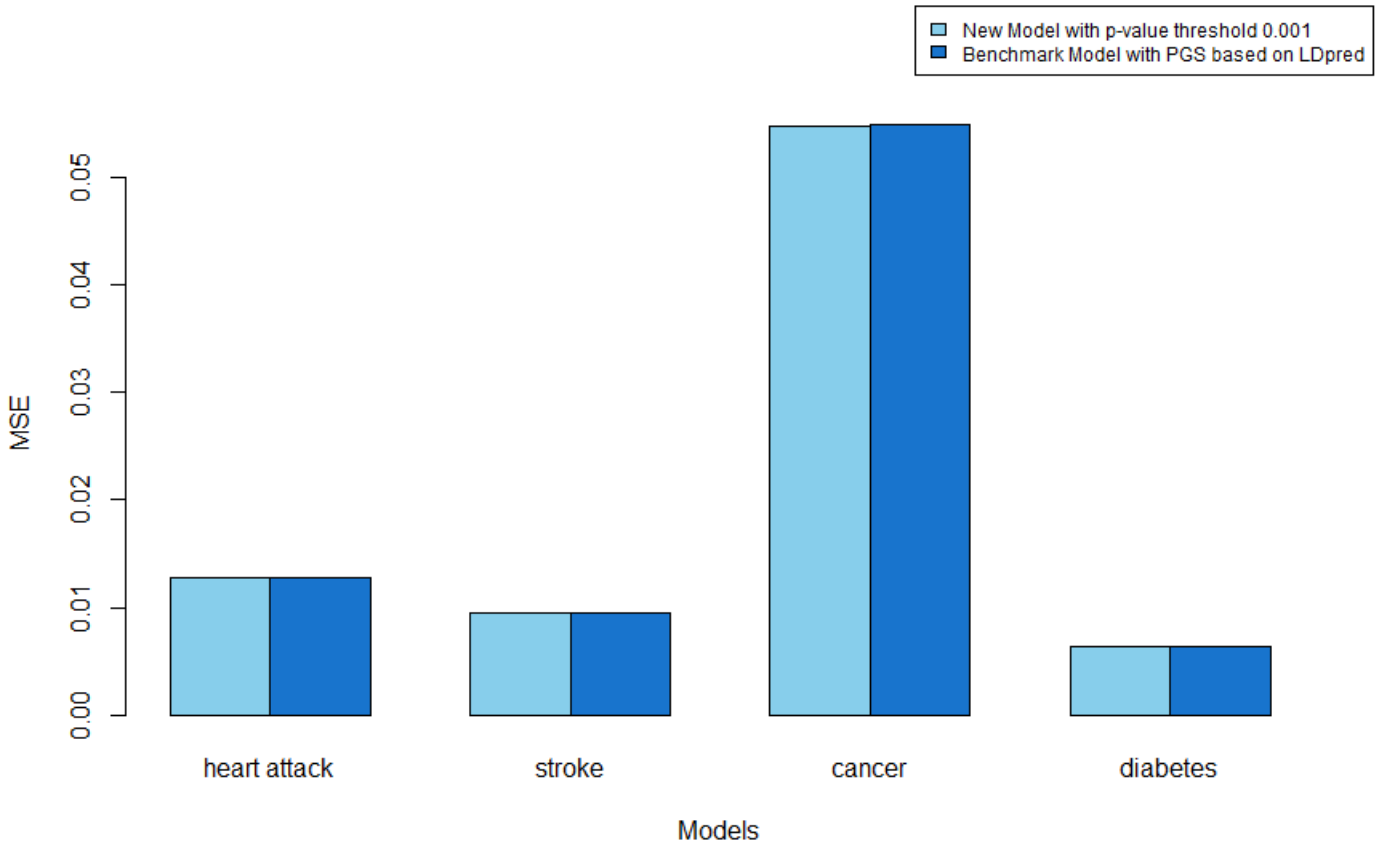


Figure 1.8: Mean Squared Error Comparison Between the Benchmark Method and my new method with Selected SNPs: Binary Indicators

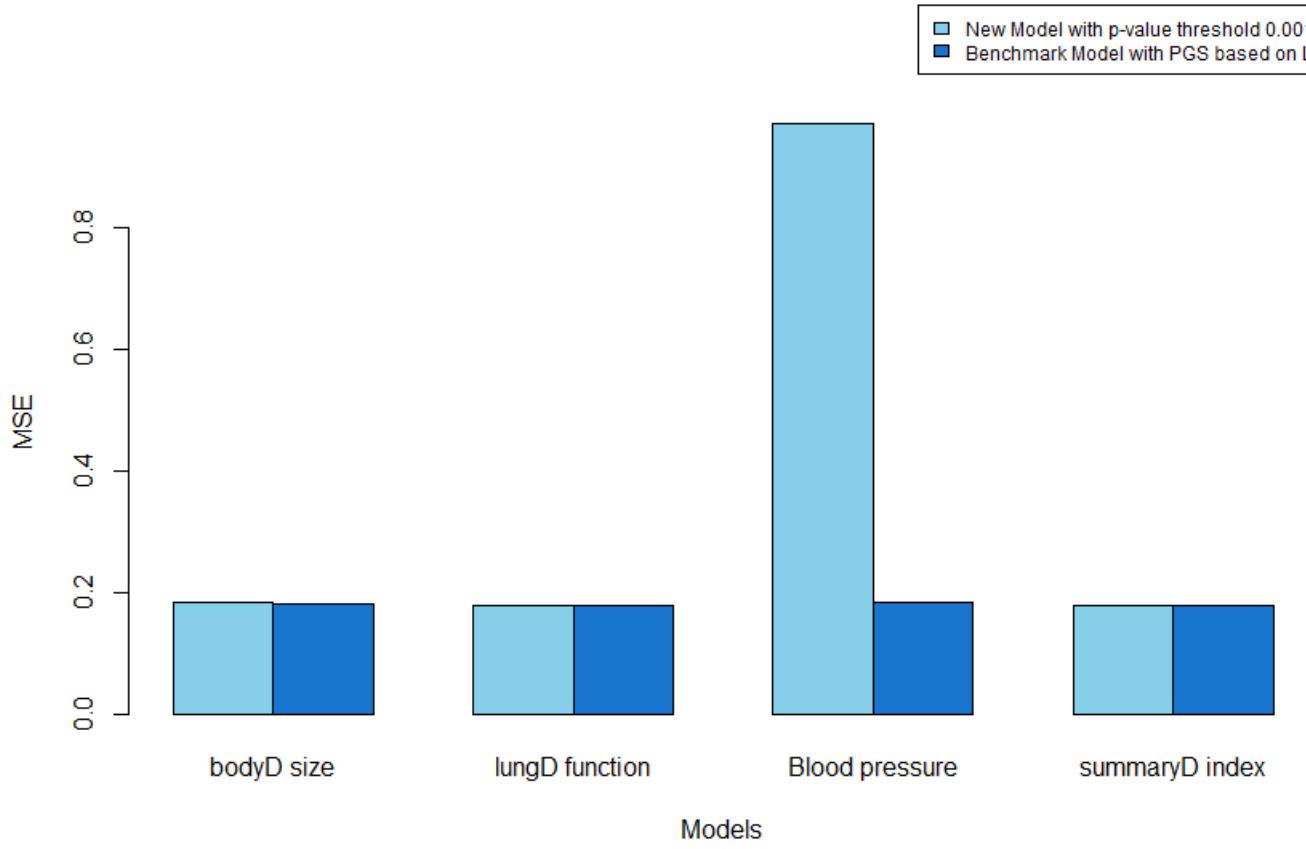


Figure 1.9: 3 Year vs 10 Year Bandwidth of the RDD for GxE Estimates of the Model for Continuous Indicators

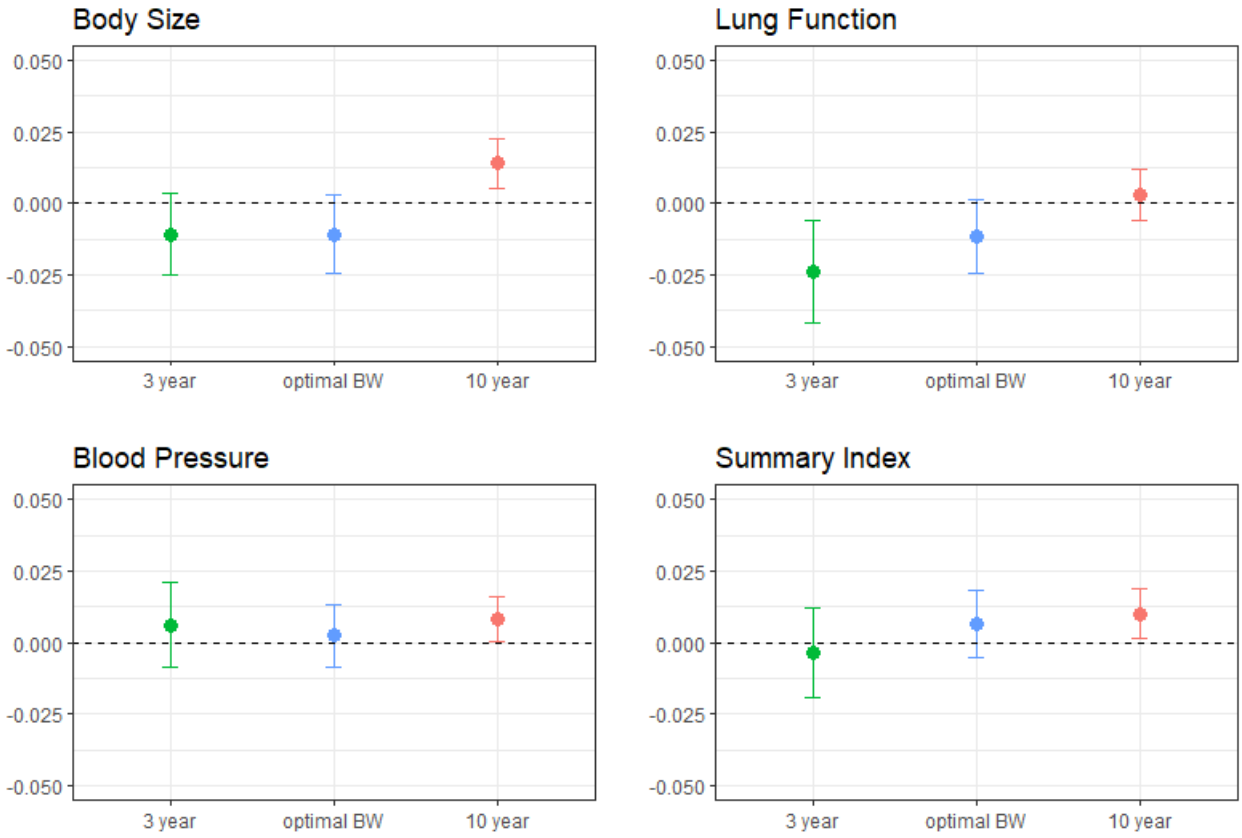
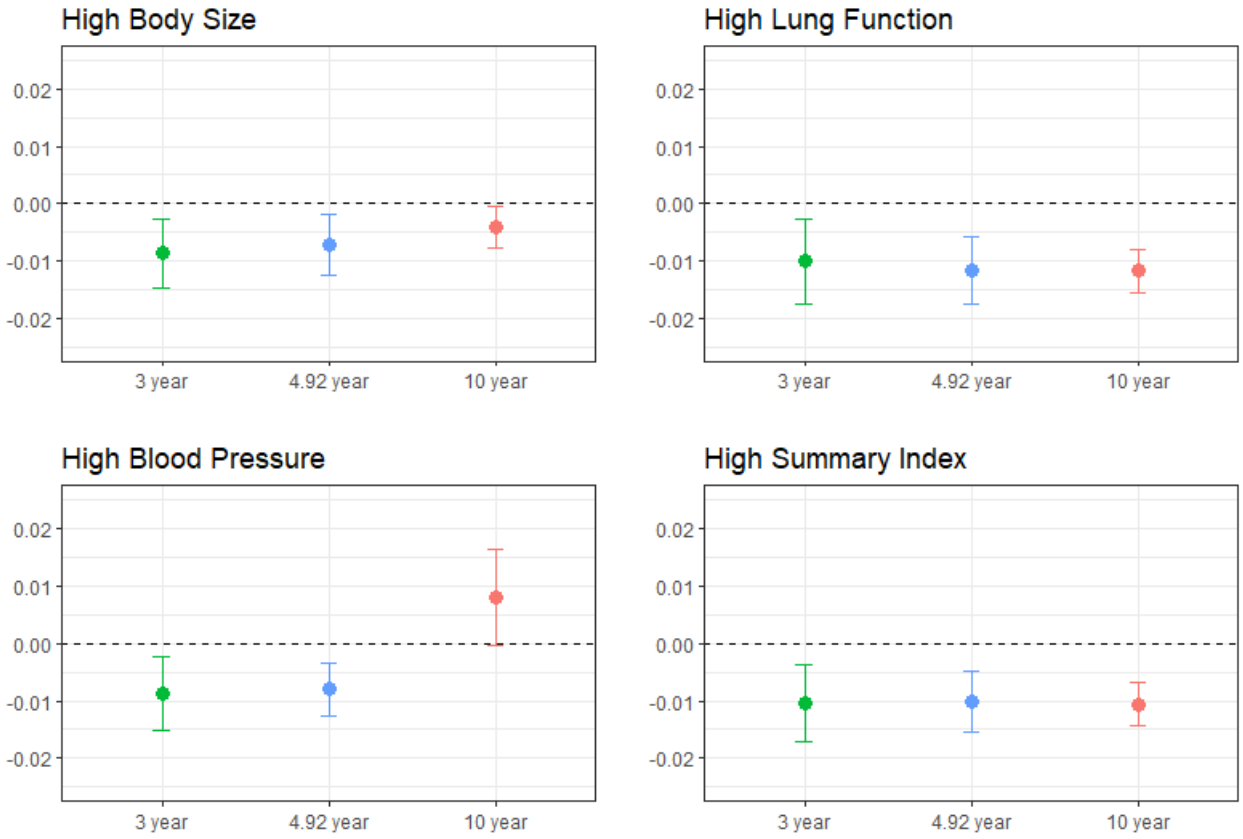


Figure 1.10: 3 Year vs 10 Year Bandwidth of the RDD for GxE Estimates of the Model for Indicator Values Higher than 3^{rd} Quartile of the Respective Indicator Distribution



Chapter 2

The Role of Gene-Environment Interaction in the Formation of Risk Attitudes

Abstract

Risk preferences are an important feature of every individual's decision-making process, which has been treated as an exogenous and fixed parameter of economic models for a long time. However, recent empirical economic literature suggests that risk aversion is in fact an endogenous variable that may change throughout life. Despite recent efforts to find factors that explain the formation of risk-preference, the empirical evidence is inconclusive and does not provide a clear picture of its architecture. I investigate risk-preference formation using a novel model class of gene-environment interactions. This allows me to study the relationship between genetic endowment and previous experiences of changes in the unemployment rate. This is the first study to shed light on the complementary role of socio-economic factors and genetic endowment in the formation of risk-preference. The aim of the paper is to deepen our understanding of the risk-preference formation process in a way which was not possible in previous studies that focused only on the socio-economic dimension of the problem. The results show that only individuals with low genetic predispositions for risk tolerance are affected by the changes in their unemployment-rate histories, whereas individuals with high genetic pre-

dispositions for risk tolerance are not significantly affected. Hence, this paper presents evidence that adverse economic situations accentuate the inequality in risk preferences that originates from initial genetic endowment. This may ultimately lead to an increase in inequality in health, wealth, income and other outcomes related to risk preferences.

2.1 Introduction

Attitudes towards risk (risk attitudes) fundamentally shape decisions. A risk aversion parameter is present in many economic models that aim to explain, for example, investment decisions (e.g. Brunnermeier and Nagel, 2008). Furthermore, recent empirical research has shown that measurements of fundamental risk aversion parameters are related to many important outcomes including migration (Jaeger et al., 2010), self-employment status (De Blasio et al., 2018), health outcomes (Dohmen et al., 2011), and many others.

Most current economic models assume that risk preferences are exogenous and fixed. However, the current empirical risk-aversion literature has presented evidence that changes in environments may influence preference for risk taking[1]. Specifically, the literature shows that, among other factors, changes in wealth, financial crises, and natural disasters lead to changes in risk attitudes (e.g. Page et al., 2014, Cameron and Shah, 2015, Cassar et al., 2017, De Blasio et al., 2018, Hanaoka et al., 2018). However, the empirical evidence is inconclusive and sometimes contradictory, which suggests that the relationship between environments and risk attitudes is more complex [2]. At the same time recent advances in behavioral genetics and genoeconomics document that individual heterogeneity in risk-taking behavior is influenced to some extent by genetic endowment (Cesarini et al., 2009, Benjamin et al., 2012). However, this strand of the literature does not take into account the economic dimension of the problem, or the possibility that the genetic and the economic sides may interact, which may skew their conclusions (Heckman, 2007, Mostafavi et al., 2020, Houmark et al., 2020).

To shine new light on the discussion and provide a possible explanation for the mixed evidence of the risk-aversion literature in empirical economics, this study brings the two

[1]By environments I mean the broad area of surroundings the individual is exposed to. These may include social factors, natural factors, and others. Including environments into which individuals self-select.

[2]For a comprehensive discussion of the topic see Chuang and Schechter (2015).

existing strands of literature together. It does so by investigating how risk attitudes are formed by earlier life experiences and by genes, and by the interaction of the two factors. If the genetic and the economic (or environmental) sides do interact, then not taking them both into account may lead to a skewed and incomplete explanation and picture of the architecture of risk attitudes.

Hence, the aim of this study is to help to understand the mixed evidence from the empirical risk-aversion literature and help illuminate the overall risk-attitudes formation problem. Additionally, the study addresses to what extent genetic endowment influences risk attitudes. Although there are several gene-environment interaction studies that have investigated the genetic architecture of many behavioral outcomes (i.e. phenotypes), this is, to the best of my knowledge, the first study that investigates how environment moderates the relationship between genes and elicited risk tolerance.

I use data from the Health and Retirement Study (HRS), which includes genetic information about the respondents and income lottery questions that elicit individual risk attitudes. As a measure of risk attitudes I create a risk-tolerance measure, which is the inverse of risk aversion. To measure the genetic endowment, I follow the literature and construct a single index measure called a polygenic score (PGS) (see e.g. Janssens et al., 2006, Belsky and Harden, 2019), which captures individual genetic predisposition for risk taking. As a measure of earlier life experiences I use individual life experiences on the labor market. To overcome the potential endogeneity arising from individual selection into experiences based on their risk tolerance, I exploit the variation across birth cohorts.

Specifically, I merge the HRS data with the national unemployment rate from the Bureau of Labor and Statistics and create a variable that captures the nationwide unemployment rate growth that each individual faced from the year they were born until the year of the survey. Moreover, following the seminal work of Malmendier and Nagel (2011) I allow for each historical unemployment growth rate to have a different weight, which is captured by a single parameter. This approach allows investigation of which life experiences are the most formative. The other benefit of this approach is that it allows the model to control for age and time fixed effects while estimating the role of life experiences based on the birth cohort variation.

The effects of life experiences related to economic conditions on risk attitudes are

subject to many studies. These life experiences can include financial crises, changes in wealth, GDP growth, and adverse development of stock markets. However, similarly to other domains the empirical evidence is mixed. For instance, in closely related papers, Malmendier and Nagel (2011) and Malmendier et al. (2011) suggest that experiencing unfavorable development of the stock market and financial crisis lead to higher risk aversion.

Another closely related paper by Levin and Vidart (2020) finds no evidence for the experience of GDP growth on risk attitudes but does find evidence that the volatility of GDP growth has a positive effect on risk averse behavior. In contrast, a substantial part of the literature suggests that the main source of heterogeneity in risk attitudes is due to persistent differences between individuals and that risk attitudes do not change in response to changes in income, unemployment status, or wealth (e.g. Sahm, 2012, Brunnermeier and Nagel, 2008).

The goal of this study is to investigate whether individuals with different genetic predispositions differ in their responsiveness to adverse economic shocks. The individual heterogeneity in responsiveness may be one source of the inconclusive evidence, because if the effect of an environment is moderated by genotype, then the estimates of the treatment effects of the environment depend on the genetic composition of the sample. More importantly, the current empirical models are not able to identify the effects of adverse economic conditions for individuals who were endowed with lower genetic predispositions for risk taking. The model I investigate may further illuminate this heterogeneity, which is important in order to better understand how adverse economic conditions affect inequality in risky behavior.

The results document that genetic predisposition for risk tolerance positively affects the elicited risk tolerance in my sample, and this effect does not disappear even if I control for past experiences. In terms of magnitude, the estimate of the effect of genes is similar to those for females, veteran status, and education. I do not find any evidence that life experiences have a dire impact on elicited risk tolerance. However, I do find evidence that the effect of life experiences varies with genotype. Specifically, I find that individuals at and below the median of the distribution of the standardized polygenic score for risk tolerance (low PGS) are substantially affected by past experiences, while individuals above the median of the PGS distribution for risk tolerance (high PGS) are

almost not responsive to past development of growth in the unemployment rate. The low PGS individuals are, on average, willing to sacrifice around 30 percentage points less of their income in order to play the risky option when they experience a one percentage point increase in unemployment growth than high PGS individuals.

Furthermore, I show that the results do not change substantively when I change the specification of the PGS group to quartiles. In this case the two bottom-quartile groups are significantly less risk tolerant than the highest PGS group (above the third quartile). Hence, my results show how genes and the environment jointly shape outcomes, implying that adverse life experiences can amplify genetic differences This may lead to an increase in the inequality in risk tolerance, which can ultimately be reflected in health, income and many other outcomes related to risk tolerance. Thus, the results help us to understand how changes in environments may further alter the differences between individual risk tolerance that arise due to genetic endowment.

The rest of the paper is structured as follows. Section 2.2 describes the background of genetic data analysis and related challenges. Section 2.3 describes the data and the main variables of interest and presents robustness checks. Section 2.4 presents the model and our identification strategy. Section 2.5 presents our findings and discusses their implications. Finally, section 2.7 concludes.

2.2 Genetic Markers in Economic Research

For many decades, scientists have discussed the respective roles of genes and environments in the formation of human outcomes. This debate is often summarised by the nature vs nurture dichotomy. However, recent evidence shows that this debate is obsolete and imprecise. Instead, new models have been proposed that capture a more nuanced relationship between outcomes, genes and socio-economic variables. These models allow for genes and environments to interact in the formation of important human outcomes (phenotypes) (Turkheimer et al., 2003, Rutter, 2006, Heckman, 2007, Biroli, 2015a, Houmark et al., 2020). Although the notion of gene-environment interaction has been discussed in the past, it was not then feasible to investigate their importance in empirical models due to the high cost of obtaining genetic data.

Recent decline in the price of genetic data collection and advances in genetics have

lead to projects called genome-wide-asscociation studies (GWAS) (Okbay and Rietveld, 2015, Locke et al., 2015, Linnér et al., 2018, Lee et al., 2018). These are hypothesis-free studies that aim to find robust relationships between genetic markers called single-nucleotide polymorphisms (SNPs) and a phenotype (Schmitz and Conley, 2017a). This and the recent availability of genetic data in many socio-economic surveys has opened doors to investigate the questions that were previously not feasible to explore.

The new empirical studies have started to provide more insights into how genes and environments complement or substitute each other in the vast areas of outcomes, which include smoking behavior (e.g. Schmitz and Conley, 2016d), education (e.g. Conley et al., 2015, Schmitz and Conley, 2017b), obesity (e.g. Biroli, 2015a, Schmitz and Conley, 2016b, Barcellos et al., 2018), or skills (Houmark et al., 2020). This paper is one of the first attempts to investigate how genes moderate the relationship between an environment and elicited risk tolerance.

One of the main challenges of incorporating genetic data into social-science research is their large dimensionality. Chabris et al. (2015) show that all behavioral traits are polygenic in their nature, which means that most outcomes are affected by many genetic markers with small effect sizes. Fortunately, external analyses called GWAS usually work with samples of hundreds of thousands or millions of individuals, so they are suitably powerful to estimate robust relatioships between individual SNPs and outcomes. The results of GWAS are used by practitioners who work with survey data to construct an index called a polygenic score (PGS)(see e.g. Janssens et al., 2006, Belsky and Harden, 2019). Many survey data now provide rich information about the respondents but lack the number of observations necessary in order to find robust relationships between high-dimensional SNPs and outcomes.

The attractiveness of a polygenic score is that it is a single index that captures individual genetic predispositions for a given trait. Hence, using a polygenic score instead of all the SNPs in empirical models, substantially decreases the dimensionality of the models (from several thousand to only one variable). The basic ingredients of the PGS are the SNPs from the survey and SNP association coefficients from the GWAS. Humans possess a total of 23 pairs of chromosomes. This means that we have 2 versions of each SNP (one per chromosome). Consequently, SNPs can take on only three possible values: 0,1, or 2. The concrete realization of the SNP variable depends on how many risky alleles

a person has at a given SNP[3]. The PGS index is then commonly used as a variable that captures genetic endowment or genetic predispositions for a given outcome and can be interacted with a measure of an environment in the gene-environment (GxE) models.

2.3 Data

This paper uses data from the Health and Retirement Study (HRS). The HRS is a nationally representative sample of the elderly US population over the age of 50. It was launched in 1992 and its participants have been surveyed every two years since then. Each wave of interviews contains information about approximately 20 000 individuals. Additionally, every six years a new cohort of participants is introduced to the survey. This paper uses information about individuals belonging to the following cohorts: "Children of the Depression" cohort (CODA), born 1924-1930, "HRS" cohort (HRS), born 1931-1941, "War Babies" cohort (WB), born 1942-1947, "Early Baby Boomers" cohort (EBB),born 1948-1953, and "Mid Baby Boomers" cohort (MBB), born 1954-1959. The purpose of the HRS survey is to monitor the changes in economic conditions, health, and cognition in the aging population in the US (Sonnega and Weir, 2014). This paper only studies individuals born between 1929 and 1959. I do not include individuals born before 1929 because the imputed data on the unemployment rate from the Bureau of Labor and Statistics ends in 1929.

I employ an easy-to-use version of the publicly available HRS data, which is available at RAND HRS [4], together with information about individual genotype from the HRS and nationwide unemployment-rate data from the Bureau of Labor Statistics. The Bureau of Labor and Statistics provides information about the US unemployment rate only from 1947. In order to obtain precedent unemployment rates, I link the information from Lebergott (1948) about the measured unemployment rates to the updated information about the imputed unemployment rate from 1929-1946 [5].

[3]A risky allele represents a specific realisation of a SNP that increases the chance or amount of an outcome. For more information see Mills et al. (2020)

[4]The RAND HRS Data (Version P, 2016) was developed by the RAND company with funding from the National Institute on Aging and the Social Security Administration, Santa Monica.

[5]The imputation study is available in the publications section of the Bureau of Labor and Statistics

2.3.1 Analytical Sample

The HRS collected DNA samples from respondents during interviews in 2006, 2008, and 2010. The DNA was extracted from saliva and genotyping was conducted using the Illumina Omni2.5 Beadchip. Before the analysis, standard quality checks were applied to the genotype data. These checks include the SNP Hardy-Weinberg Equilibrium (HWE) p-value: $p < 0.0001$, the SNP missing rate $\leq$1%, individual missing rates $\leq$ 10%, and minor allele frequency $\geq$ 1%.

Besides the quality control mentioned above, this paper also uses information about the ancestry group provided by the HRS to select only individuals from European ancestry. Hence, the final sample consists only of individuals of European descent. This is because, as mentioned for example in Tishkoff et al. (2009), individuals with African ancestry form a group of genetically diverse populations. This diversity is reflected in different linkage disequilibrium patterns and different minor allele frequencies across populations. Consequently, these differences alter the relationship between a phenotype and SNPs that was established by a GWAS performed on a population of European descent. If the GWAS results were applied to construct a genetic risk score, also known as a polygenic score, in a sample of individuals from different ancestry groups, the resulting index of genetic risk would be noisy and lose a great deal of its predictive power.

In 1998 respondents from the birth cohorts used in the analysis were asked risk attitudes related questions. In 2000, one in twelve individuals were randomly selected to answer the questions. In 2004 and 2006 only individuals younger than 65 years were asked the questions, and in 2002 only individuals from the EBB cohort were asked these questions (Bugliari et al., 2016). Hence the final sample consists of 5243 individuals observed across 5 interview waves, which amounts to a total of 9937 observations. Table 2.1 summarizes the main variables of interest.

2.3.2 Measurement of the Variables of Interest

A. Risk Aversion measure

Unlike the traditional approach in economics, where preferences are perceived as fixed starting points and not malleable throughout life, a modern approach relaxes the fixed nature of preferences and rather perceives them as malleable psychological traits (e.g.

Table 2.1: Summary Statistics

	All	
	mean	sd
Risk Attitudes		
Gamble Questions	0.165	0.201
Genes		
PGS(standardized)	0.000	1.000
Demographics		
Educ(years)	13.591	2.390
Age	57.512	5.333
Household Income (yearly. thousands of USD)	89.849	189.191
Sex	0.593	0.491
Behavior		
Smoking Now	0.185	0.388
Smoking Ever	0.577	0.494
Self-employed	0.192	0.394
Risky Assets (thousands of USD)	69.371	452.622
Observations		

Loewenstein and Angner, 2003). For instance, Becker et al. (2012) show that economic preferences measures are likely to be complements of measures of the big five psychological traits in explaining real behavior. Although many economic studies treat preferences as malleable outcomes, it is still not clear which mechanisms lead to heterogeneity in individual risk aversion and to changes in it. Moreover, there is lack of consensus in the empirical literature about the effects of changes in various environments on risk aversion. One reason for the contradictory results is the imperfect measurement of risk aversion. Often the risk-aversion-survey measures also capture other factors such as expectations about the future, beliefs, status quo biases, and others. Therefore, to properly study risk aversion (or risk tolerance), it is necessary to have a clean measure of risk-taking behavior. In this study I use income-gamble responses, which allow me to estimate the effect of life experiences on a risk attitude measure net of expectations and other potential confounding factors. This is because the income gamble questions explicitly state the

probabilities of every scenario.

The main outcome variable of interest is a measurement of risk tolerance[6], which is derived from the responses to hypothetical HRS income lottery questions. The individual's risk attitudes are then induced from the responses. From 1992 until 2006 the HRS questionnaire included hypothetical lottery questions that asked participants to choose between a job with a certain income and one that has a 50% chance of doubling the individual's income and a 50% chance of cutting it by a certain amount. The risk attitudes can be induced from the switching point, where the respondents switch from the risky option to the safe option. From 1992 until 1996 the safe option meant staying in their current hypothetical job or switching to a risky job. In 1998 both alternatives were presented as new job offers and the questions included two additional categories. For the 1992-1996 waves, the specific wording of the question is:

-1cm-1cm *"Suppose that you are the only income earner in the family, and you have a good job guaranteed to give you your current (family) income every year for life. You are given the opportunity to take a new and equally good job, with a 50-50 chance it will double your (family) income and a 50-50 chance that it will cut your (family) income by a third. Would you take the new job?"*

If the individual accepts the new job then a new question is presented with a higher potential income cut. If the individual declines the new job a new question is presented with a lower potential income cut. In the 1998-2006 the questions were changed to:

-1cm-1cm *"Suppose that you are the only income earner in the family. Your doctor recommends that you move because of allergies, and you have to choose between two possible jobs. The first would guarantee your current total family income for life. The second is possibly better paying, but the income is also less certain. There is a 50-50 chance the second job would double your total lifetime income and a 50-50 chance that it would cut it by a third. Which job would you take − the first job or the second job?"*

If the safe option is chosen, then the individual is asked to choose again between the two jobs but this time the income cut is reduced to 20%. If the individual still prefers the safe option, he or she is presented with a new question where the potential income cut is reduced even further to 10%. If the individual chooses the risky option in the first

[6]Risk tolerance is the inverse of risk aversion.

question, then the potential income cut in the next question is increased to a half. If the individual chooses the risky option again then the potential cut is increased further to 75%.

This paper uses only information from the 1998-2006 waves because the questions are framed as a choice between two new hypothetical jobs, while the older version was framed as a choice between new and current hypothetical jobs. Hence, the older version is likely to capture both risk aversion and status quo bias (Kimball et al., 2008). This paper aims to obtain the cleanest available measurement of risk aversion that is available and hence uses only the later version of the questions.

The main outcome variable is constructed from the responses presented in table 2.2. From table 2.2 it follows that the values of the risk-attitudes variable coincide with the point at which the individual switches from the safe to the risky option (e.g. individuals with value 0.33 accept the safe option if the income cut is higher than one third and when it is one third, they switch to the risky option). Hence, the higher the value of this variable, the higher the degree of risk tolerance.

Although the income-gamble questions provide a clean measure of risk preferences, the measure may still contain other factors including e.g. beliefs (Levin and Vidart, 2020). Hence, it may be the case that the income gable questions are still an imperfect measure of risk-related behavior. However, Dohmen et al. (2011) show that the income gamble questions that are often used in surveys to elicit risk aversion are correlated with real risk taking behavior across many different dimensions. Furthermore, the authors validated similar survey questions to mine with experiments and showed them to be valid measures of risk attitudes. To present more evidence on this matter, I perform simple tests to examine to what extent the HRS risk attitudes measure is related to real outcomes. Table 2.3 shows correlations between risky investment decisions and hypothetical lottery outcomes. Table 2.3 shows that the income gamble questions are significantly related to actual behavior.

B. Life Experiences

The aim of this paper is to investigate how life experience, together with genetic endowment, shape risk tolerance. The risk attitudes measure that we introduce above is induced from gamble lotteries regarding hypothetical job offers. Therefore, it is natural

Table 2.2: Risk Tolerance measures

Value of the Outcome	Income Cut Accepted	Income cut Rejected
0.75	75%	-
0.5	50%	75%
0.33	33%	50%
0.2	20%	33%
0.1	10%	20%
0	-	10%

to choose a life-experience measure that is related to the labour market. One class of possible environmental variables would be individual job histories. However, these measures can suffer from endogeneity because, for example, more risk-averse individuals may select themselves into jobs with lower risk of being laid off and consequently experience fewer spells of unemployment. To deal with the bias emerging from the selection of individuals into labor market environments based on their risk attitudes, I use an aggregate measure of changes in unemployment rate as a life experience measure. To construct the environmental measure, I merge the data on the unemployment rate from the Bureau of Labor and Statistics with the HRS Data. Including the imputed data, I have information about the US unemployment rate from 1929 to 2006, which was the year of the last wave of the survey that contained the income-gamble questions. Figure 2.1 shows the development of the unemployment-rate growth in the US in this time period. Note that the paper focuses on changes in the unemployment rate rather than on its levels. This is because evidence from behavioral economics suggests that individuals are more susceptible to changes in their environments than to levels (e.g. Kahneman, 1979, Malmendier and Nagel, 2011).

I link the aggregate unemployment measure to every individual, such that the life experience of each individual is captured as the history of aggregate changes in the unemployment rate that they experienced since birth. Thus, the main source of variation in the life-experience measure comes from the differences in birth years. The resemblance of this approach to cohort studies implies several potential challenges as the main issue with cohort studies is the linear dependence of year, age, and cohort effects. Furthermore, omitting one of the effects may confound the estimate of the remaining two. For

Table 2.3: Correlation of Risk Tolerance and Various Real Life Outcomes

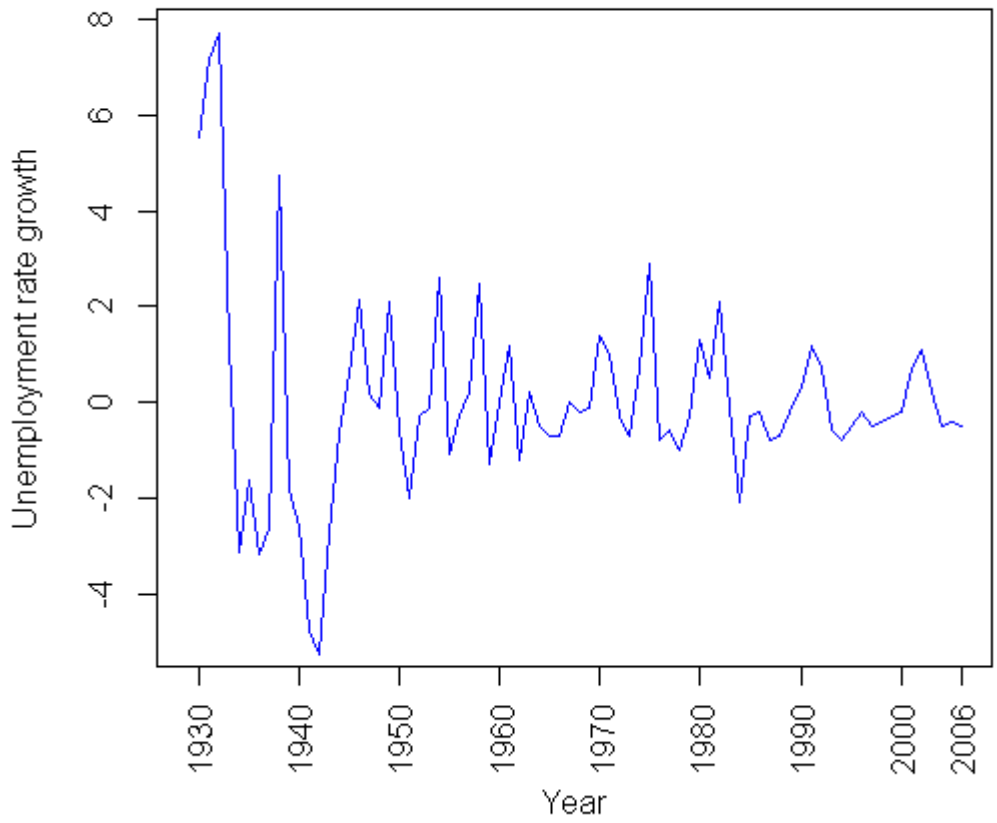
	Risk Tolerance
Intercept	0.042**
	(0.020)
Female	-0.053***
	(0.007)
Veteran	-0.022**
	(0.009)
Risky Assets	0.000
	(0.000)
Income	0.000
	(0.000)
Education Years	0.010***
	(0.001)
Ever Smoked	0.012*
	(0.006)
Smoke Now	0.003
	(0.008)
Self-employed	0.066***
	(0.008)
Nobs	9937
Adjusted R squared	0.048

Note: *Individual level clustered standard errors in parenthesis.* $***p < 0.01, **p < 0.05, *p < 0.1$.

example Dohmen et al. (2017) find that risk aversion increases with age, which is by definition correlated with birth year and the year of an interview. Similarly, the time of the interview may matter for the response. Some periods may, for example be less volatile than others, which could be projected into a higher degree of risk tolerance for everyone. Therefore, in order to separate the cohort effects from the time and age effects, it is necessary to control for age and time in the analysis. However, this is not feasible as the three variables are perfectly collinear (Heckman and Robb, 1985).

Another challenge linked to the historical unemployment rates is the dependence of the effect of life experiences on the stage of the life cycle the individual is currently in. For example, a significant increase in the unemployment rate may have a different effect for a

Figure 2.1: Growth of the US Unemployment Rate



forty-year old worker with a secured job than for a young adult who is entering the labor market. It follows that the experience measure should not be a simple average of growth in past unemployment rates, but the specification should rather capture the possibility that the unemployment rate shocks may have different effects in different periods of life. One extreme specification would be to allow for each year of realized unemployment rate growth to have a different effect. This specification is difficult to implement because not only would it imply a high-dimensional empirical model, but also the number of parameters to estimate would differ for each individual.

To address the two challenges mentioned in the previous two paragraphs, I use the

methodology developed by Malmendier and Nagel (2011). The idea is to build a single-parameter function of age, life experiences, and how long ago the experience was realized. This function is essentially a weighted average of past experiences where the weight depends on the age of the respondent, how long ago the experience was realized, and a parameter λ that allows for different weighting schemes:

$$A_{i,t}(\lambda) = \sum_{k=1}^{age_{i,t-1}} w_{i,t}(\lambda, k) * \Delta Unemployment_{i,t-k}$$

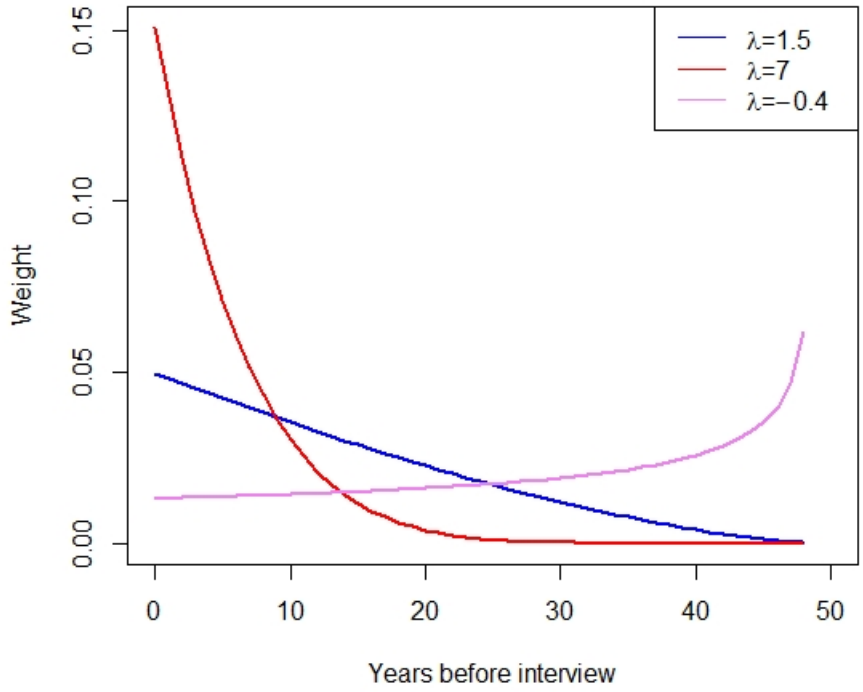
$$Where \quad w_{i,t} = \left[\frac{(age_{i,t} - k)^{\lambda}}{\sum_{k}^{age_{i,t-1}} (age_{i,t} - k)^{\lambda}} \right] \tag{2.1}$$

Equation (2.1) can flexibly capture the differential importance of historical growth in unemployment rates without imposing too complex model that may be difficult to estimate. The weight assigned to each past experience is captured by a parameter λ that can be estimated from the data and serves as a discount rate of past experiences. If $\lambda < 0$, then the weight is monotonically increasing and convex. Such a profile would imply that early life experiences matter more than more-recent ones. If $\lambda = 0$, then each past experience would obtain equal weight and equation (2.1) would become a simple average of past experiences. If $\lambda > 0$, then the weight is monotonically decreasing and concave, which means that the more-recent experiences matter more than the past ones. Finally, if $\lambda = 1$, then the weight is linearly decreasing. To better illustrate the logic behind the parameter λ, in figure 2.2 I show the weights as a function of time lags for different lambdas for a representative fifty-year-old individual.

Finally, because the expression in equation (2.1) is a function of the unemployment rate and a nonlinear function of age, it is possible to include controls for age and time in the main specification, as I show in section 2.4.

C. Polygenic Score As described in section 2.2 most of the outcomes (phenotypes) of interest to social scientists are in their nature polygenic. Hence, it is not generally feasible to estimate the effect of the individual genes in survey data because of the lack of power in these data sets. Instead, I adopt the standard approach from the behavioral genetic literature that studies the effects of gene-environment interactions and construct a single index measure that captures individual genetic risk, also known as a polygenic score.

Figure 2.2: Weights with different values for λ



To construct the polygenic score, I use the summary statistic from the recent GWAS conducted by Karlsson Linnér et al. (2019), together with information about genes from the HRS. I then construct the polygenic score as follows

$$PGS_i = \sum_j^J \beta_j SNP_{i,j} \tag{2.2}$$

where i stands for individual and j stands for a SNP. The β coefficients are taken from the GWAS summary statistic.

The discovery sample of the GWAS did not include the HRS data and was performed on individuals of European descent. I do not use any trimming methods for the score construction. Thus, the polygenic score includes all the SNPs that are in the HRS and overlap with the SNPs from the GWAS. Moreover, in the analysis, I work with a

standardized polygenic score. Table 2.4 presents evidence on the predictive power of the HRS polygenic score on both income-gamble questions and for some real-life outcomes. Each row of table 2.4 shows a coefficient on the polygenic score from a regression of an outcome on the polygenic score and the first ten principal components of the genetic-relatedness matrix, which controls for population stratification (Price et al., 2006).

Table 2.4: PGS on Risk Tolerance and Real Life Outcomes

Outcome:	PGS(standardized) Coefficient:
Lottery	0,008**
	(0.003)
Self Employed	0.018**
	(0.008)
Education (years)	0.05
	(0.04)
Ever Smoked	0.03***
	(0.009)
Smoke Now	0.013**
	(0.006)
Nobs	9937

Note: *Individual level clustered standard errors in parenthesis.* $***p < 0.01, **p < 0.05, *p < 0.1$.

2.4 Empirical Model

In the analysis I explore the effects of life experiences, and polygenic score, and their interaction, on the responses to income gamble questions. To model the life experiences, I adopt the method developed by Malmendier and Nagel (2011) (described in more detail in section 2.3.2). This section describes the heterogeneity in individual perception of the past within the framework of the established model described above. There are many potential reasons why individuals may remember their past differently, for example, they may idealize their past (Schutz, 1962). Alternatively, the heterogeneity in how individuals react to their experiences may come from their genetic endowment. This paper investigates whether individuals who have a higher genetic predisposition for risk

tolerance are resilient when it comes to adverse developments in the labor market, while individuals who are genetically predisposed to be less risk tolerant are more careful and more responsive to their past experiences. To investigate the heterogeneity in responsiveness to life experiences by genotype, I extend Malmendier and Nagel's (2011) current model by introducing a measure of genetic predisposition for risk tolerance and its interaction with the past-experiences measure. The general form of the model is described by equation (2.3).

$$\text{Risk}_{i,t} = \alpha_1 + \gamma_1 A_{i,t}(\lambda) + \sum_g^G \delta_{1,g} 1_{PGS_g}(PGS_i) + \sum_g^G \theta_{1,g} A_{i,t}(\lambda) \times 1_{PGS_g}(PGS_i) + X_{i,t}\beta_1 + \psi_{1,a} + \mu_{1,t} + \epsilon_{i,t} \tag{2.3}$$

where $A_{i,t}(\lambda)$ is the measure of life experiences defined by equation (2.1). Because in the case of this paper $A_{i,t}$ captures the individual unemployment rate history, in what follows I address this measure as "Unemployment" or "Unemployment history". $1_{PGS_g}(PGS_i)$ stands for the polygenic score dummy variable, which is equal to 1 if individual i belongs to a specific part of the distribution of polygenic score PGS_g and 0 otherwise. Finally, X stands for other covariates, including the first 10 principal components of the genetic-relatedness matrix, to control for population stratification. $\psi_{1,a}$ and $\psi_{2,a}$ stand for age fixed effects and $\mu_{1,t}$ and $\mu_{2,t}$ stand for time fixed effects. Hence, the model allows me to study the effects of the cohort differences in unemployment-rate growth net of time and age effects. Finally, $Risk_{i,t}$ stands for the responses to the income-gamble questions. I estimate the model using nonlinear least squares. To capture the nonlinear relationship between the interaction term and the risk-attitudes outcome, I assign individuals into G genetic groups based on their polygenic score and estimate the main effect and the effect of the interaction with unemployment rate histories for every polygenic score group.

$$1_{PGS_g}(PGS_i) = \begin{cases} 1, & \text{if } PGS_i \in PGS_g \\ 0, & \text{if } PGS_i \notin PGS_g \end{cases}$$

The step-function approach is meant to approximate the true conditional mean $E[Y|PGS, A(\lambda)]$. I choose to use the step function of the PGS because it fits the relationship better than a simple linear function of the PGS. Specifically, as I show in more detail in section 2.5, the relationship between the interaction of the PGS and unemploy-

ment rate history and risk attitudes is rather flat for high values of the PGS distribution and spikes upwards for lower values of the PGS distribution. To capture different possible functional forms, I consider 3 versions of model (2.3). First, I split the PGS by median into two PGS groups. Second, I split the PGS by tercile into 3 groups. Finally, I split the PGS by quartiles into 4 groups.

2.4.1 Identification and Basic Concepts

As stated above, the aim of the empirical model is to shed more light on how genetic endowment and life experiences shape human risk attitudes. To better understand the model a useful mental exercise is to link it to an ideal experiment. The simplest experiment I have in mind is to compare four hypothetical scenarios for individual risky behavior such that in all the scenarios individuals differ only in terms of their experience and genetic predisposition for risk tolerance. In the first scenario individuals would have low genetic predispositions for risk tolerance and experience adverse situations on labor markets[7]. In the second scenario the individuals would experience favorable situations on labor-markets but have low genetic predispositions for risk tolerance. In the third scenario the individuals would experience adverse labor market development and have high genetic risk for risk tolerance. Finally, in the fourth scenario the individuals would have high genetic predispositions for risk tolerance and experience favorable situations on the labor market. It is useful to think of the model as an approximation to the ideal experiment because it simplifies the analysis of the potential shortcomings of the model and its strength.

Model 2.3 allows for the heterogeneous effect of past experiences by genotype. Hence, it provides more insight into discrepancies between the results from previous empirical literature. The core idea behind gene-environment interaction is that some traits are mainly affected if both suitable environments and suitable genes occur at the same time (see e.g. Rutter, 2006). It follows that some environments may affect outcomes only for people who are genetically predisposed to be susceptible to changes in those outcomes.

[7]Adverse situations on the labor market may be embodied by several factors including a higher amount of the individual's unemployment spells, and a bad match between an employer and an employee that may lead to income loss or unemployment. In general I think of an adverse situation as a series of adverse shocks that can lead to either income losses and or unemployment

Thus, one possible reason for the mixed evidence in the empirical literature that investigates the causes of risk aversion may be the heterogeneous treatment effect by genotype. So far the main goal of the empirical risk-aversion literature was to find an average effect on the treated (ATT) which is an average of the effects over all genetic groups:

$$ATT = E_{G|T}[E[Y_1 - Y_0|T = 1, G]|T = 1]$$

However, if the fixed nature of risk aversion is relaxed, it is possible to perceive it as a skill that can be harnessed. Therefore, as has been shown, for example, in Heckman (2007), Houmark et al. (2020), the environment side and the genetic side may dynamically complement each other. If true then omitting either of the two dimensions from the main specification may lead to a skewed and incomplete picture of the risk preference formation. The ATT estimates from the contemporaneous empirical literature may, to some extent, depend on the genetic composition of the sample if, for instance, only individuals with a certain genetic predisposition for risk tolerance are susceptible to the change in the environment, or if the effects have an opposite sign for individuals with different genetic predispositions. At the same time, the behavioral genetics literature, which at best considers the environmental channel and the genetic channel to contribute independently to the risk-aversion preference, misses the point of possible dynamic complementarity of the two dimensions. The proposed empirical model (2.3) aims to illuminate the complementary relationship between initial genetic predispositions, environment, and risk attitudes. In the rest of this section I discuss potential problems and constraints of the empirical model, which allows me to better define the realm of insights the model can provide.

A potential problem that may arise when estimating individual life experiences from the survey data is reverse causality. In this setting, individuals may select themselves into environments based on their degree of risk tolerance and consequently differ in their life experiences. The empirical model overcomes this issue by focusing on the unemployment rate at a national level. Although it is possible that individuals may affect their environments I assume that no individual from the sample has sufficient power to affect the aggregate unemployment rate. The downside of this approach is that the model is only sensitive to nationwide historical changes in unemployment rates that

affected everyone in the given cohort.

The nature of an empirical model that focuses on differences between cohorts implies that the effect of life experiences may include other factors that differ between cohorts. For instance Levin and Vidart (2020) argue that differences in smoking behavior across cohorts may lead to differences in death rates and therefore also to differences in potential risk tolerances between treated and control groups [8]. I include in the model other explanatory variables that may capture additional differences between cohorts. However, it possible that some unobserved cohort differences are still captured by my measure of earlier life experiences.

Finally, given that this paper is a gene-environment interaction study, it also embodies the threats that arise from using genetic data as described in section 2.2. The main concern is the spurious correlation between genes and a phenotype due to population stratification, which is essentially a special case of omitted variable bias. To address the potential confounders that arise due to population stratification, I adopt the standard approach of the gene-environment interaction studies and include the first ten principal components in the regressions.

2.5 Results

In this section I document that genetic predispositions and past labor-market experiences contribute together to the formation of risk tolerance. Furthermore, I show that not taking into account the possibility of the complementary role of the two dimensions may lead to incorrect inference about the architecture of risk tolerance. To support this claim, I first consider a scenario that does not allow for the gene-environment interaction, thus reflecting the nature vs nurture dichotomy. Such a model allows life experiences and genetic predispositions to contribute independently to the risk tolerance measure. The results are shown in table 2.5. The results suggest that the polygenic score influences the responses to income-gamble questions. Specifically, a one standard deviation increase in the polygenic score leads to a 3.5 percentage points increase in the share of income the individual is willing to sacrifice in order to play the risky option in the lottery.

[8] I discuss the problem of differences in survival probabilities between birth cohorts in more detail in section 2.6

The coefficient of the past-life-experience measure is not statistically significant. Thus, the results based on a simple model that does not take into account gene-environment interactions suggest that past labor-market experiences play little role in forming risk tolerance compared to the genetic factor. The result that earlier life experiences do not contribute to risk attitudes, be it risk tolerance or risk aversion, is not uncommon in the empirical literature (Chuang and Schechter, 2015). At the same time the results in table 2.5 are in accordance with the traditional notion that risk preferences are predetermined and do not change over the lifetime. However, as I document in the rest of this section, this conclusion is imprecise as it does not allow for the more dynamic relationship between predetermined genetic predispositions and life experiences.

Following the discussion on the complementary role between genetic endowment and environments (or choices), I next estimate the gene-environment model 2.3, which captures the idea that genes and life experiences may complement or substitute each other in the formation of risk attitudes. Table 2.6 shows the results for 2 PGS groups[9]. Individuals below the median polygenic score are labeled as a group with low genetic predisposition for risk tolerance and individuals with a score higher than the median genetic score are labeled as a group with high predispositions for risk tolerance.

The main effect of the low polygenic score group is negative and significant, which is consistent with the results from table 2.5. The results show that individuals with low-polygenic-score predispositions to risk tolerance are willing to sacrifice, by 3 percentage points, less of their income in order to play the lottery than the high polygenic score individuals. The coefficient of life experiences, which is captured by past unemployment rate changes, and the lambda coefficient are not statistically significant. However, the coefficient on the interaction term is both statistically significant and large in magnitude.

Taken together, the results provide a more detailed picture of the formation of risk attitudes than models that take into account only either the genetic or socio-economic part of the problem and do not allow for these two strands to interact. The results from the gene-environment interaction model document that individuals with lower genetic predispositions to risk tolerance who experienced unemployment growth are willing to risk less than their counterparts who also experienced high unemployment growth but

[9]In section 2.6 I discuss additional model specifications and show that the main results hold across these.

belong to the group with high genetic predispositions for risk tolerance. The difference in responsiveness to an unfavorable past between the two genetic groups is a 37 percentage points share of income they are willing to sacrifice in order to play the lottery.

The results imply that individuals with a low polygenic score for risk tolerance are willing to sacrifice less of their income in order to play a lottery when they experience an increase in the overall unemployment rate than the high polygenic score group. At the same time, individuals with a higher polygenic score are almost non-responsive to unfavorable past experiences. While low-polygenic-score individuals are, on average, willing to sacrifice less of their income, by 37 percentage points, when they experience a 1 percentage point increase in the unemployment rate, high polygenic score individuals seem to be affected less by the past experiences[10].

The last parameter of interest is λ, which captures the relative importance of more-recent experiences relative to more distant ones. Table 2.6 presents suggestive evidence that individuals discount past histories relatively more than more recent ones. However, the estimate is rather noisy, which is reflected in its large standard error.

It follows from the discussion above that the effect of past experiences on risk attitudes is present only for individuals who are strongly genetically predisposed to be risk averse. Individuals who have average or even high genetic predispositions for risky behavior have a "thick skin", so for them past experiences play little role when they decide if they should take a risk or not. An important implication of these results is that negative life experiences may increase inequality between individuals who were born with different genetic endowment. The inequality may arise because risk attitudes are related to many real-life outcomes. The results also help to clarify the inconsistency of previous studies that try to estimate the effects of changes in environments on risk attitudes. By allowing for the interaction between genetics and the environment, the model uncovers an important feature of the role of past life experiences on risk attitudes that would not be possible under a simpler version of the model that does not include both dimensions and their interaction.

[10]The difference in the responsiveness by the PGS group can be seen by adding the estimates of the past unemployment-rate effect and the interaction effect together. However, this simple arithmetic does not account for the confidence intervals. Therefore, I conclude that high polygenic score individuals are, on average, less responsive to the unfavorable past experiences.

2.6 Robustness Checks

This section discusses potential threats to identification of the coefficients of interest and provides tests of these concerns. Moreover, this section tests different model specifications to test for robustness of the results.

To address the robustness of the results, I examine additional possible PGS step-functions. For that purpose, I next consider two additional versions of model (2.3). First, I split the polygenic score distribution into terciles. The low genetic predisposition risk group corresponds in this case to the observations below the first tercile; the average polygenic score group corresponds to observations between the first and the second tercile and the high polygenic score group corresponds to observations above the second tercile. Formally:

$$PGS_{Low} = \{PGS_i : PGS_i \leq Q(1/3)\}$$
$$PGS_{Average} = \{PGS_i : Q(1/3) < PGS_i < Q(2/3)\}$$
$$PGS_{High} = \{PGS_i : Q(2/3) \geq PGS_i\}$$

where $Q(p)$ is the corresponding quantile corresponding to probability p from $F[PGS]$. The finer structure helps to identify the nature of the relationship between the variables of interest. It also allows to further investigation of the relationship between polygenic score and risk tolerance [11].

The results for the PGS step function by terciles are presented in the first column of table 2.7. The results indicate that the low polygenic score group is less risk tolerant than the high polygenic score group. The average polygenic score group is more similar to the high polygenic score group in terms of risk tolerance. Table 2.7 demonstrates that there is neither statistical nor substantial difference between the average and high polygenic score groups. Thus, the results confirm the previous finding that the effect of polygenic score on risk tolerance is driven mainly by the low polygenic-score individuals. The interaction term between both the average and low polygenic score groups are both statistically

[11] To uncover the relationship precisely, we would need to have a dummy variable for each value of the polygenic score, but the score is continuous so this specification is not feasible.

insignificant due to low precision of the estimates. However, the point estimates of the two interaction terms suggest that the negative effect of past unemployment rate history is greater for the lowest PGS group, which is in accordance with the results based on the main specification presented in section 2.5. Individuals from the low PGS group are willing to sacrifice, by 25.6 percentage points on average, less of their income in order to play the lottery when they experience a percentage point increase in unemployment history.

Next, I consider the quartile based PGS step function. Formally:

$$PGS_{Low} = \{PGS_i : PGS_i \leq Q(1/4)\}$$
$$PGS_{LowAverage} = \{PGS_i : Q(1/4) < PGS_i < Q(1/2)\}$$
$$PGS_{HighAverage} = \{PGS_i : Q(1/2) < PGS_i < Q(3/4)\}$$
$$PGS_{High} = \{PGS_i : Q(3/4) \geq PGS_i\}$$

The results for the model with PGS by quartiles are presented in column 2 of table 2.7. The pattern of the effects of the PGS score groups is again in accordance with previous findings. Specifically, individuals from the lowest PGS group (below the first quartile of PGS distribution) are, on average, willing to sacrifice 4 percentage points less of their income to play the lottery than the highest PGS group (above the third quartile of the PGS distribution). Individuals from the low-PGS group (between the first and second quartile of the PGS distribution) are, on average, willing to sacrifice 3 percentage points less of their income to play the lottery than the highest PGS group. The average PGS group (between second and third quartiles of the PGS distribution) is very similar to the highest PGS group in terms of risk tolerance. The coefficient for the average group presents evidence that the difference in the income the individuals are willing to sacrifice in order to play the lottery, compared to the highest PGS group, is -0.8 percentage points. The coefficient is statistically insignificant and substantially small compared to estimates of the lower PGS groups.

The interaction effects present further evidence on the heterogeneous response to the unfavorable unemployment shocks by genotype. The effect of past unemployment histo-

ries is largest for the lowest PGS group (-0.47). The estimate suggests that individuals from the lowest PGS group (below the first quartile of the PGS distribution) are willing to sacrifice 47 percentage points less of their income to play the lottery, than the high PGS group when they have experienced a 1 percentage point increase in unemployment rate history. The interaction of the low-PGS group (between first and second quartile of the PGS distribution) suggests that individuals from this group are willing to sacrifice 43 percentage points less of their income on average to play the lottery, than the highest PGS group when they experienced a 1 percentage point increase in unemployment rate history. Both estimates are of substantial magnitude and similar to the interaction coefficient estimate from the main specification in table 2.6 (PGS groups by median). The interaction-effect estimates from the two low-PGS quartile groups are statistically significant on the ten percent significance level. This is mainly due to the low number of observations in both groups compared to the main specification where these two groups are merged into 1 (below median).

Nevertheless, together with the interaction effect estimate from column one, the results suggest that the interaction between the risk aversion PGS and unemployment rate history is negative and quite substantial in magnitude. Finally, the estimate of the interaction effect of the high average group (between the second and third quartiles of the PGS distribution) suggests that there is no significant difference between this group and the highest PGS group (above the third quartile of the PGS distribution). The discussion of different model specifications implies that individuals who are below the median of the PGS distribution are the most responsive to shocks in their unemployment histories, while those above the median are less so. Hence, the main specification includes the PGS step function by median.

Next I discuss a potential threat to identification that comes from the nature of the HRS data. Because the survey focuses on population aged 50+ and, as mentioned in section 2.3, this paper includes individuals born between 1929 and 1959, it is possible that some respondents may not have survived until the data collection phase. Domingue et al. (2016) show that the genotyped HRS sample differs in observables from the non genotyped sample. This suggests that the genotyped HRS sample suffers from a mortality related sample-selection problem. Moreover, Domingue et al. (2016) show that averages of various polygenic scores differ by birth cohort which indicates a genotype-based sample

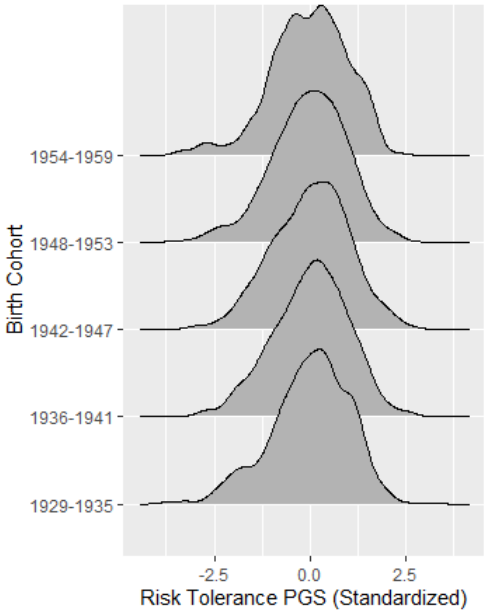
selection.

The mortality-based selection that may arise due to non-random deaths of HRS respondents across cohorts, may lead to a survival bias of the estimated GxE coefficient. The main concern is that more-risk-tolerant individuals may be more likely to die from their risky behavior. Consequently, the resulting bias of the GxE coefficient would be positive. The other possible scenario is that the less-risk-tolerant individuals may be more likely to die prematurely[12]. In this case the bias of the GxE coefficient would be negative. To mitigate the survival bias, it is more appropriate to use variation across birth cohorts that are close to each other in terms of birth year. Table 2.8 presents a simple test of differential survival rates of individuals with different risk attitudes.

First, I address the concern that genetic predispositions for risk tolerance may lead to higher risk tolerance, which may lead to higher death rates. This claim would be reflected in different polygenic score distributions among the cohorts. Table 2.8 does not support this claim as the means and standard deviations of polygenic scores are rather stable across birth cohorts. Additionally, figure 2.3 presents a simple comparison between PGS distributions across cohorts and further supports the claim that the PGS distribution is rather stable across cohorts. Second, table 2.8 suggests that the risk tolerance, measured by the income-gamble questions, is also stable across cohorts.

[12]For example they may achieve less wealth due to their low risk tolerance

Figure 2.3: Risk Tolerance PGS Distribution by Cohorts

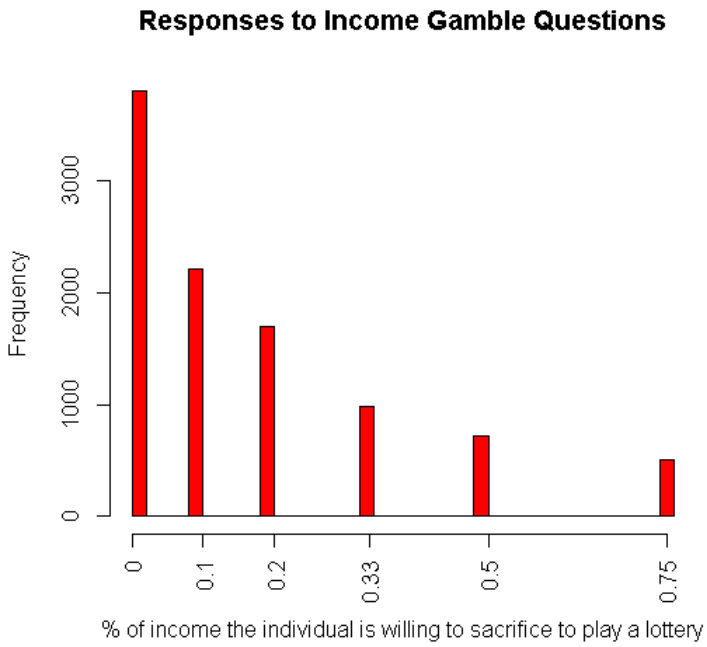


Another potential problem may arise in situations when individuals do not wish to answer all the risk-aversion related questions or if they answer them fast, without thinking too much about the lottery questions. Specifically in the income-gamble question, the individuals may incline to always choose one of the extremes. Such a measurement would be too noisy to provide any meaningful information about individual risk tolerance. In figure 2.4 I provide evidence that this is not the case in the HRS survey.

2.7 Conclusion

This paper investigates how adverse life experiences influence the inequality in risk tolerance that stems from genetic endowment. To identify the role of adverse experiences, I use data concerning the historical US unemployment rate. To capture the multidimensional problem of genetic predispositions for risk tolerance, I construct a polygenic score variable. Thus, the paper exploits the variation on the birth cohort level. I adopt a non-linear model developed by Malmendier and Nagel (2011) that allows me to estimate an effect of a cohort level variable while at the same time controlling for the age and time fixed effects.

Figure 2.4



Previous research shows that elicited risk tolerance is related to many real life outcomes, which means that heterogeneity in the elicited risk tolerance has consequences for inequality in health, income, or wealth. Consequently a substantial body of empirical economic research has investigated whether and how changes in the environment influence risk attitudes. Despite this effort, there is little concensus on the matter. Hence, the literature does not provide a clear message about whether preferences for risk taking can change during life and, if they do, in what direction. Parallel to the discussion in economics, there is a debate in behavioral genetics that investigates the genetic component of the variation in risk attitudes. However, this strand of literature does not take into account the potential interaction of genetic endowment with choices and environments, which makes the conclusion incomplete and skewed.

This paper combines the knowledge of both strands of the literature and aims to estimate a gene-environment model in the setting of risk attitudes, providing a more precise picture of the architecture of risk attitudes. Specifically, the paper investigates whether previous life experiences of changes in the unemployment rate affect risk tolerance and

how this effect varies for individuals with high and low genetic predispositions for risk taking. Thus, the paper provides an important insight into the relationship between adverse economic conditions and risk attitudes.

The results show that individuals with low genetic predispositions for risk tolerance are, on average, 30 percentage points more susceptible to adverse shocks on the labor market than individuals with high genetic predispositions for risk tolerance. Thus, the paper demonstrates that individuals who have low genetic predisposition to risk tolerance became less risk-tolerant in response to adverse changes on the labor market. In contrast, individuals with high genetic predisposition to risk tolerance are not significantly by such changes. Hence, unfavorable development on the labor market may further accentuate the inequality in attitudes towards risk that arise because of differences in initial genetic endowment. The increased inequality in risk aversion may lead to a further increase in income inequality. The findings also suggest that, in order to uncover a more detailed and precise picture of risk attitudes formation, it is necessary to take into account both genetic and socio-economic factors. Although these two factors may contribute independently to risk attitudes, an important feature uncovered by the paper is that they complement or substitute each other in the risk-attitudes-formation model.

Table 2.5: Role of life experiences and the PGS

Outcome: Risk Tolerance	
Constant	-0.093
	(0.084)
Unemployment	-0.119
	(0.540)
λ	24.67
	(0.997)
PGS(standardized)	0.035
	(0.00)
Female	-0.053
	(0.000)
Veteran	-0.016
	(0.000)
Income	2.894×10^{-8}
	(0.000)
Education(YRS)	0.008
	(0.001)
MSE	0.039
Age FE	Yes
Year FE	Yes
Principal Components	10
Observations	9937

Note: p-values in parenthesis. To calculate the p-values I used clustered bootstrapping as described in Cameron and Miller (2015).

Table 2.6: Life experiences shape risk attitudes for individuals with low genetic predispositions for risk tolerance

Gene-Environment Model	
PGS below median	-0.032
	(0.001)
PGS below median × Unemployment	-0.372
	(0.043)
λ	1.749
	(0.292)
Unemployment	1.027
	(0.412)
MSE	0.039
Age FE	Yes
Time FE	Yes
Principal Components	10
Observations	9937

Note: p-values in parenthesis. To calculate the p-values I used clustered bootstrapping as described in Cameron and Miller (2015).

Table 2.7: Model Specification check: Different versions of the PGS step function

	(1)	(2)
Gene-Environment Model	PGS by tercile	PGS by quartiles
PGS below 1^{st} tercile	-0.032	
	(0.004)	
PGS between 1^{st} and 2^{nd} tercile	-0.006	
	(0.583)	
PGS below 1^{st} tercile × Unemployment	-0.256	
	(0.240)	
PGS between 1^{st} and 2^{nd} tercile × Unemployment	0.066	
	(0.812)	
PGS below 1^{st} quartile		-0.041
		(0.001)
PGS between 1^{st} and 2^{nd} quartile		-0.033
		(0.005)
PGS bewteen 2^{nd} and 3^{rd} quartile		-0.008
		(0.536)
PGS below 1^{st} quartile × Unemployment		-0.477
		(0.072)
PGS between1^{st} and 2^{nd} quartile × Unemployment		-0.436
		(0.080)
PGS between 2^{nd} and 3^{rd} quartile × Unemployment		-0.155
		(0.614)
λ	1.818	1.700
	(0.327)	(0.249)
Unemployment	0.836	1.116
	(0.531)	(0.382)
MSE	0.039	0.39
Age FE	Yes	Yes
Time FE		Yes
Yes		
Principal Components	10	10
Observations	9937	9937

Note: p-values in parenthesis. To calculate the p-values I used clustered bootstrapping as described in Cameron and Miller (2015).

Table 2.8: Cohort Comparison

	All		1929-1935		1936-1941		1942-1947		1948-1953		1954-1959	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
Risk Attitudes												
Gamble Questions	0.165	0.201	0.146	0.201	0.170	0.208	0.167	0.205	0.165	0.194	0.153	0.184
Genes												
PGS(standardized)	0.000	1.000	-0.003	1.030	0.016	0.990	0.019	1.009	-0.028	0.992	-0.017	0.991
Demographics												
Educ(years)	13.591	2.390	12.874	2.504	13.214	2.480	13.580	2.423	13.967	2.230	13.662	2.226
Age	57.512	5.333	67.199	1.923	62.486	1.931	57.986	3.673	53.905	2.548	48.401	2.822
Household Income (yearly. thousands of USD)	89.849	189.191	56.352	63.534	79.982	192.318	85.135	94.735	103.808	273.237	109.574	162.435
Sex	0.593	0.491	0.637	0.481	0.529	0.499	0.586	0.493	0.580	0.494	0.843	0.364
Behavior												
Smoking Now	0.185	0.388	0.121	0.326	0.158	0.365	0.192	0.394	0.193	0.395	0.239	0.427
Smoking Ever	0.577	0.494	0.610	0.488	0.613	0.487	0.603	0.489	0.524	0.500	0.541	0.499
Self-employed	0.192	0.394	0.275	0.447	0.237	0.426	0.195	0.397	0.172	0.378	0.143	0.350
Risky Assets (thousands of USD)	69.371	452.622	106.783	367.301	65.064	216.952	81.038	636.184	56.262	298.878	46.846	407.299
Observations												

Chapter 3

New Method to study Gene-Environment Interaction in Empirical Economics Models

Abstract

Many socio-economic surveys have started to include genetic data about their respondents, which has lead to new studies that investigate how environments and choices interact with genetic endowments to form important economic, behavioral, or health outcomes. To cope with the high dimensionality of genetic data, researchers often summarize individual genetic information using an index for genetic predisposition called a polygenic score (PGS). The index exploits information from genome-wide association studies (GWAS), which establish robust correlations between genes and determinants of economic wellbeing, health, and inequality: including preferences, smoking, obesity, and education. The GWAS correlations are then used to construct a PGS for a given outcome, which then often serves as a variable in empirical economic models. This paper revisits the validity of the usage of PGSs in the framework of the widely used gene-environment models and in the non-interacted models. First, I demonstrate that gene-environment (GxE) interactions can severely distort the PGS index and thereby skew the results of important parameters of GxE studies. To correct the bias that stems from omitted GxE interaction in the GWAS, I propose a new two-step method to estimate GxE models and

their non-interacted counterparts. The new method requires only information from a GWAS to select the relevant genetic variables in the first step, and in the second step it estimates the full GxE model jointly. Unlike the standard method, this procedure does not rely on the GWAS estimates, which are often derived from a different population than in the survey used for the main empirical specification. Hence, the new method does not suffer from biases that stem from using GWAS estimates in the PGS index. In the empirical application I show that measurement error bias can significantly distort inference based on the standard GxE modelling approach. By not relying on the GWAS estimates, the new method expands the scope of the current survey-based studies that aim to incorporate genetic data into social research. The new method allows the study of outcomes for which suitable GWAS are not yet available.

Introduction

Recent technological progress has drastically reduced the cost of collecting genetic data, allowing genetic information to be part of a wide range of medical and social-science data sets. This technological progress has opened the door to new studies that aim to integrate genetic data into social-science research. Recent studies examine whether factors such as socio-economic status, education, or policy changes can mediate or amplify genetic dispositions (Turkheimer et al., 2003, Liu and Guo, 2015, Biroli et al., 2017, Barcellos et al., 2018, Barban et al., 2021). These studies often adopt methods from the natural sciences, which when applied to social-science difficulties may lead to methodological problems caused by endogeneity of choices.

This paper investigates the validity of the current modelling approach of genetic data in social-science research and proposes an alternative technique that aims to solve some of the shortcomings of the current approach. The two major classes of models that introduce genetic data into economics are gene-environment interaction models (GxE) and non-interacted models, which control for genetic endowment. In this paper I study the properties of these two model types under typical social-science settings, which may often suffer from omitted variable, selection, and measurement-error bias. Furthermore, I propose a new two-step method that aims to solve the issues of the current approach by using a split-sample approach. The two-step approach is necessary due to the high dimensionality of genetic data. Hence, the first step serves as a variable selection step and the second step estimates the actual model at hand.

Whereas my method deals with the high dimensionality of genetic data by estimating the model of interest in 2 steps, the current approach relies on external results, and therefore on additional assumptions about model specification that may lead to biased estimates when not met. First, the standard method uses results from external analyses called genome-wide association studies (GWAS) that establish robust correlations between many genetic variables called single-nucleotide polymorphisms (SNPs) and outcomes. Second, the standard approach constructs a polygenic score index (PGS) (see e.g. Janssens et al., 2006, Belsky and Harden, 2019) using the information about the relationship between SNPs and an outcome from a GWAS. The PGS is a weighted average of SNP variables weighted by their relative importance, which is captured by GWAS

coefficients. Finally, in a GxE model, the PGS is applied together with the interaction with an environment[1] factor such as mother's education.

This paper adds to the discussion presented in Groero (2022) by focusing more on the properties of standard gene-environments models when the outcomes of GWAS and the study of interest do not substantially differ. This paper shows that even when this condition holds, the standard use of PGS in the gene-environment models leads to measurement error bias that stems from applying GWAS coefficients as weights in the PGS construction. Moreover, contrary to the method introduced in Miao et al. (2022), this method still adopts the logic of a single index measure that captures genetic predispositions towards a certain trait. The method proposed in this paper imposes a standard PGS x E structure to the gene-environment model, which allows for simple interpretation of the interaction coefficient.

In the simpler model, the PGS is added as an additional control variable in an empirical model. The problem with this approach is that the first step and the model estimation step are built under different model specification assumptions and for different populations. Hence, the relationships between SNPs and an outcome in the first stage are likely to differ from the relationship in the second stage. For example, if a scientist wants to estimate whether mothers' education moderates or amplifies the role of genetic predispositions for education, they need to rely on established estimates of the relationship between individual SNPs and educational attainment to construct a PGS. However, these estimates could be based on populations with a different distribution of mothers' education than the survey population the scientist uses. For instance, in an extreme case, if a GWAS was performed on a population of individuals with only low-educated mothers, then using these estimates to create a PGS for individuals with highly-educated mothers would lead to a measurement error in the PGS and biased estimates of the model at hand.

I examine the nature of the bias the mis-measured PGS introduces to the GxE models. Specifically, I show that under GxE heterogeneity and correlation between genes and environments, the GWAS coefficients are biased. This bias depends on the GxE relationship with an outcome and on the variance-covariance structure of the genetic data.

[1]Although I talk about environments in the text, other socio-econmic variables including choices also belong into this category.

This implies that, under GxE heterogeneity, GWAS coefficients are not valid weights for the PGS used in GxE models. Consequently, including such a PGS in a GxE model will generally lead to systematic non-classical measurement error bias. To overcome the problem, I propose a new two-step approach, which produces a better measure of PGS by estimating the individual SNP weights within the GxE model. The improved PGS only requires information about the individual SNP significance level, which the GWAS provides. Hence, the new approach treats a GWAS as a variable selection step similarly to the principles of the sample splitting literature (e.g. Wasserman and Roeder, 2009).

Finally, in two empirical applications I demonstrate that such gene-environment (GxE) interactions exist and can severely distort the predicted PGS, thereby skewing the results of studies that simply interact a PGS with other variables. In the empirical part of the paper, I focus on two types of samples. First, I test the differences between the new method and the standard method in the Health and Retirement Study sample, which is relatively smaller and far from the respective GWAS population. Second, I run comparisons on UK Biobank sample, which is relatively large and closer to the target GWAS population. Comparing the performance of the two methods on these two types of samples allows me to assess whether the size of the bias of the standard method differs when the target sample is closer or more distant from the target GWAS sample. This provides more insight into which situation it still reasonable to use the standard approach and when it is preferable to employ the new method proposed in this paper.

3.1 The Role and Usage of Genetic Data in Social Sciences

For many decades scientists from a variety of fields have been interested in the role of genes versus the environment on human behavior, health, skills, and other outcomes. While early studies focused mainly on the dichotomy of the two sources (i.e. nature versus nurture), more recent studies studies show that the relationship between environments, or choices, and genes is much more complicated and in most cases the two sources may even complement or substitute each other (Turkheimer, 2000, Turkheimer et al., 2003, Rutter, 2006, Heckman, 2007). Hence, recent studies focus on the gene-environment interaction

(GxE) models to better understand how genes, together with environments and choices, contribute to the architecture of important economic, health, or behavioral outcomes. For recent examples of this literature see e.g. Turkheimer et al. (2003), Rutter (2006), Ridley (2003), Barcellos et al. (2018), Biroli (2015a), Liu and Guo (2015), Schmitz and Conley (2016d,c, 2017a,b), Domingue et al. (2015), Wedow et al. (2018), Barcellos et al. (2018), Bierut et al. (2018).

One of the largest challenges of incorporating genetic data into economic research has been their high dimensionality, cost of collection, and scarcity. This has changed in recent years. Due to the drastic decrease in the costs of genotyping individuals, scientists have formed large consortia that conduct large genome-wide association studies (GWAS). These studies establish robust correlations between genetic markers and outcomes including education (Okbay and Rietveld, 2015, Lee et al., 2018), obesity (Locke et al., 2015), smoking (The Tobacco and Genetics Consortium et al., 2010), attitudes towards risk (Linnér et al., 2018) and others. However, these studies establish only a basic correlation between individual genetic markers called single-nucleotide polymorphisms (SNP) and an outcome, without taking into account any interaction with the environment. Moreover, GWAS do not account for the correlation between socioeconomic variables with genes, which leads to omitted variable bias. Such a correlation may arise if, for instance, individuals self-select into environments based on their genetic endowment or make choices based on their innate abilities.

Another problem of the GWAS is that they are generally conducted using a large sample pool that does not necessarily represent the same population that the social scientist is usually interested in. Finally, the outcomes analysed in a GWAS often do not correspond exactly to those analysed by social scientists in the GxE models (or in simpler non-interacted models, which include only the PGS without the interaction term). Such an outcome mismatch leads to additional problems, discussed in Groero (2022). Hence, when applied to different populations, indices based on GWAS results may provide inaccurate measures of genetic endowment.

Nevertheless, social scientists use the information a GWAS provides to construct an index of genetic predisposition for an outcome called a polygenic score (PGS). The need to use the single index derives from the polygenic nature of most human traits (outcomes). Chabris et al. (2015) establish that most human outcomes are affected by many genetic

markers with small effect sizes. This presents a problem for standard empirical studies, which usually work with surveys that contain at most several thousands of observations, while there are billions of SNPs that need to be tested, and often hundreds or thousands of SNPs contribute to any given outcome. Therefore, the PGS is a neat way of reducing dimensionality (see e.g. Janssens et al., 2006, Belsky and Harden, 2019).

The basic version of the PGS is rather simple to construct. A reasearcher usually only needs access to a survey that includes genetic data about their respondents and GWAS summary statistics. The PGS is then constructed in the following way:

$$PGS_i = \sum_{j}^{J} \gamma_j^{GWAS} SNP_{j,i}$$

Where γ_j are the GWAS coefficients, $SNP_{j,i}$ is a particular realization of SNP j for individual i in a survey, and J stands for the total number of SNPs in the survey[2].

Even though PGSs are widely used, they have several shortcomings when applied to economic models. Some recent studies have pointed to the low predictive power of PGSs, which is a common consequence of measurement error. It has been shown that the predictive power of a PGS varies with the specification of the outcome model and the population it is applied to. This problem is likely to arise if the survey population of interest differs from the GWAS population (Mostafavi et al., 2020, Tropf et al., 2017). The heterogeneity of the genetic effects among environments raise important questions about the external validity of the GWAS coefficients. The problem of external validity arises from differences between the samples the genetic information was drawn from (survey sample) and the GWAS sample from which the correlations with outcome variables are obtained. These conditions are unlikely to hold in the presence of gene-environment interactions. This implies that, under GxE heterogeneity, the GWAS weights of the PGS are not generally applicable to other populations. This paper shows that, if the PGS is mismeasured, the GxE model estimated from a survey's data will generally suffer

[2]Note that each SNP represents a position on the DNA. Humans are diploid organisms, which means that we all have 2 versions of each SNP (one per chromosome). Hence, every human SNP can take on only three possible values: 0,1, or 2. The specific realization depends on how many risky alleles a person has at a given SNP, where risky allele means a specific realisation of a SNP that contributes to an outcome. For more information about genetic markers, see e.g. Mills et al. (2020)

from a measurement error bias that will depend on the relationship between genetic predispositions and environments. This bias differs from classical measurement error bias, which would generally lead to attenuation. Rather, the direction of the measurement error bias will depend on complementarity and the covariance structure of the genetic endowment and the environment.

3.2 The current PGS approach in the GxE models

Although studying the role of genes and their interactions with socio-economic variables is an important strand of research that connects biological markers to social-science outcomes. This research raises many conceptual and practical challenges. In the previous section I mentioned several potential problems that the usage of GWAS results in PGS construction may lead to. Nevertheless, PGSs are widely used and often interpreted as a genetic predisposition for a certain trait. As such, they are used in the (interacted) GxE models and in the (non-interacted) models that include only the PGS as a control variable without interaction.

Section 3.1 noted that a PGS is a weighted average of survey SNPs where the population of the survey generally differs from the GWAS population. As mentioned above, this may lead to problems of performance of the PGS in terms of predictive power but it also hinders the interpretation of the GxE model coefficients. The construction of a PGS implicitly assumes that the GWAS coefficients are valid for samples where both interacted GxE or non-interacted models are conducted. Moreover, using a PGS in both interacted and non-interacted models raises some substantive difficulties in the interpretation of the results. In Groero (2022) I argue that this may be a serious problem when the outcome in a GWAS differs from the outcome studied in an analysis for which the PGS was constructed. In this section I show that using GWAS coefficients as weights in PGS construction generally leads to measurement error bias of the GxE model coefficients.

The heterogeneity of the genetic effects among environments raises important questions of the external validity of the GWAS coefficients. The problem of external validity arises from differences between the samples which the genetic information was drawn from (survey sample) and the sample from which the correlations with outcome variables are obtained (the GWAS sample). These conditions are unlikely to hold in the pres-

ence of gene-environment interactions and if the outcome in the GWAS sample does not correspond to that in the survey sample.

To illustrate the point, consider a simple example where a researcher wants to study the effect of mothers' education on years of schooling for individuals with high and low genetic predispositions to achieve high academic outcomes [3]. To simplify the example further consider the case when mothers' education is binary and distinguishes mothers with a college degree from those with lower educational attainment. Following the standard method, a scientist would construct a PGS based on the educational attainment GWAS and interact it with mothers' education in a survey such as the Health and Retirement Study (HRS). Also, for the sake of argument, assume that the GWAS population consists of individuals with low-educated mothers only. Then, at best, the GWAS identifies the effect of genes for individuals whose mothers have low educational attainment. This means, that a follow-up study that aims to investigate the GxE interaction model can feasibly estimate the role of the PGS on education for individuals with low-educated mothers but is not able to identify this effect for these individuals, unless the role of genes is not heterogeneous in mothers' education. This is because if the role of genes depends on mothers' education, then the PGS for individuals with highly-educated mothers includes GWAS coefficients that are valid only for individuals whose mothers have low education. Hence, the PGS for individuals with highly educated mothers would be measured with an error. This case would then lead to a biased estimate of the GxE coefficient in the follow-up GxE study.

Below I discuss in detail the two reasons the GWAS weights are not generally applicable to the PGS in the survey samples that researchers typically use to estimate GxE models[4]. I study the consequences of not accounting for the heterogeneous genetic effects for the GWAS parameters $\hat{\gamma}_{GWAS}$ and for the GxE model parameters, β and ρ from equation (3.1). First, I consider the GWAS step by analyzing the asymptotic properties of the individual SNP coefficients under GxE heterogeneity. Second, I use the results from the GWAS step to analyze the consequences of the heterogeneity of genetic effects for the GxE analyses.

[3] A similar analysis is performed in section 3.5 and sheds more light on whether mothers' education complements or substitutes the role of genes in the education formation process.

[4] In what follows I consider the GxE model. However, the arguments also hold for a simpler non-interacted model.

A) GWAS step

The first step of the PGS construction is the GWAS step. Let N, J, and K denote the number of observations, the number of SNPs and the number of environments, respectively. Next, denote the SNP matrix as $G_{N\times J}$, the environment matrix as $E_{N\times K}$ and the interaction matrix as $(E \times G\Gamma)_{N\times K}$, where $G\Gamma$ represents the PGS with $\Gamma_{J\times 1}$ being the matrix of J SNP coefficients (genetic effects) γ_j. Then, a general true and estimated GWAS model can be described as follows:

$$
\begin{aligned}
&GWAS \quad Stage: \\
&\qquad Y = G\Gamma\beta + E\theta + E \times G\Gamma\rho + \epsilon && (3.1)\\
&\qquad Y = G\Gamma^{gwas} + v && (3.2)\\
&\qquad \mathbb{E}[W\epsilon] = \mathbf{0} && (3.3)\\
&\qquad \mathbb{E}[Gv] = \mathbf{0} && (3.4)\\
&\qquad W = [G \quad E \quad G \times E]
\end{aligned}
$$

where equation (3.1) represents the standard GxE model estimated in many gene-environment applications (e.g. Schmitz and Conley, 2016d,c, Barcellos et al., 2018) and is here assumed to be the true data-generating process, and equation (3.2) represents the GWAS model. In equations (3.1) - (3.4) E stands for the Environment, G for the matrix of SNPs, and $G\Gamma$ for the PGS. Hence, β, θ, ρ are coefficients of the PGS, environment, and the GxE interaction respectively. One way to think about the current GxE models is that they implicitly test for model mispecification of the GWAS model. However, they do so by taking the coefficients from the seemingly mispecified GWAS to construct the PGS. Equation (3.2) represents a multivariate GWAS that generates SNP coefficients. These are then used to construct the PGS [5]. Consequently, the probability limit of the estimated GWAS

[5]Note that we abstract from the additional problem that the GWAS coefficients from univariate regressions may be biased. Additional or different biases may arise if the true model does not conform to the assumption of a simple linear interaction. These issues are likely to make the problem even worse. Moreover, we acknowledge that the GWAS usually also includes other variables such as sex, age, and principal components of the genetic relationship matrix. In this example we abstract from this as it does not affect the results in any significant way.

coefficients approach is as follows[6]:

$$\text{plim} \quad \widehat{\Gamma^{gwas}} = \Gamma\beta + plim\left(\sum_i G_i^T G_i\right)^{-1} \sum_i G_i^T E_i \theta + plim\left(\sum_i G_i^T G_i\right)^{-1} \sum_i G_i^T \left(E_i \times G_i \Gamma\right) \rho \tag{3.5}$$

Equation (3.5) shows that not including the environment variables from the GxE studies may lead to omitted variable bias of the GWAS coefficients [7]. The bias depends on two terms. First, if the environment is correlated with genetic endowment G, then the GWAS coefficients suffer from omitted variable bias. The specific form of the bias in $\widehat{\Gamma^{gwas}}$ depends on the relationship between E with G, E with Y, E with G, and $E \times G$ with Y. Thus, the bias in the estimated SNP coefficients $\widehat{\Gamma^{gwas}}$ is a complex function of environments or choices, including parental investments, individual life experiences, and initial genetic conditions. For example, a possible omitted variable may arise if individuals with unfavorable genetic predispositions for education may receive more attention from their parents, and hence receive more parental investments in terms of time or other resources.

In a model that would investigate the link between a SNP and education the lower genetic predispositions for education are negatively correlated with the amount of parental investment. This would lead to a negative omitted variable bias, provided that higher parental investment leads to higher education. Under the typical omitted variable scenario, it would be enough to either include the omitted variable in the regression or to exploit the variation in G that is exogenous. GWAS studies aim to partially solve the issue by including the principal components of the genetic relatedness matrix that controls for population stratification (Price et al., 2006), which is one factor that leads to gene-environment correlation. More recently, GWAS research has focused on family data samples in order to better control for confounding factors such as population stratification, assortative mating or omitted parental genotype (Kong et al., 2018, Young et al., 2019). Even though these solutions alleviate some biases, neither of these adjustments solves the omitted variable bias effectively. For a more rigorous analysis, one would need to understand the data underlying the model and ideally include E in the baseline GWAS

[6]The derivation of equation (3.5) is in section A of the appendix

[7]Note that in the best case scenario of no bias, GWAS are able to identify the $\Gamma\beta$ term, which is a joint effect of any given SNP multiplied by the overall effect of the PGS.

model.

However, it is important to note that the bias in equation (3.5) also depends on the interaction term. Analogous to the argument presented in Solon et al. (2015) and Deaton (1997), if the true genetic effect is heterogeneous then the GWAS identifies a sample weighted average of the heterogeneous genetic effects that is generally not the true average genetic effect.

Consider the case where the recent adjustments (e.g. including family fixed effects) deal with the omitted variable bias discussed above. Then the environment is uncorrelated with the genetic endowment.

$$Assumption \quad 1: \quad \mathbb{E}[EG] = 0$$

Moreover, assume that the true data generating process (DGP) can be described by equation (3.1). Then the population average genetic effect is equal to:

$$\mathbb{E}[\delta_i] = \Gamma\beta + \Gamma\mathbb{E}[E]\rho \tag{3.6}$$

Then equation (3.5) writes as:

$$plim \quad \widehat{\Gamma^{gwas}} = \Gamma\beta + plim\left(\sum_i G_i^T G_i\right)^{-1} \sum_i [G_i^T G_i \Gamma E_i]\rho \tag{3.7}$$

Equation (3.7) implies that under the GxE model (3.1) the SNP coefficient estimates are a weighted average of the heterogeneous genetic effects $\Gamma_i = \Gamma\beta + \Gamma E_i\rho$. Importantly, the average genetic effects identified by a GWAS depend on the distribution of the environment in the respective sample and on the conditional SNP variance-covariance matrix $G^T G$ [8].

Equation (3.7) suggests that under the GxE heterogeneity described by model (3.1) the estimated average genetic effects in a GWAS sample may not identify the population-average genetic effects. To better illustrate the problem and to analyze the sources of the potential bias, consider a simpler case where E is just one discrete variable. For example

[8] In section B of the appendix consider a case of independence of E and G and show that even under this strong assumption the estimated coefficients may still lead to biased average genetic effects

E can be a policy that increases years of schooling. In a typical GxE study a researcher may be interested in estimating the heterogeneous effect of such a policy by genotype on future income or health. However, if genetic effects on outcomes of interest depend on years of schooling during childhood and adolescence, then the estimated average genetic effects identified by equation (3.2) depend on the distribution of years of education in the GWAS sample.

To formalise the argument suppose that E is discrete and can take on a limited amount of possible values E_l such that $l = \{1, 2, ..., L\}$, representing for example whether a person was treated by the policy and was consequently exposed to more years of education. Then $\Gamma_i^{gwas} = \Gamma_l^{gwas} = \Gamma\beta + \Gamma E_l\rho$ [9]. Next, denote the GWAS sample size by N^{gwas} and the sample size of each l group by N_l^{gwas}. Finally, suppose that as the sample size N^{gwas} grows, the proportions of each l groups remain the same. Then following the result of Deaton (1997), equation (3.7) can be rewritten as a weighted sum of genetic effects in the GWAS sample.

$$\text{plim} \quad \widehat{\Gamma^{gwas}} = \Gamma\beta + plim\left(\sum_{l=1}^{L} \frac{N_l^{gwas}}{N^{gwas}}\Omega_l^{gwas}\right)^{-1}\left(\sum_{l=1}^{L} \frac{N_l^{gwas}}{N^{gwas}}\Omega_l^{gwas}\Gamma E_l\rho\right) \tag{3.8}$$

where I denote the variance-covariance matrix of SNPs as $\frac{1}{N_l^{gwas}}G_l^T G_l$ as Ω_l^{gwas} [10].

Formula (3.8) illustrates several important points. First, the GWAS estimates of the genetic effects are not consistent estimates of the population average genetic effect (3.6) unless $\rho = 0$, which is equivalent to saying that the genetic effects are homogeneous and equal to $\mathbb{E}[\delta_l] = \Gamma\beta$. Next, if $\rho > 0$ then the environment and the genetic endowment complement and reinforce each other, which will lead to an overestimation of the average genetic treatment effect. If $\rho < 0$, then the environment and genetic endowment mitigate each other, which will in turn lead to underestimation of the average genetic effect. This suggests that the GWAS estimates and the biases of the PGS weights depend on the a priori complementarity between genetic endowment and the environment.

On a more technical note, the bias in the estimated average genetic effects also de-

[9] Note that in this case ρ is just a scalar. Furthermore, for simplicity I consider a case where the heterogeneous effect is linear in E. The argument presented here applies even to a more realistic example of a fully saturated model where $\Gamma_l^{gwas} = \Gamma\beta + \sum_{l=1}^{L}\Gamma E_l\rho_l$

[10] In section B of the appendix I consider a special case where E takes on only 2 values.)

pends on the different genetic variance-covariance structure Ω_l^{gwas} among different groups of the environment. Interestingly, the bias does not disappear even if the genetic correlation structure is similar for different values of the environment. To see this, consider a case where $\Omega_l^{gwas} = \Omega^{gwas}\ \forall\ \text{l}$. Then $plim\ \ \widehat{\Gamma^{gwas}} = \Gamma\beta + \sum_{l=1}^{L} \frac{N_l^{gwas}}{N^{gwas}} \Gamma E_l \rho$. Therefore, even if the variance-covariance structure of the SNPs is the same across environments, the GWAS may still yield inconsistent estimates of the average genetic effects because the proportions of the environmental groups may differ from those of the population. However, if the G and E terms are independent, then a proper weighting of the inverse population shares would lead to a consistent estimate of the population-average genetic effect.

Finally, equation (3.8) implies that (unless $\rho = 0$) the GWAS estimates of the genetic effects will in general identify a different weighted average of the individual genetic effects than a GWAS performed on a survey sample used to estimate the GxE models. Consequently, the GxE applications, which aim to test for $\rho \neq 0$ paradoxically use PGSs that are built using SNP coefficients, which would be correct for a given survey sample only if $\rho = 0$. It is important to note that while ρ captures the true relationship between the GxE interaction and an outcome, for example the role of a policy that increases years of schooling on the relationship between high academic achievements and genetic predispositions for education, the social scientist is only able to estimate $\hat{\rho}$ in a GxE model. An important consequence of this approach is that $\hat{\rho} = 0, \rho \neq 0$ may arise. I illustrate the problem in more detail below.

B) GxE analysis step

The problems presented above imply that the PGS weights researchers should use for a GxE model in a survey, $\widehat{\Gamma^{Survey}}$, likely often differ from those that are actually used, $\widehat{\Gamma^{gwas}}$. In consequence, the researcher estimates a mispecified model of the following form:

$$\text{Y}_i = \alpha + \left[\text{PGS}_i^* + G(\Gamma^{gwas} - \Gamma^{Survey})\right]\beta + E_i \times \left[\text{PGS}_i^* + G(\Gamma^{gwas} - \Gamma^{Survey})\right]\rho + E_i\theta + \varepsilon_i \tag{3.9}$$

where $\text{PGS}^* = G\Gamma^{Survey}$ is the true PGS for the population of interest. This formalization shows that the PGS as currently constructed can be seen as a version of the correct PGS

that is affected by systematic measurement error. In line with this observation, some recent studies have pointed to the low predictive power of a PGS, which is a common consequence of measurement error. Its predictive power varies with the specification of the outcome model and the population it is applied to, which is likely to occur if the measurement error arises from differences in the model specification or if the population of interest differs from the GWAS population (Mostafavi et al., 2020, Tropf et al., 2017). Importantly, if the PGS is mismeasured, as described by equations (3.14) and (3.9), the estimated coefficients of the GxE model estimated on survey data will generally suffer from a measurement error bias that depends on the relationship between genetic endowment and environments. In section 3.5 I present an empirical example where the measurement error leads to attenuation of the PGS and the GxE interaction coefficients. However, in general the bias may not only lead to attenuation. Rather, the direction of the measurement error bias will depend on complementarity and the covariance structure of the genetic endowment and environment [11].

3.3 New Method to Estimate GxE Models

To address the problems raised in section 3.2 I propose to estimate the GxE model together with the SNP weights on the survey data. The new method consists of two steps that follow the logic of Wasserman and Roeder (2009). In the first step I select relevant SNPs using GS summary statistics. Contrary to the standard method, I do not rely on the estimated coefficients but rather on the p-values, which provide information about which SNPs are important predictors of a given outcome. Hence, the new method treats a GWAS as a variable selection device. A potential problem of GWAS is that their analyses usually run a SNP by SNP univariate regression, which may lead to omitted variable bias. Therefore, I adjust the GWAS SNP coefficients (and consequently the p-values) using a method presented in Yang et al. (2011).

In the second step I estimate the GxE model of the form:

$$Y = \beta_0 + \beta_1 SNP\Gamma + \beta_2 E + \rho E \times SNP\Gamma + \theta X + \epsilon$$

[11] For more technical details about the specifics of the measurement error bias in the GxE model please see section C of the appendix

$$Var(SNP\Gamma) = 1 \tag{3.10}$$

where $SNP\Gamma$ stands for the polygenic score index, which is a function of SNP variables and SNP coefficients Γ. The main difference between the standard PGS and the PGS in the new method is that while the standard method takes the vector of coefficients Γ from a GWAS, the new method estimates them in the main model specification. E represents the environment of interest and X other covariates. The normalization condition $Var(SNP\Gamma) = 1$ is necessary for identification of the model. This normalization is also applied in the standard approach.

Although the new method solves the issues connected with using the GWAS SNP weights in the PGS construction, it still relies on the GWAS p-values. Hence, as with any method, the new approach faces several limitations. The largest limitation stems from the fact that the new method relies on the ability of the GWAS to properly select important SNP variables. In the ideal case, where GWAS are able to identify all relevant SNPs, the new method leads to unbiased estimates of all coefficients of interest. However, if the GWAS are not perfect, then the new method will not include all the relevant SNPs and may attenuate the effect of the genetic component. Fortunately, the ability of GWAS to properly select relevant genetic variables has been gradually improving over the years (Tropf et al., 2017). Additionally, the new method does not use an efficient way to estimate standard errors. Therefore, I propose to make inference based on bootstrapping, which leads to inefficient but still consistent estimates and inference.

Equation 3.10 represents a general GxE model specification. In the rest of this paper I apply the method on two cases and compare the results to the standard method. First, I apply the method to a survey with a small sample size, which is representative of surveys social scientists usually work with. The important feature of the first survey is that its target population is not close to the GWAS population. Then I apply the method to a survey with a larger sample size, which also represents a population more similar to the GWAS sample. These two settings allow me to study the differences between the new and standard methods in two different settings, which are both relevant to applied researchers. First, I analyze the differences between the two methods in a population that differs from the GWAS population. Second, I analyze the differences in a population that forms part of a GWAS. In the analysis I expect the differences between the methods

to be larger in a sample that represents a population distant from the GWAS population. However, I show that there are significant differences between the methods even in the case when the GWAS and survey populations overlap to some extent.

3.3.1 Empirical Application

First, I estimate a simple GxE model using the Health and Retirement Study survey and information from the educational attainment GWAS (Lee et al., 2018). Following equation 3.10, I estimate the following GxE model:

$$\begin{aligned} &Educ = \beta_0 + \beta_1 SNP_{educ}\Gamma_{educ} + \beta_2 MEduc + \rho MEduc \times SNP_{educ}\Gamma_{educ} + \theta X + \epsilon \\ &Var(SNP_{educ}\Gamma_{educ}) = 1 \end{aligned} \tag{3.11}$$

where $Educ$ stands for respondents' education and $Meduc$ stands for mother's education. This is a standard model that investigates the interaction between genetic predispositions for education and mother's education level (e.g. Conley et al., 2015).

Next, I estimate the GxE model on UK Biobank data following the empirical strategy of Barcellos et al. (2018).

$$\begin{aligned} &BMI = \beta_0 + \beta_1 SNP_{bmi}\Gamma_{bmi} + \beta_2 ROSLA + \rho ROSLA \times SNP_{bmi}\Gamma_{bmi} + \\ &t_1 DoB + t_2 DoB^2 + t_3 DoB \times ROSLA + t_4 DoB^2 \times ROSLA + \alpha X + \epsilon \\ &Var(SNP_{bmi}\Gamma_{bmi}) = 1 \end{aligned} \tag{3.12}$$

Where ROSLA stands for a rise in school leaving age policy in the UK in 1972. The policy increased the earliest school leaving age from 15 to 16 years old. DoB stands for date of birth. The model is essentially a regression discontinuity model with the ROSLA cutoff representing the September 1957 birth cohort, which was the first to experience the policy change. The model includes a second order polynomial of the running variable date of birth (DoB).

The two applications represent two types of studies. First, the example that uses HRS

data represents a general setting in which the sample of interest to the scientist is different from the GWAS population that was used to generate the SNP coefficients. Hence, the amount of bias should be large in this setting. The second application represents a general setting where the sample of interest lies close to the GWAS population. The purpose of this exercise is to investigate whether the standard method can deliver estimates close to the new method estimates if the research of interest reasonably resembles the GWAS population. Such evidence would suggest that using the standard method can still deliver useful results in samples that are sufficiently large and close to the target GWAS population.

3.4 Data

In the empirical part of the paper I use 2 data sources to investigate how the method performs in data sets with different sample-size magnitudes and samples of populations with different relatedness to the GWAS population. First, I employ an easy-to-use version of the the publicly available Health and Retirement Study (HRS)[12]. The HRS is a nationally representative sample of the US population aged over 50. It first launched in 1992, since when its respondents have been interviewed on a biannual basis (Sonnega and Weir, 2014). An important feature of the HRS is that it contributes only rarely to GWAS samples. Hence, the target HRS population differs from the GWAS population quite substantially.

Second, I use information from the UK Biobank. The UK Biobank is a prospective cohort study of more than 500,000 respondents aged between 37 and 73 years at the time of recruitment between 2006 and 2010 (Sudlow et al., 2015). Due to the large amount of UK Biobank samples and its inclusion in GWAS its data represents a population that is closer to the GWAS target population.

[12]The RAND HRS Data (Version P, 2016) was developed by the RAND company with funding from the National Institute on Aging and the Social Security Administration, Santa Monica.

3.4.1 HRS Sample

Following a standard procedure in the genetic data analysis, I first select only individuals with European ancestry. This is because individuals from e.g. an African ancestry group differ in the distribution of individual genetic markers (Tishkoff et al., 2009). Hence, using a data set that combines multiple ancestries may lead to incorrect results. Additionally, I apply the following filters to the HRS genetic data: (i) SNP Hardy-Weinberg Equilibrium (HWE) p-value: $p < 0.0001$; (ii) SNP missing rate $\leq$1%; (iii) individual missing rates $\leq$ 10%; and (iv) minor allele frequency $\geq$ 1%. Finally, I use information from the 2010 wave, which includes information about both genetic data and the HRS survey questions. Thus, the final HRS sample contains 8260 observations.

I use the HRS data to investigate whether the proposed method of estimating GxE models differs from the standard GxE method in smaller samples. The specific empirical model for the application of the new method follows the debate on whether parental education interacts with genetic predisposition for education in the education formation process (e.g. Conley et al., 2015). Therefore, The model uses data about respondents' years of education, mothers' education, and genetic data.

Table 3.1: Descriptive Statistics: HRS

Variable	Mean	Standard deviation
Years of education	13.482	2.452
Mother's years of education	10.654	2.990
Female sex	0.580	0.494
Age	69.638	10.991
Number of observations: 8260		

3.4.2 UK Biobank Sample

The application of the new method follows the empirical strategy used in Barcellos et al. (2018, 2019). I exploit the increase in the ROSLA policy in the UK, which started in 1972 and affected individuals born after September 1957. The analytical sample consists of 255,395 individuals of European ancestry born in England, Scotland, or Wales between

September 1, 1947 and September 1, 1967. Finally, I apply filters to the genetic data that are similar to the HRS sample.

Following Barcellos et al. (2018), the outcome of interest is body mass index (BMI), constructed as in the original study. This version of the empirical GxE model uses a regression discontinuity design with the birth month year being the running variable and with the cutoff at September 1957.

Table 3.2: Descriptive Statistics: UK Biobank

Variable	Mean	Standard deviation
Body Size	-0.003	0.992
Male	0.448	0.497
Age	52.82	5.842
Wales	0.048	0.215
Scotland	0.089	0.285
Number of observations: 255,395		

3.5 Results

This section provides insight into the differences between the new two-step non-linear GxE method and the standard linear GxE model. In section 3.2, I argued that the coefficients of interest in the GxE model are likely biased due to measurement error in the PGS that stems from the inappropriate SNP weights γ. Section 3.3 introduced a new method that aims to deliver unbiased estimates of the GxE model under standard NLS consistency assumptions. This section applies the new and standard method into 2 different settings to investigate the nature of the measurement error bias. As mentioned above, the nature of the measurement error bias is not clear, since it depends on complex relationships between genes and environments. Thus, it is not clear if the bias generally leads to attenuation of the coefficients of interest.

First, I apply my method to the HRS data set and estimate a simple model described by equation 3.11. The model describes the relationship between the educational attainment (EA) PGS, mother's education, and years of education. The results, presented in table 3.3, suggest that The EA PGS positively affects individual years of education.

Specifically, a one standard deviation unit increase in the EA PGS leads to an increase in years of education by 0.9 years on average. At the same time the results suggest that a one year increase in a mothers' education decreases the role of the EA PGS by 0.03 years, which suggests that longer mothers' education compensates for poorer genetic endowment. However, the results on the interaction provide only suggestive evidence, as mothers' education is most likely endogenous. More importantly, table 3.3 documents that the standard method in this case leads to attenuation bias of both the EA PGS and the interaction coefficients. The third column of table 3.3 presents results from the statistical test of equality of coefficients generated by the new and the standard methods and shows that there is a statistically significant difference between the two methods. Moreover, as the first 2 columns of table 3.3 show, the difference in the coefficients is also substantial in magnitude.

Table 3.3: HRS: Standard GxE model attenuates the relationship between PGS and Education

Outcome: Education (Years)	New Method	Standard Method	difference P-value
PGS EA	0.935***	0.783***	2.119×10^{-11}
	(3.608×10^{-8})	(0.083)	-
PGS EA x Mother Education	-0.031***	-0.010	0.001
	(0.003)	(0.007)	

Significance levels: ***0.01 **0.05 *0.1
P-values of the coefficient tests are based on bootstrapped T test using 1000 bootstrap samples.
Standard errors: (i) OLS heteroskedasticity robust (ii) NLS bootstrapped with 1000 resamples.

Second, I apply the new GxE method to a larger UK Biobank data set. Table 3.4 presents the results of model 3.12. This specification captures the relationship between body mass index (BMI), genetic predispositions for high BMI, and a policy that raised the school leaving age (ROSLA). Table 3.4 shows two sets of results. First, I apply the method to a continuous and standardized BMI measure [13]. The results reported in the first two rows of table 3.4, show that the BMI PGS has a positive effect on BMI. Specifically, a one standard deviation increase in the BMI measure before the reform leads on average, to a 0.141 standard deviation increase in in the BMI measure. Furthermore, after the reform the relationship between the BMI PGS and BMI strengthened by 0.014

[13]The measure is constructed as in Barcellos et al. (2018)

standard deviation units on average. The first column of table 3.4 show the results based on the new GxE method, while the second column presents the results based on the standard GxE method. A close examination of the first two columns of table 3.4 implies that in this case the measurement error bias in the PGS leads to a positive bias of the PGS. While the PGS estimate generated by the new method is 0.141, the corresponding estimate of the standard method is 0.160. Interestingly, the measurement error bias of the interaction coefficient (PGS BMI x ROSLA Policy) is negative. While the new method delivers an interaction coefficient of 0.014, the standard method delivers 0.009. The differences in both the PGS and the interaction coefficients are substantially large and statistically significant.

The last two rows of table 3.4 show the results for BMI being larger than the third quartile of the BMI distribution, which points to an obesity status. The results again show that the PGS BMI has a positive effect on the probability of being obese. However, in this case the results suggest that the ROSLA policy decreased the role of genetic predispositions for obesity, which is consistent with the findings of Barcellos et al. (2018). As in previous cases, the more important analysis in this study is the difference of estimates generated by the new GxE method and the standard method. In this case the differences in both the PGS and the interaction coefficients are similar to the continuous BMI case, although the substantive difference in this case is smaller than in the model with a continuous measure of BMI.

The results from tables 3.3 and 3.4 confirm the findings of section 3.2, which argues that the standard GxE method leads to measurement error bias that is systematic and does not generally lead only to attenuation of the estimated coefficients. This section shows that the measurement error bias stemming from the mismeasured PGS has both substantive and statistical impact on the results. Therefore, a method that directly estimates the SNP coefficients in the main GxE model specification is preferable.

Next, I analyze the relationship between the individual SNP coefficients produced by the new and standard GxE methods. Figure 3.1 documents the relationship between the GWAS SNP coefficients and the new GxE method coefficients in the case of model 3.11, estimated using the HRS data. Figure 3.1 demonstrates that there is no statistically significant relationship between individual SNP coefficient estimates from the GWAS and from the GxE model estimated by the new method. Hence, the individual SNP weights

Table 3.4: UK Biobank: The standard GxE model leads to positive bias in the relationship between PGS and Education

Outcome: BMI (standardized)	New Method	Standard Method	difference P-value
PGS BMI	0.141***	0.160***	0.000
	(1.638×10^{-6})	(0.004)	-
PGS BMI x ROSLA Policy	0.014***	0.009**	0.998
	(0.004)	(0.004)	
Outcome: BMI (BMI $\geq 3^{rd}$ quartile)			
PGS BMI	0.054***	0.059***	0.000
	(6.602×10^{-9})	(0.002)	-
PGS BMI x ROSLA Policy	-0.004**	-0.008***	0.021
	(0.002)	(0.002)	

Significance levels: ***0.01 **0.05 *0.1
P-values of the coefficient tests are based on bootstrapped T test using 1000 bootstrap samples.
Standard errors: (i) OLS heteroskedasticity robust (ii) NLS bootstrapped with 1000 resamples.

generated by the two methods differ quite substantially.

Figure 3.2 repeats the analysis for model 3.12 applied to the UK Biobank data. In this case the relationship between the individual SNP weights produced by the new and the standard GxE methods are positively correlated. However, the correlation is only moderate, which again suggests that a PGS based on the GWAS coefficients leads to measurement error.

Figure 3.1: The relationship between the SNP GWAS coefficients and the new GxE method SNP coefficients

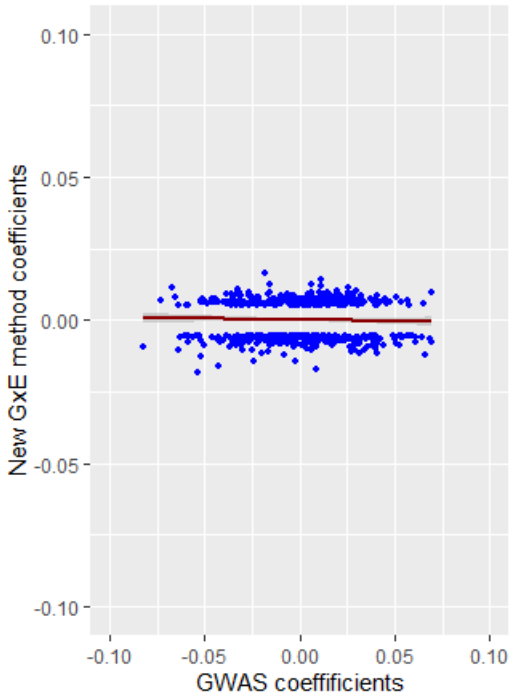
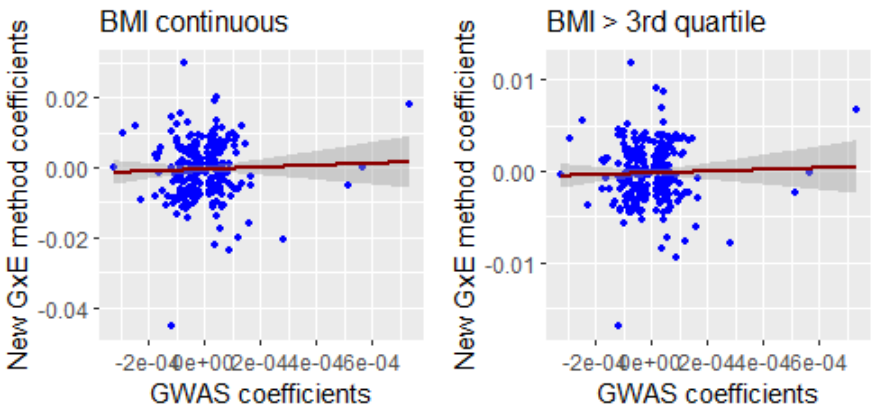


Figure 3.2: The relationship between the SNP GWAS coefficients and the new GxE method SNP coefficients



3.6 Conclusion

Recent technological progress has allowed researchers to incorporate genetic data into social science research, which offers unique possibilities to investigate questions that were not feasible to answer in the past. However, the new research often incorporates techniques from genetic data analysis that are valid within the natural science paradigm but may be challenging to use in the social sciences. This paper examines the properties of widely-used GxE models, which aim to research important questions about how genes and environments mitigate or amplify each other. Due to the high dimensionality of genetic data the standard GxE models rely on external GWAS, which provide information about correlation between every genetic variable SNP and an outcome such as education,

obesity, smoking, risk aversion, and others. These correlation coefficients are then used to construct a weighted average called polygenic score, which is a sum of risky alleles at each SNP weighted by their respective GWAS coefficients.

This paper argues that using GWAS weights in the PGS construction leads to measurement error bias in the GxE models and skewed results of important GxE model parameters. The GWAS weights are conducted on different populations from those in surveys that are often of interest to social scientists. Hence, if gene-environment interactions exist, the genetic correlation coefficients are not generally portable across populations or samples. Additionally, GWAS models, by construction, only provide information about average genetic treatment effects, while GxE models investigate heterogeneous effects by environment. This leads to additional problems as the models that choose PGS weights and the GxE model rely on different model specifications.

This paper shows that under GxE heterogeneity, GWAS coefficients are generally not consistent estimates of the population average genetic effects. Thus, GWAS implicitly work with the assumption of homogeneous genetic effects. The bias depends on the sign of the relationship between the GxE and an outcome and on the variance-covarinace structure of the genetic variables. Consequently, a PGS that uses biased GWAS weights is measured with error, which leads to a systematic measurement error bias in a GxE model.

To correct for the measurement error bias, I propose a new two-step approach that relies only on GWAS p-values to select a subset of significant SNPs with coefficients that are feasible to estimate in a survey. Thus, the new approach treats a GWAS as a variable selection step. In the second step of the new method I estimate the GxE model together with the individual SNP coefficients. Provided that the GWAS selects correct SNPs for a given outcome, the new approach leads to consistent estimates under traditional non-linear least squares consistency assumptions and under the argument presented in Wasserman and Roeder (2009).

In the empirical application of the paper I show that the GxE exists and that the standard method produces significantly different results from the new method. This suggests that the measurement error bias of the standard method leads to skewed results.

Apendix

A Derivations of formulae of section 3.2

Derivation of equation (3.5)

$$
\begin{aligned}
plim \quad \widehat{\Gamma^{gwas}} &= plim \left(\sum_i G_i^T G_i \right)^{-1} \left(\sum_i G_i^T Y_i \right) \\
&= plim \left(\sum_i G_i^T G_i \right)^{-1} \left(\sum_i G_i^T (G_i \Gamma^{gwas} + v) \right) \\
&= plim \left(\sum_i G_i^T G_i \right)^{-1} \left(\sum_i G_i^T (G_i \Gamma \beta + E_i \theta + (E_i \times G_i \Gamma)\rho + \epsilon_i) \right) \\
&= \Gamma \beta \\
&+ plim \left(\sum_i G_i^T G_i \right)^{-1} \left(\sum_i G_i^T E_i \theta \right) \\
&+ plim \left(\sum_i G_i^T G_i \right)^{-1} \left(\sum_i G_i^T (E_i \times G_i \Gamma)\rho \right) \\
&+ plim \left(\sum_i G_i^T G_i \right)^{-1} \left(\sum_i G_i^T \epsilon_i \right)
\end{aligned}
$$

Under assumption (3.3)

$$\text{plim } \widehat{\Gamma^{gwas}} = \Gamma\beta + plim \left(\sum_i G_i^T G_i\right)^{-1} \sum_i G_i^T E_i \theta + plim \left(\sum_i G_i^T G_i\right)^{-1} \sum_i G_i^T \left(E_i \times G_i \Gamma\right) \rho$$

B Extensions of section 3.2

Extension of the omitted interaction effect formula (3.7). Assume a special case where E and G are independent.

$$Assumption \quad B.1: \quad E \perp G = 0$$

The independence assumption allows further simplification of equation (3.7)[14]:

$$\begin{aligned} plim \quad \widehat{\Gamma^{gwas}} &= \Gamma\beta + plim\left(\sum_i G_i^T G_i\right)^{-1} \sum_i G_i^T G_i \Gamma \sum_i E_i\rho \\ &= \Gamma(\beta + plim \sum_i E_i\rho) \end{aligned} \tag{3.13}$$

Hence, in this case each of the estimated J SNP coefficients $\hat{\gamma}_j^{gwas}$ from a GWAS converges in probability to

$$plim \quad \gamma_j^{gwas} = \gamma_j \left(\beta + plim \sum_i E_i\rho\right)$$

Equation (3.13) shows that the independence assumption alleviates the problem presented in section 3.2 because the estimated average genetic effects do not depend on the conditional variance-covariance genetic matrix $G^T G$, Nevertheless, the estimated average genetic effects still depend on the distribution of the environment in the sample. Therefore, even under the independence assumption the GWAS estimates do not generally identify the population-average genetic effect and there is no reason to believe that the weighted average of the genetic effects identified by a GWAS is the correct weighted average of these effects that a researcher should use in GxE analyses performed in different survey samples.

Formula (3.8) represents a case where the heterogeneous effect is linear in E. To illustrate that the problem applies also to the saturated models, consider a case when E is categorical and can take only 2 values (i.e. 0 or 1) for $l = \{1, 2\}$, the above formula then

[14]The independence is needed because the bias term in equation (3.7) is a matrix of higher order moments. Therefore, the standard mean independence assumption is not enough to simplify the equation.

rewrites as:

$$plim \quad \widehat{\Gamma^{gwas}} = \left(\frac{N_1^{gwas}}{N^{gwas}}\Omega_1^{gwas} + \frac{N_2^{gwas}}{N^{gwas}}\Omega_2^{gwas}\right)^{-1}\left(\frac{N_1^{gwas}}{N^{gwas}}\Omega_1^{gwas}\Gamma\beta + \frac{N_2^{gwas}}{N^{gwas}}\Omega_2^{gwas}(\Gamma\beta + \rho\Gamma)\right)$$
$$= \Gamma\beta + \left(\frac{N_1^{gwas}}{N^{gwas}}\Omega_1^{gwas} + \frac{N_2^{gwas}}{N^{gwas}}\Omega_2^{gwas}\right)^{-1}\frac{N_2^{gwas}}{N^{gwas}}\Omega_2^{gwas}\Gamma\rho \tag{3.14}$$

C Measurement Error Bias in the GxE Model

To analyze the nature of the bias consider a case of only one environment interaction. First, introduce some notation. Denote the PGS constructed from the GWAS weights as follows:

$$\widetilde{PGS_i} = PGS_i^* + G_i(\Gamma^{gwas} - \Gamma^{Survey})$$

Then the measurement error in the interaction term is

$$E \times \widetilde{PGS_i} = E_i \times PGS_i^* + E_i \times G_i(\Gamma^{gwas} - \Gamma^{Survey})$$

Denote the PGS measurement error as $G_i(\Gamma^{gwas} - \Gamma^{Survey}) = G_i\Gamma^{\Delta}$ and the matrix of covariates as $\widetilde{X_i} = \begin{bmatrix}\widetilde{PGS_i} & E & E \times \widetilde{PGS_i}\end{bmatrix}$.

Then the asymptotic measurement error bias of the GxE model (3.9) $\mathbb{B} = \mathbb{E}[\begin{bmatrix}\widehat{\beta} & \widehat{\rho} & \widehat{\theta}\end{bmatrix}]^T - \begin{bmatrix}\beta & \rho & \theta\end{bmatrix}^T$ depends on the variance-covariance structure of the genetic matrix, the environment and the difference between the true SNP coefficients and the GWAS SNP coefficients. Given that we consider only one environment, the probability limit of the model estimates is written as

$$plim \begin{bmatrix}\widehat{\beta}\\ \widehat{\rho}\\ \widehat{\theta}\end{bmatrix} = \begin{bmatrix}\beta\\ \rho\\ \theta\end{bmatrix} + plim\left(\sum_i \widetilde{X_i}^T\widetilde{X_i}\right)^{-1}\sum_i \widetilde{X_i}^T(-G_i\Gamma^{\Delta})\begin{bmatrix}\beta\\ \rho\\ \theta\end{bmatrix}$$
$$= \begin{bmatrix}\beta\\ \rho\\ \theta\end{bmatrix} - plim\left\{\left(\sum_i \widetilde{X_i}^T\widetilde{X_i}\right)^{-1}\begin{bmatrix}\sum_i \widetilde{PGS_i}G_i\Gamma^{\Delta} & 0 & \sum_i \widetilde{PGS_i}E_iG_i\Gamma^{\Delta}\\ \sum_i E_iG_i\Gamma^{\Delta} & 0 & \sum_i E_i^2G_i\Gamma^{\Delta}\\ \sum_i \widetilde{PGS_i}E_iG_i\Gamma^{\Delta} & 0 & \sum_i \widetilde{PGS_i}E_i^2G_i\Gamma^{\Delta}\end{bmatrix}\begin{bmatrix}\beta\\ \rho\\ \theta\end{bmatrix}\right\}$$

Therefore, the bias $\mathbb{B}$ amounts to

$$\mathbb{B}=-plim\left\{(\sum_i \widetilde{X_i}^T\widetilde{X_i})^{-1}\begin{bmatrix} \sum_{i=1}^{N}\sum_{k=1}^{J}\sum_{j=1}^{J} G_{i,k}G_{i,j}\gamma_k^{gwas}\gamma_j^{\Delta} & 0 & \sum_{i=1}^{N}\sum_{k=1}^{J}\sum_{j=1}^{J} G_{i,k}G_{i,j}\gamma_k^{gwas}\gamma_j^{\Delta}E_i \\ \sum_{i=1}^{N} E_iG_i\Gamma^{\Delta} & 0 & \sum_{i=1}^{N} E_i^2G_i\Gamma^{\Delta} \\ \sum_{i=1}^{N}\sum_{k=1}^{J}\sum_{j=1}^{J} G_{i,k}G_{i,j}\gamma_k^{gwas}\gamma_j^{\Delta}E_i & 0 & \sum_{i=1}^{N}\sum_{k=1}^{J}\sum_{j=1}^{J} G_{i,k}G_{i,j}\gamma_k^{gwas}\gamma_j^{\Delta}E_i^2 \end{bmatrix}\begin{bmatrix}\beta\\ \rho\\ \theta\end{bmatrix}\right\} \tag{3.15}$$

Equation (3.15) has several important implications for the estimated coefficients of the GxE model. Note that if $\Gamma^{gwas} = \Gamma^{\Delta}$ then the bias $\mathbb{B}$ goes to 0 because all the terms inside the middle matrix of $\mathbb{B}$ will be 0. Importantly, using results from equations (3.8), (3.14), and (3.15) it is easy to see that this condition holds if

$$\begin{aligned} &plim\left(\sum_{l=1}^{L}\frac{N_l^{gwas}}{N^{gwas}}\Omega_l^{gwas}\right)^{-1}\left(\sum_{l=1}^{L}\frac{N_l^{gwas}}{N_l^{gwas}}\Omega_l^{gwas}\Gamma E_l\rho\right) = \\ &plim\left(\sum_{l=1}^{L}\frac{N_l^{Survey}}{N^{Survey}}\Omega_l^{Survey}\right)^{-1}\left(\sum_{l=1}^{L}\frac{N_l^{Survey}}{N_l^{Survey}}\Omega_l^{Survey}\Gamma E_l\rho\right) \end{aligned} \tag{3.16}$$

Therefore, the measurement error bias will disappear if there is no heterogeneity in the genetic effects ($\rho = 0$) or if the structure of the genetic relatedness matrix $\sum_{l=1}^{L}\frac{N_l^{gwas}}{N^{gwas}}\Omega_l^{gwas}$ resembles the structure of the survey genetic relatedness matrix $\sum_{l=1}^{L}\frac{N_l^{Survey}}{N_l^{Survey}}\Omega_l^{Survey}$. An important implication of the above thought experiment is that even if the environment is exogenous in both samples, i.e. the GWAS sample and the survey sample, the measurement error will not disappear. To see this, note that if the environment is orthogonal to the genetic structure of the two populations (or samples) then $\Omega_l^{gwas} = \Omega_l^{Survey} = \Omega$ for all l, which is not enough to satisfy the equality in equation (3.16), which in turn does not guarantee the measurement error bias $\mathbb{B}$ to be 0.

In the previous section I discussed that in the GWAS step it is unlikely that the environment is orthogonal to SNPs in the genetic matrix G. However, in the survey, researchers often employ identification strategies from econometrics that are built to identify causal effects. Note that if, in the survey, a researcher manages to satisfy assumption 2 (i.e. $E \perp G_j, \forall j$), then the middle matrix of equation (3.15) simplifies and

the measurement bias in $\widehat{\rho}$ and $\widehat{\beta}$ asymptotically approaches the following:

$$\mathbb{B}(\widehat{\beta}) = \beta \frac{1}{\sigma_G^2} \left(\sum_{j=1}^{J} \gamma_j^{gwas} \gamma_j^{\Delta} \sigma_{g,j}^2 + (J-1) \sum_{k=1}^{J} \sum_{j=1}^{J} \gamma_k^{gwas} \gamma_j^{\Delta} \sigma_{g,k,j} \right) \tag{3.17}$$

$$\mathbb{B}(\widehat{\rho}) = \rho \frac{1}{\sigma_E^2} \left(\sum_{j=1}^{J} \gamma_j^{gwas} \gamma_j^{\Delta} \sigma_{g,j}^2 + (J-1) \sum_{k=1}^{J} \sum_{j=1}^{J} \gamma_k^{gwas} \gamma_j^{\Delta} \sigma_{g,k,j} \right) \tag{3.18}$$

where $\sigma_{g,k,j}$ denotes the covariance between SNPs k and j and $\sigma_{g,j}^2$ denotes the variance of SNP j. Note that in order to obtain the results in (3.17) an(3.18) it is not enough to assume no correlation between E and G. The results in (3.17) and (3.18) are most likely to hold in an experimental setting where treatment and control groups are chosen completely randomly. Although equations (3.17) and (3.18) imply that even under independence the estimates of ρ and β in a GxE study would yield inconsistent estimates, in this special case it still possible to test for $\rho = 0$ even if the PGS is measured with error as described above. Even though ρ and β are generally biased, their fraction will, in probability limit, identify the true fraction up to a scale that is equal to the E and G respective variances.

$$\frac{\widehat{\rho}}{\widehat{\beta}} = \frac{\rho}{\beta} \frac{\sigma_G^2}{\sigma_E^2}$$

Therefore, if G and E are independent at least in the survey sample, a researcher may conduct a statistical test for $\frac{\rho}{\beta} = 0$, which would essentially test for $\rho = 0$[15]. It is important to acknowledge that this is a very special case that relies on a strong assumption that is unlikely to hold outside an experimental setting. Hence, a researcher should present strong evidence that assumption 2 is likely to hold in his or her setting, using propper tests such as the test of conditional independence introduced in Mittag (2018) . In the general case the measurement will lead to biased estimates of ρ, β, and θ in the GxE studies.

[15]Note that in this test it is important to assume that $\beta \neq 0$ which is a condition that is likely to hold in most GxE applications.

Bibliography

Albouy, V. and L. Lequien (2009). Does compulsory education lower mortality? *Journal of health economics 28*(1), 155–168.

Almond, D., J. Currie, and V. Duque (2018). Childhood circumstances and adult outcomes: Act ii. *Journal of Economic Literature 56*(4), 1360–1446.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association 103*(484), 1481–1495.

Attanasio, O., T. Boneva, and C. Rauh (2018). Parental beliefs about returns to different types of investments in school children: Hceo working paper 2018–032.

Barban, N., E. De Cao, and M. Francesconi (2021). Gene-environment effects on female fertility.

Barcellos, S. H., L. S. Carvalho, and P. Turley (2018, oct). Education can reduce health differences related to genetic risk of obesity. *Proceedings of the National Academy of Sciences 115*(42), E9765–E9772.

Barcellos, S. H., L. S. Carvalho, and P. Turley (2019). Distributional effects of education on health. Technical report, National Bureau of Economic Research.

Bates, G. P., R. Dorsey, J. F. Gusella, M. R. Hayden, C. Kay, B. R. Leavitt, M. Nance, C. A. Ross, R. I. Scahill, R. Wetzel, et al. (2015). Huntington disease. *Nature reviews Disease primers 1*(1), 1–21.

Becker, A., T. Deckers, T. Dohmen, A. Falk, and F. Kosse (2012). The relationship between economic preferences and psychological personality measures. *Annu. Rev. Econ. 4*(1), 453–478.

Becker, J., C. A. Burik, G. Goldman, N. Wang, H. Jayashankar, M. Bennett, D. W. Belsky, R. K. Linnér, R. Ahlskog, A. Kleinman, et al. (2021). Resource profile and user guide of the polygenic index repository. *Nature human behaviour*, 1–15.

Belsky, D. W. and K. P. Harden (2019, jan). Phenotypic Annotation: Using Polygenic Scores to Translate Discoveries From Genome-Wide Association Studies From the Top Down. *Current Directions in Psychological Science*, 096372141880772.

Benjamin, D. J., D. Cesarini, M. J. H. M. van der Loos, C. T. Dawes, P. D. Koellinger, P. K. E. Magnusson, C. F. Chabris, D. C. Conley, D. I. Laibson, M. Johannesson, and P. M. Visscher (2012, may). The genetic architecture of economic and political preferences. *Proceedings of the National Academy of Sciences*.

Bickel, H., J. Gerrard, and E. M. Hickmans (1953, oct). Influence of phenylalanine intake on phenylketonuria. *Lancet 265*(6790), 812–3.

Bierut, L. J., P. Biroli, T. Galama, and K. Thom (2018, jun). Childhood socioeconomic status moderates genetic predisposition for peak smoking. *bioRxiv*, 336834.

Biroli, P. (2015a). Genetic and Economic Interaction in Health Formation: the Case of Obesity. *Working Paper*.

Biroli, P. (2015b). Health and Skill Formation in Early Childhood. *Working Paper*.

Biroli, P., L. Bierut, T. Galama, and K. Thom (2017). GxSES in health behaviors: theory and empirics. *Working Paper*.

Bone, W. P., K. M. Siewert, A. Jha, D. Klarin, S. M. Damrauer, K.-M. Chang, P. S. Tsao, T. L. Assimes, M. D. Ritchie, and B. F. Voight (2021). Multi-trait association

studies discover pleiotropic loci between alzheimer's disease and cardiometabolic traits. *Alzheimer's research & therapy 13*(1), 1–14.

Boneva, T. and C. Rauh (2018). Parental beliefs about returns to educational investments—the later the better? *Journal of the European Economic Association 16*(6), 1669–1711.

Brunnermeier, M. K. and S. Nagel (2008). Do wealth fluctuations generate time-varying risk aversion? micro-evidence on individuals. *American Economic Review 98*(3), 713–36.

Bugliari, D., N. Campbell, C. Chan, O. Hayden, M. Hurd, R. Main, J. Mallett, C. McCullough, E. Meijer, M. Moldoff, et al. (2016). Rand hrs data documentation, version p. *RAND Center for the Study of Aging*.

Bulik-Sullivan, B., H. K. Finucane, V. Anttila, A. Gusev, F. R. Day, P.-R. Loh, L. Duncan, J. R. Perry, N. Patterson, E. B. Robinson, et al. (2015). An atlas of genetic correlations across human diseases and traits. *Nature genetics 47*(11), 1236–1241.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2015). Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association 110*(512), 1753–1769.

Cameron, A. C. and D. L. Miller (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources 50*(2), 317–372.

Cameron, L. and M. Shah (2015). Risk-taking behavior in the wake of natural disasters. *Journal of Human Resources 50*(2), 484–515.

Cassar, A., A. Healy, and C. Von Kessler (2017). Trust, risk, and time preferences after a natural disaster: experimental evidence from thailand. *World Development 94*, 90–105.

Cattaneo, M. D., M. Jansson, and X. Ma (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association 115*(531), 1449–1455.

Cesarini, D., C. T. Dawes, M. Johannesson, P. Lichtenstein, and B. Wallace (2009, may). Genetic Variation in Preferences for Giving and Risk Taking. *Quarterly Journal of Economics 124*(2), 809–842.

Chabris, C. F., J. J. Lee, D. Cesarini, D. J. Benjamin, and D. I. Laibson (2015). The Fourth Law of Behavior Genetics. *Current Directions in Psychological Science 24*(4), 304–312.

Chuang, Y. and L. Schechter (2015). Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. *Journal of Development Economics 117*, 151–170.

Clark, D. and H. Royer (2013). The effect of education on adult mortality and health: Evidence from britain. *American Economic Review 103*(6), 2087–2120.

Conley, D., B. W. Domingue, D. Cesarini, C. Dawes, C. A. Rietveld, and J. D. Boardman (2015). Is the effect of parental education on offspring biased or moderated by genotype? *Sociological Science 2*, 82.

Conti, D. V., B. F. Darst, L. C. Moss, E. J. Saunders, X. Sheng, A. Chou, F. R. Schumacher, A. A. Al Olama, S. Benlloch, T. Dadaev, et al. (2021). Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nature genetics 53*(1), 65–75.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Davies, N. M., M. Dickson, G. Davey Smith, G. J. Van Den Berg, and F. Windmeijer (2018). The causal effects of education on health outcomes in the uk biobank. *Nature human behaviour 2*(2), 117–125.

De Blasio, G., M. De Paola, S. Poy, and V. Scoppa (2018). Risk aversion and entrepreneurship: New evidence exploiting exposure to massive earthquakes in italy. Technical report, IZA Discussion Papers.

Deaton, A. (1997). *The analysis of household surveys: a microeconometric approach to development policy*. The World Bank.

Deaton, A. (2008). Income, health, and well-being around the world: Evidence from the gallup world poll. *Journal of Economic perspectives 22*(2), 53–72.

Dixon, J. B. (2010). The effect of obesity on health outcomes. *Molecular and cellular endocrinology 316*(2), 104–108.

Dohmen, T., A. Falk, B. H. Golsteyn, D. Huffman, and U. Sunde (2017). Risk attitudes across the life course.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011, 06). Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association 9*(3), 522–550.

Domingue, B., D. Belsky, A. Harrati, D. Conley, D. Weir, and J. Boardman (2016). Mortality selection in a genetic sample and implications for association studies. biorxiv, 049635.

Domingue, B. W., D. C. Conley, J. M. Fletcher, and J. D. Boardman (2015, jul). Cohort Effects in the Genetic Influence on Smoking. *Behavior genetics*.

Glied, S. A. and A. Lleras-Muney (2003). Health inequality, education and medical innovation.

Groero, J. (2022). Leveling health inequalities: Raising the school leaving age reduces the risk of diseases and severe medical conditions related to genetic endowmentleveling health inequalities: Raising the school leaving age reduces the risk of diseases and severe medical conditions related to genetic endowment.

Grossman, M. (1972). On the concept of health capital and the demand for health, 80 j. *Pol. Econ 223*(10.2307), 1830580223.

Hanaoka, C., H. Shigeoka, and Y. Watanabe (2018). Do risk preferences change? evidence from the great east japan earthquake. *American Economic Journal: Applied Economics 10*(2), 298–330.

Hartiala, J. A., Y. Han, Q. Jia, J. R. Hilser, P. Huang, J. Gukasyan, W. S. Schwartzman, Z. Cai, S. Biswas, D.-A. Trégouët, et al. (2021). Genome-wide analysis identifies novel susceptibility loci for myocardial infarction. *European Heart Journal 42*(9), 919–933.

Heckman, J. and R. Robb (1985). Using longitudinal data to estimate age, period and cohort effects in earnings equations. In *Cohort analysis in social research*, pp. 137–150. Springer.

Heckman, J. J. (2007, aug). The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences 104*(33), 13250–5.

Houmark, M. A., V. Ronda, and M. Rosholm (2020). The nurture of nature and the nature of nurture: How genes and investments interact in the formation of skills. Technical report, IZA Discussion Papers.

Hubert, H. B., M. Feinleib, P. M. McNamara, and W. P. Castelli (1983). Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the framingham heart study. *Circulation 67*(5), 968–977.

Jaeger, D. A., T. Dohmen, A. Falk, D. Huffman, U. Sunde, and H. Bonin (2010). Direct evidence on risk attitudes and migration. *The Review of Economics and Statistics 92*(3), 684–689.

Janssens, A. C. J. W., Y. S. Aulchenko, S. Elefante, G. J. J. M. Borsboom, E. W. Steyerberg, and C. M. van Duijn (2006, jul). Predictive testing for complex diseases using multiple genes: Fact or fiction? *Genetics in Medicine 8*(7), 395–400.

Kahneman, D. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica 47*, 278.

Karlsson Linnér, R., P. Biroli, E. Kong, S. F. W. Meddens, R. Wedow, M. A. Fontana, M. Lebreton, S. P. Tino, A. Abdellaoui, A. R. Hammerschlag, M. G. Nivard, A. Okbay, C. A. Rietveld, P. N. Timshel, M. Trzaskowski, R. de Vlaming, C. L. Zünd, Y. Bao, L. Buzdugan, A. H. Caplin, C.-Y. Chen, P. Eibich, P. Fontanillas, J. R. Gonzalez, P. K. Joshi, V. Karhunen, A. Kleinman, R. Z. Levin, C. M. Lill, G. A. Meddens,

G. Muntané, S. Sanchez-Roige, F. J. van Rooij, E. Taskesen, Y. Wu, F. Zhang, A. Auton, J. D. Boardman, D. W. Clark, A. Conlin, C. C. Dolan, U. Fischbacher, P. J. F. Groenen, K. M. Harris, G. Hasler, A. Hofman, M. A. Ikram, S. Jain, R. Karlsson, R. C. Kessler, M. Kooyman, J. MacKillop, M. Männikkö, C. Morcillo-Suarez, M. B. McQueen, K. M. Schmidt, M. C. Smart, M. Sutter, A. R. Thurik, A. G. Uitterlinden, J. White, H. de Wit, J. Yang, L. Bertram, D. I. Boomsma, T. Esko, E. Fehr, D. A. Hinds, M. Johannesson, M. Kumari, D. Laibson, P. K. E. Magnusson, M. N. Meyer, A. Navarro, A. A. Palmer, T. H. Pers, D. Posthuma, D. Schunk, M. B. Stein, R. Svento, H. Tiemeier, P. R. H. J. Timmers, P. Turley, R. J. Ursano, G. G. Wagner, J. F. Wilson, J. Gratten, J. J. Lee, D. Cesarini, D. J. Benjamin, P. D. Koellinger, and J. P. Beauchamp (2019, feb). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics 51*(2), 245–257.

Kimball, M. S., C. R. Sahm, and M. D. Shapiro (2008). Imputing risk tolerance from survey responses. *Journal of the American statistical Association 103*(483), 1028–1038.

Kong, A., G. Thorleifsson, M. L. Frigge, B. J. Vilhjalmsson, A. I. Young, T. E. Thorgeirsson, S. Benonisdottir, A. Oddsson, B. V. Halldorsson, G. Masson, et al. (2018). The nature of nurture: Effects of parental genotypes. *Science 359*(6374), 424–428.

Lebergott, S. (1948). Labor force, employment and unemployment, 1929-1939: Estimating methods. *Monthly Labor Review 67*, 50–53.

Lee, J. J., R. Wedow, A. Okbay, E. Kong, O. Maghzian, M. Zacher, T. A. Nguyen-Viet, P. Bowers, J. Sidorenko, R. K. Linnér, et al. (2018). Gene discovery and polygenic prediction from a 1.1-million-person gwas of educational attainment. *Nature genetics 50*(8), 1112.

Levin, R. and D. Vidart (2020). Risk-taking adaptation to macroeconomic experiences: Theory and evidence from developing countries. *Available at SSRN 3748186*.

Linnér, R. K., P. Biroli, E. Kong, S. F. W. Meddens, R. Wedow, M. A. Fontana, M. Lebreton, A. Abdellaoui, A. R. Hammerschlag, M. G. Nivard, A. Okbay, C. A. Rietveld, P. N. Timshel, S. P. Tino, M. Trzaskowski, R. de Vlaming, C. L. Zünd, Y. Bao,

L. Buzdugan, A. H. Caplin, C.-Y. Chen, P. Eibich, P. Fontanillas, J. R. Gonzalez, P. K. Joshi, V. Karhunen, A. Kleinman, R. Z. Levin, C. M. Lill, G. A. Meddens, G. Muntané, S. Sanchez-Roige, F. J. van Rooij, E. Taskesen, Y. Wu, F. Zhang, a. R. Team, E. Consortium, I. C. Consortium, P. G. Consortium, S. S. G. A. Consortium, A. Auton, J. D. Boardman, D. W. Clark, A. Conlin, C. C. Dolan, U. Fischbacher, P. J. Groenen, K. M. Harris, G. Hasler, A. Hofman, M. A. Ikram, S. Jain, R. Karlsson, R. C. Kessler, M. Kooyman, J. MacKillop, M. Männikkö, C. Morcillo-Suarez, M. B. McQueen, K. M. Schmidt, M. C. Smart, M. Sutter, A. R. Thurik, A. G. Uitterlinden, J. White, H. de Wit, J. Yang, L. Bertram, D. Boomsma, T. Esko, E. Fehr, D. A. Hinds, M. Johannesson, M. Kumari, D. Laibson, P. K. Magnusson, M. N. Meyer, A. Navarro, A. A. Palmer, T. H. Pers, D. Posthuma, D. Schunk, M. B. Stein, R. Svento, H. Tiemeier, P. R. Timmers, P. Turley, R. J. Ursano, G. G. Wagner, J. F. Wilson, J. Gratten, J. J. Lee, D. Cesarini, D. J. Benjamin, P. Koellinger, and J. P. Beauchamp (2018, feb). Genome-wide study identifies 611 loci associated with risk tolerance and risky behaviors. *bioRxiv*, 261081.

Liu, H. and G. Guo (2015, aug). Lifetime Socioeconomic Status, Historical Context, and Genetic Inheritance in Shaping Body Mass in Middle and Late Adulthood. *American Sociological Review 80*(4), 705–737.

Locke, A. E., B. Kahali, S. I. Berndt, A. E. Justice, T. H. Pers, F. R. Day, C. Powell, S. Vedantam, M. L. Buchkovich, J. Yang, D. C. Croteau-Chonka, T. Esko, T. Fall, T. Ferreira, S. Gustafsson, Z. Kutalik, J. Luan, R. Mägi, J. C. Randall, T. W. Winkler, A. R. Wood, T. Workalemahu, J. D. Faul, J. A. Smith, J. Hua Zhao, W. Zhao, J. Chen, R. S. N. Fehrmann, Å. K. Hedman, J. M. Karjalainen, E. M. Schmidt, D. Absher, N. Amin, D. Anderson, M. Beekman, J. L. Bolton, J. L. Bragg-Gresham, S. Buyske, A. Demirkan, G. Deng, G. B. Ehret, B. Feenstra, M. F. Feitosa, K. Fischer, A. Goel, J. Gong, A. U. Jackson, S. Kanoni, M. E. Kleber, K. Kristiansson, U. Lim, V. Lotay, M. Mangino, I. Mateo Leach, C. Medina-Gomez, S. E. Medland, M. A. Nalls, C. D. Palmer, D. Pasko, S. Pechlivanis, M. J. Peters, I. Prokopenko, D. Shungin, A. Stančáková, R. J. Strawbridge, Y. Ju Sung, T. Tanaka, A. Teumer, S. Trompet, S. W. van der Laan, J. van Setten, J. V. Van Vliet-Ostaptchouk, Z. Wang, L. Yengo, W. Zhang, A. Isaacs, E. Albrecht, J. Ärnlöv, G. M. Arscott, A. P. Attwood,

S. Bandinelli, A. Barrett, I. N. Bas, C. Bellis, A. J. Bennett, C. Berne, R. Blagieva, M. Blüher, S. Böhringer, L. L. Bonnycastle, Y. Böttcher, H. A. Boyd, M. Bruinenberg, I. H. Caspersen, Y.-D. Ida Chen, R. Clarke, E. Warwick Daw, A. J. M. de Craen, G. Delgado, M. Dimitriou, A. S. F. Doney, N. Eklund, K. Estrada, E. Eury, L. Folkersen, R. M. Fraser, M. E. Garcia, F. Geller, V. Giedraitis, B. Gigante, A. S. Go, A. Golay, A. H. Goodall, S. D. Gordon, M. Gorski, H.-J. Grabe, H. Grallert, T. B. Grammer, J. Gräßler, H. Grönberg, C. J. Groves, G. Gusto, J. Haessler, P. Hall, T. Haller, G. Hallmans, C. A. Hartman, M. Hassinen, C. Hayward, N. L. Heard-Costa, Q. Helmer, C. Hengstenberg, O. Holmen, J. J. Hottenga, A. L. James, J. M. Jeff, Å. Johansson, J. Jolley, T. Juliusdottir, L. Kinnunen, W. Koenig, M. Koskenvuo, W. Kratzer, J. H. Laitinen, C. Lamina, K. Leander, N. R. Lee, P. Lichtner, L. Lind, J. Lindström, K. Sin Lo, S. Lobbens, R. Lorbeer, Y. Lu, F. Mach, P. K. E. Magnusson, A. Mahajan, W. L. McArdle, S. McLachlan, C. Menni, S. Merger, E. Mihailov, L. Milani, A. Moayyeri, K. L. Monda, M. A. Morken, A. Mulas, G. Müller, M. Müller-Nurasyid, A. W. Musk, R. Nagaraja, M. M. Nöthen, I. M. Nolte, S. Pilz, N. W. Rayner, F. Renstrom, R. Rettig, J. S. Ried, S. Ripke, N. R. Robertson, L. M. Rose, S. Sanna, H. Scharnagl, S. Scholtens, F. R. Schumacher, W. R. Scott, T. Seufferlein, J. Shi, A. Vernon Smith, J. Smolonska, A. V. Stanton, V. Steinthorsdottir, K. Stirrups, H. M. Stringham, J. Sundström, M. A. Swertz, A. J. Swift, A.-C. Syvänen, S.-T. Tan, B. O. Tayo, B. Thorand, G. Thorleifsson, J. P. Tyrer, H.-W. Uh, L. Vandenput, F. C. Verhulst, S. H. Vermeulen, N. Verweij, J. M. Vonk, L. L. Waite, H. R. Warren, D. Waterworth, M. N. Weedon, L. R. Wilkens, C. Willenborg, T. Wilsgaard, M. K. Wojczynski, A. Wong, A. F. Wright, Q. Zhang, E. P. Brennan, M. Choi, Z. Dastani, A. W. Drong, P. Eriksson, A. Franco-Cereceda, J. R. Gådin, A. G. Gharavi, M. E. Goddard, R. E. Handsaker, J. Huang, F. Karpe, S. Kathiresan, S. Keildson, K. Kiryluk, M. Kubo, J.-Y. Lee, L. Liang, R. P. Lifton, B. Ma, S. A. McCarroll, A. J. McKnight, J. L. Min, M. F. Moffatt, G. W. Montgomery, J. M. Murabito, G. Nicholson, D. R. Nyholt, Y. Okada, J. R. B. Perry, R. Dorajoo, E. Reinmaa, R. M. Salem, N. Sandholm, R. A. Scott, L. Stolk, A. Takahashi, T. Tanaka, F. M. Van't Hooft, A. A. E. Vinkhuyzen, H.-J. Westra, W. Zheng, K. T. Zondervan, A. C. Heath, D. Arveiler, S. J. L. Bakker, J. Beilby, R. N. Bergman, J. Blangero, P. Bovet, H. Campbell, M. J. Caulfield, G. Cesana, A. Chakravarti, D. I. Chasman, P. S. Chines, F. S. Collins, D. C. Crawford,

L. Adrienne Cupples, D. Cusi, J. Danesh, U. de Faire, H. M. den Ruijter, A. F. Dominiczak, R. Erbel, J. Erdmann, J. G. Eriksson, M. Farrall, S. B. Felix, E. Ferrannini, J. Ferrières, I. Ford, N. G. Forouhi, T. Forrester, O. H. Franco, R. T. Gansevoort, P. V. Gejman, C. Gieger, O. Gottesman, V. Gudnason, U. Gyllensten, A. S. Hall, T. B. Harris, A. T. Hattersley, A. A. Hicks, L. A. Hindorff, A. D. Hingorani, A. Hofman, G. Homuth, G. Kees Hovingh, S. E. Humphries, S. C. Hunt, E. Hyppönen, T. Illig, K. B. Jacobs, M.-R. Jarvelin, K.-H. Jöckel, B. Johansen, P. Jousilahti, J. Wouter Jukema, A. M. Jula, J. Kaprio, J. J. P. Kastelein, S. M. Keinanen-Kiukaanniemi, L. A. Kiemeney, P. Knekt, J. S. Kooner, C. Kooperberg, P. Kovacs, A. T. Kraja, M. Kumari, J. Kuusisto, T. A. Lakka, C. Langenberg, L. Le Marchand, T. Lehtimäki, V. Lyssenko, S. Männistö, A. Marette, T. C. Matise, C. A. McKenzie, B. McKnight, F. L. Moll, A. D. Morris, A. P. Morris, J. C. Murray, M. Nelis, C. Ohlsson, A. J. Oldehinkel, K. K. Ong, P. A. Madden, G. Pasterkamp, J. F. Peden, A. Peters, D. S. Postma, P. P. Pramstaller, J. F. Price, L. Qi, O. T. Raitakari, T. Rankinen, D. C. Rao, T. K. Rice, P. M. Ridker, J. D. Rioux, M. D. Ritchie, I. Rudan, V. Salomaa, N. J. Samani, J. Saramies, M. A. Sarzynski, H. Schunkert, P. E. H. Schwarz, P. Sever, A. R. Shuldiner, J. Sinisalo, R. P. Stolk, K. Strauch, A. Tönjes, D.-A. Trégouët, A. Tremblay, E. Tremoli, J. Virtamo, M.-C. Vohl, U. Völker, G. Waeber, G. Willemsen, J. C. M. Witteman, M. Carola Zillikens, L. S. Adair, P. Amouyel, F. W. Asselbergs, T. L. Assimes, M. Bochud, B. O. Boehm, E. Boerwinkle, S. R. Bornstein, E. P. Bottinger, C. Bouchard, S. Cauchi, J. C. Chambers, S. J. Chanock, R. S. Cooper, P. I. W. de Bakker, G. Dedoussis, L. Ferrucci, P. W. Franks, P. Froguel, L. C. Groop, C. A. Haiman, A. Hamsten, J. Hui, D. J. Hunter, K. Hveem, R. C. Kaplan, M. Kivimäki, D. Kuh, M. Laakso, Y. Liu, N. G. Martin, W. März, M. Melbye, A. Metspalu, S. Moebus, P. B. Munroe, I. Njølstad, B. A. Oostra, C. N. A. Palmer, N. L. Pedersen, M. Perola, L. Pérusse, U. Peters, C. Power, T. Quertermous, R. Rauramaa, F. Rivadeneira, T. E. Saaristo, D. Saleheen, N. Sattar, E. E. Schadt, D. Schlessinger, P. Eline Slagboom, H. Snieder, T. D. Spector, U. Thorsteinsdottir, M. Stumvoll, J. Tuomilehto, A. G. Uitterlinden, M. Uusitupa, P. van der Harst, M. Walker, H. Wallaschofski, N. J. Wareham, H. Watkins, D. R. Weir, H.-E. Wichmann, J. F. Wilson, P. Zanen, I. B. Borecki, P. Deloukas, C. S. Fox, I. M. Heid, J. R. O'Connell, D. P. Strachan, K. Stefansson, C. M. van Duijn, G. R. Abecasis, L. Franke, T. M. Frayling, M. I. McCarthy, P. M. Visscher, A. Scherag,

C. J. Willer, M. Boehnke, K. L. Mohlke, C. M. Lindgren, J. S. Beckmann, I. Barroso, K. E. North, E. Ingelsson, J. N. Hirschhorn, R. J. F. Loos, and E. K. Speliotes (2015, feb). Genetic studies of body mass index yield new insights for obesity biology. *Nature 518*(7538), 197–206.

Loewenstein, G. and E. Angner (2003). Predicting and indulging changing preferences. *Time and decision: Economic and psychological perspectives on intertemporal choice*, 351–391.

MacKinnon, J. G. (2006). Bootstrap methods in econometrics. *Economic Record 82*, S2–S18.

Malmendier, U. and S. Nagel (2011). Depression babies: do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics 126*(1), 373–416.

Malmendier, U., G. Tate, and J. Yan (2011). Overconfidence and early-life experiences: the effect of managerial traits on corporate financial policies. *The Journal of finance 66*(5), 1687–1733.

Manski, C. F. (2011). Genes, Eyeglasses, and Social Policy. *Journal of Economic Perspectives*.

Meyer, B. D. and N. Mittag (2017). Misclassification in binary choice models. *Journal of Econometrics 200*(2), 295–311.

Meyer, B. D., N. Mittag, and R. M. George (2020). Errors in survey reporting and imputation and their effects on estimates of food stamp program participation. *Journal of Human Resources*, 0818–9704R2.

Miao, J., G. Song, Y. Wu, J. Hu, Y. Wu, S. Basu, J. S. Andrews, K. Schaumberg, J. M. Fletcher, L. L. Schmitz, et al. (2022). Reimagining gene-environment interaction analysis for human complex traits. *bioRxiv*, 2022–12.

Michailidou, K., S. Lindström, J. Dennis, J. Beesley, S. Hui, S. Kar, A. Lemaçon, P. Soucy, D. Glubb, A. Rostamianfar, et al. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature 551*(7678), 92–94.

Mills, M. C., N. Barban, and F. C. Tropf (2020). *An Introduction to Statistical Genetic Data Analysis*. MIT Press.

Mittag, N. (2018). A nonparametric k-sample test of conditional independence.

Mostafavi, H., A. Harpak, I. Agarwal, D. Conley, J. K. Pritchard, and M. Przeworski (2020). Variable prediction accuracy of polygenic scores within an ancestry group. *Elife 9*.

Okbay, A. and C. A. Rietveld (2015, aug). On improving the credibility of candidate gene studies: A review of candidate gene studies published in Emotion. *Emotion 15*(4), 531–537.

Page, L., D. A. Savage, and B. Torgler (2014). Variation in risk seeking behaviour following large losses: A natural experiment. *European Economic Review 71*, 121–131.

Plomin, R. (1990). *Nature and nurture: An introduction to human behavioral genetics.* Wadsworth Publishing Company.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich (2006, aug). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics 38*(8), 904–9.

Purcell, S. M., N. R. Wray, J. L. Stone, P. M. Visscher, M. C. O'Donovan, P. F. Sullivan, P. Sklar, D. M. Ruderfer, A. McQuillin, D. W. Morris, C. T. O'Dushlaine, A. Corvin, P. A. Holmans, M. C. O'Donovan, S. Macgregor, and G (2009, jul). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature 460*(7256), 748–752.

Ridley, M. (2003). *Nature Via Nurture: Genes, Experience, and What Makes Us Human.*

Rosales-Rueda, M. (2018). The impact of early life shocks on human capital formation: Evidence from el niño floods in ecuador. *Journal of health economics 62*, 13–44.

Rosales-Rueda, M. F. (2014). Family investment responses to childhood health conditions: Intrafamily allocation of resources. *Journal of health economics 37*, 41–57.

Rutter, M. (2006). *Genes and Behavior: Nature-Nurture Interplay Explained.* Oxford, UK: Blackwell Publishers.

Sahm, C. R. (2012). How much does risk tolerance change? *The quarterly journal of finance 2*(04), 1250020.

Sanz-de Galdeano, A. and A. Terskaya (2019). Sibling differences in educational polygenic scores: How do parents react? Technical report.

Schmitz, L. and D. Conley (2016a). The long-term consequences of vietnam-era conscription and genotype on smoking behavior and health. *Behavior genetics 46*(1), 43–58.

Schmitz, L. and D. C. Conley (2016b, jun). The Impact of Late-Career Job Loss and Genotype on Body Mass Index. *NBER working paper 1*(22348).

Schmitz, L. L. and D. C. Conley (2016c, jun). The Impact of Late-Career Job Loss and Genotype on Body Mass Index. *NBER Working Paper* (22348).

Schmitz, L. L. and D. C. Conley (2016d, jan). The Long-Term Consequences of Vietnam-Era Conscription and Genotype on Smoking Behavior and Health. *Behavior Genetics 46*(1), 43–58.

Schmitz, L. L. and D. C. Conley (2017a). Modeling Gene-Environment Interactions With Quasi-Natural Experiments. *Journal of Personality 85*(1), 10–21.

Schmitz, L. L. and D. C. Conley (2017b, dec). The effect of Vietnam-era conscription and genetic potential for educational attainment on schooling outcomes. *Economics of Education Review 61*, 85–97.

Schutz, A. (1962). Common-sense and scientific interpretation of human action. In *Collected papers I*, pp. 3–47. Springer.

Solon, G., S. J. Haider, and J. M. Wooldridge (2015). What are we weighting for? *Journal of Human resources 50*(2), 301–316.

Sonnega, A. and D. R. Weir (2014, oct). The Health and Retirement Study: A Public Data Resource for Research on Aging. *Open Health Data 2*(1).

Sudlow, C., J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, B. Liu, P. Matthews, G. Ong, J. Pell, A. Silman, A. Young, T. Sprosen, T. Peakman, R. Collins, D. Clayton, P. McKeigue, W. Willett, W. Blot, G. Colditz, A. Folsom, B. Henderson, M. Stampfer, F. Collins, T. Manolio, R. Doll, R. Peto, W. Ollier, T. Peakman, P. Burton, A. Hansell, I. Fortier, M. Khoury, A. Hattersley, M. McCarthy, R. Collins, P. Elliott, T. Peakman, B. Liu, H. Young, F. Crowe, V. Benson, E. Spencer, R. Clarke, M. Shipley, S. Lewington, L. Youngman, J. Pell, J. Valentine, H. Inskip, P. Downey, and V. Barbour (2015, mar). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine 12*(3), e1001779.

The Tobacco and Genetics Consortium, H. Furberg, Y. Kim, J. Dackor, E. Boerwinkle, N. Franceschini, D. Ardissino, L. Bernardinelli, P. M. Mannucci, F. Mauri, P. A. Merlini, D. Absher, T. L. Assimes, S. P. Fortmann, C. Iribarren, J. W. Knowles, T. Quertermous, L. Ferrucci, T. Tanaka, J. C. Bis, C. D. Furberg, T. Haritunians, B. McKnight, B. M. Psaty, K. D. Taylor, E. L. Thacker, P. Almgren, L. C. Groop, C. Ladenvall, M. Boehnke, A. U. Jackson, K. L. Mohlke, H. M. Stringham, J. Tuomilehto, E. J. Benjamin, S.-J. Hwang, D. Levy, S. R. Preis, R. S. Vasan, J. Duan, P. V. Gejman, D. F. Levinson, A. R. Sanders, J. Shi, E. H. Lips, J. D. McKay, A. Agudo, L. Barzan, V. Bencko, S. Benhamou, X. Castellsagué, C. Canova, D. I. Conway, E. Fabianova, L. Foretova, V. Janout, C. M. Healy, I. Holcátová, K. Kjaerheim, P. Lagiou, J. Lissowska, R. Lowry, T. V. Macfarlane, D. Mates, L. Richiardi, P. Rudnai, N. Szeszenia-Dabrowska, D. Zaridze, A. Znaor, M. Lathrop, P. Brennan, S. Bandinelli, T. M. Frayling, J. M. Guralnik, Y. Milaneschi, J. R. B. Perry, D. Altshuler, R. Elosua, S. Kathiresan, G. Lucas, O. Melander, C. J. O'Donnell, V. Salomaa, S. M. Schwartz, B. F. Voight, B. W. Penninx, J. H. Smit, N. Vogelzangs, D. I. Boomsma, E. J. C. de Geus, J. M. Vink, G. Willemsen, S. J. Chanock, F. Gu, S. E. Hankinson, D. J. Hunter, A. Hofman, H. Tiemeier, A. G. Uitterlinden, C. M. van Duijn, S. Walter, D. I. Chasman, B. M. Everett, G. Paré, P. M. Ridker, M. D. Li, H. H. Maes, J. Audrain-McGovern, D. Posthuma, L. M. Thornton, C. Lerman, J. Kaprio, J. E.

Rose, J. P. A. Ioannidis, P. Kraft, D.-Y. Lin, and P. F. Sullivan (2010, may). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genetics 42*(5), 441–447.

Tishkoff, S. A., F. A. Reed, F. R. Friedlaender, C. Ehret, A. Ranciaro, A. Froment, J. B. Hirbo, A. A. Awomoyi, J.-M. Bodo, O. Doumbo, M. Ibrahim, A. T. Juma, M. J. Kotze, G. Lema, J. H. Moore, H. Mortensen, T. B. Nyambo, S. A. Omar, K. Powell, G. S. Pretorius, M. W. Smith, M. A. Thera, C. Wambebe, J. L. Weber, and S. M. Williams (2009, may). The genetic structure and history of Africans and African Americans. *Science 324*(5930), 1035–44.

Trejo, S. and B. W. Domingue (2019). Genetic nature or genetic nurture? quantifying bias in analyses using polygenic scores. *BioRxiv*, 524850.

Tropf, F. C., S. H. Lee, R. M. Verweij, G. Stulp, P. J. Van Der Most, R. De Vlaming, A. Bakshi, D. A. Briley, C. Rahal, R. Hellpap, et al. (2017). Hidden heritability due to heterogeneity across seven populations. *Nature human behaviour 1*(10), 757–765.

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science 9*(5), 160–164.

Turkheimer, E., A. Haley, M. Waldron, B. M. D'Onofrio, and I. I. Gottesman (2003, nov). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science 14*(6), 623–628.

Vilhjálmsson, B. J., J. Yang, H. K. Finucane, A. Gusev, S. Lindström, S. Ripke, G. Genovese, P.-R. Loh, G. Bhatia, R. Do, T. Hayeck, H.-H. Won, S. Kathiresan, M. Pato, C. Pato, R. Tamimi, E. A. Stahl, N. Zaitlen, B. Pasaniuc, G. Belbin, E. E. Kenny, M. H. Schierup, P. L. De Jager, N. A. Patsopoulos, S. McCarroll, M. J. Daly, S. M. Purcell, D. Chasman, B. Neale, M. Goddard, P. M. Visscher, P. Kraft, N. Patterson, and A. L. Price (2015, oct). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *The American Journal of Human Genetics 97*(4), 576–592.

Visscher, P. M., M. A. Brown, M. I. McCarthy, and J. Yang (2012, jan). Five years of GWAS discovery. *American journal of human genetics 90*(1), 7–24.

Visscher, P. M., N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, M. A. Brown, and J. Yang (2017). 10 years of gwas discovery: biology, function, and translation. *The American Journal of Human Genetics 101*(1), 5–22.

Wasserman, L. and K. Roeder (2009). High-dimensional variable selection. *The Annals of Statistics 37*(5A), 2178–2201.

Wedow, R., M. Zacher, B. M. Huibregtse, K. Mullan Harris, B. W. Domingue, and J. D. Boardman (2018). Education, Smoking, and Cohort Change: Forwarding a Multidimensional Theory of the Environmental Moderation of Genetic Effects. *American Sociological Review 83*(4), 802–832.

Yang, J., T. Ferreira, A. P. Morris, S. E. Medland, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. N. Weedon, R. J. F. Loos, T. M. Frayling, M. I. McCarthy, J. N. Hirschhorn, M. E. Goddard, and P. M. Visscher (2012, apr). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature genetics 44*(4), 369–75, S1–3.

Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher (2011, jan). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics 88*(1), 76–82.

Yengo, L., J. Sidorenko, K. E. Kemper, Z. Zheng, A. R. Wood, M. N. Weedon, T. M. Frayling, J. Hirschhorn, J. Yang, P. M. Visscher, et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in 700000 individuals of european ancestry. *Human molecular genetics 27*(20), 3641–3649.

Young, A. I., S. Benonisdottir, M. Przeworski, and A. Kong (2019). Deconstructing the sources of genotype-phenotype associations in humans. *Science 365*(6460), 1396–1400.

Zhao, B. and H. Zhu (2021). On genetic correlation estimation with summary statistics from genome-wide association studies. *Journal of the American Statistical Association*, 1–11.