DISCUSSION PAPER SERIES

DP16603

Gene-Environment Effects on Female Fertility

Nicola Barban, Elisabetta De Cao and Marco Francesconi

LABOUR ECONOMICS

CEPR

ISSN 0265-8003

Gene-Environment Effects on Female Fertility

Nicola Barban, Elisabetta De Cao and Marco Francesconi

Discussion Paper DP16603
Published 04 October 2021
Submitted 01 October 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Gene-Environment Effects on Female Fertility

Abstract

Fertility has a strong biological component generally ignored by economists. Using the UK Biobank, we analyze the extent to which genes, proxied by polygenic scores, and the environment, proxied by early exposure to the contraceptive pill diffusion, affect age at first sexual intercourse, age at first birth, completed family size, and childlessness. Both genes and environment exert substantial influences on all outcomes. The anticipation of sexual debut and the postponement of motherhood led by the diffusion of the pill are magnified by gene-environment interactions, while the decline in family size and the rise in childlessness associated with female emancipation are attenuated by gene-environment effects. The nature-nurture interplay becomes stronger in more egalitarian environments that empower women, allowing genes to express themselves more fully. These conclusions are confirmed by heterogenous effects across the distributions of genetic susceptibilities and exposure to environmental risks, sister fixed effects models, mother-daughter comparisons, and counterfactual simulations.

JEL Classification: D10, I14, I15, J01, J13, J16

Keywords: Fertility, Genetics, Polygenic Score, Contraceptive pill, Nature versus nurture, Social norms

Nicola Barban - n.barban@unibo.it
University of Bologna

Elisabetta De Cao - e.de-cao@lse.ac.uk
London School of Economics

Marco Francesconi - mfranc@essex.ac.uk
University of Essex and CEPR

Gene-Environment Effects on Female Fertility*

Nicola Barban[†a], Elisabetta De Cao[‡b], and Marco Francesconi[§c]

[a]University of Bologna
[b]London School of Economics and Political Science
[c]University of Essex

September 28, 2021

Abstract

Fertility has a strong biological component generally ignored by economists. Using the UK Biobank, we analyze the extent to which genes, proxied by polygenic scores, and the environment, proxied by early exposure to the contraceptive pill diffusion, affect age at first sexual intercourse, age at first birth, completed family size, and childlessness. Both genes and environment exert substantial influences on all outcomes. The anticipation of sexual debut and the postponement of motherhood led by the diffusion of the pill are magnified by gene-environment interactions, while the decline in family size and the rise in childlessness associated with female emancipation are attenuated by gene-environment effects. The nature-nurture interplay becomes stronger in more egalitarian environments that empower women, allowing genes to express themselves more fully. These conclusions are confirmed by heterogenous effects across the distributions of genetic susceptibilities and exposure to environmental risks, sister fixed effects models, mother-daughter comparisons, and counterfactual simulations.

JEL Classification: D10, I14, I15, J01, J13, J16

Keywords: Fertility, Genetics, Polygenic Score, Contraceptive Pill, Nature versus Nurture, Social Norms

*We thank Silvia Barcellos, Jonathan Beauchamp, Sonia Bhalotra, Prashant Bharadwaj, Francesco Billari, Pietro Biroli, Dora Costa, Raquel Fernández, Titus Galama, Piero Gottardi, James Heckman, Petter Lundborg, Andries Marees, Melinda Mills, Franco Peracchi, Erik Plug, Bob Pollak, Ana Rodríguez González, Victor Ronda, Kjell Salvanes, Uta Schönberg, Almudena Sevilla, Felix Tropf, Patrick Turley, Gerard van den Berg, Stephanie von Hinke, and participants in seminars and workshops at the Universities of Alicante, Barcelona, Bocconi, Bologna, HSE (Moscow), LSE, Lund, Oxford, Roma I, Royal Holloway, USC, York, and at the 2019 NBER Cohort Studies Meeting, the 2021 RES, SOLE, SEHO, ESPE, EALE, Essen Health Economics conferences, and the IEB Workshop on Public Policies for constructive discussions and useful comments on previous versions of the paper. We are grateful to the British Academy for generous financial support.

[†]n.barban@unibo.it
[‡]e.de-cao@lse.ac.uk
[§]Corresponding Author: mfranc@essex.ac.uk

In bringing economics to bear on procreation and children,
a new dialogue between data and theory has begun.
THEODORE W. SCHULTZ (1973, p. S2)

1. Introduction

Understanding fertility decisions has been at the center of economic research since the 1960s. Fundamental to this scientific endeavor are the pioneering contributions by Becker (1960, 1964, 1965, 1973, 1981/1991), Schultz (1963, 1973), Becker and Lewis (1973), Ben-Porath (1973), Willis (1973), and Heckman and Willis (1976).[1] Such contributions have paved the way to the analysis of some of the most profound transformations observed over the last 100 years linking fertility to female education, life cycle labor supply and wages, improvements in home production technology, child health and well-being, economic growth, and female empowerment (e.g., Barro and Becker, 1989; Becker et al., 1990; Heckman and Walker, 1990; Browning, 1992; Galor and Weil, 2000; Goldin and Katz, 2002; de la Croix and Doepke, 2003; Greenwood et al., 2005; Fernández and Fogli, 2009; Eckstein et al., 2019).[2] As a result, timing of childbirth, completed family size, and teenage childbearing have been the focus of a substantial body of work in economics and other cognate social sciences.[3] Although many of these studies hinted at the relevance of genetic endowments to reproductive fitness, none of them explicitly examined the interplay of genetic markers and socioeconomic factors in shaping female fertility behavior. Ours is the first paper that does this, integrating sociogenomics into the economics of fertility.

Most of what we know that relates human genetics to fertility behavior is based on noneconomic heritability studies, which typically use twins data to decompose cross-sectional variation in observed outcomes into unobservable genetic and environmental components (e.g., Kohler et al., 1999; Tropf et al., 2015). These additive nature-and-nurture models, however, do not tell us much about how genes are associated with fertility behavior, other than supplying an estimate of the proportion of variation in a phenotype (either a morphological and physiological trait, or the outcome of a decision) that is attributable to genetic differences. Moreover, by suppressing the role played by genes, heritability studies cannot go beyond the nature versus nurture distinction, which is known to oversimplify, and mischaracterize our understanding of, behavior (Heckman, 2007; Kong et al., 2018; Houmark et al., 2020).

Bringing behavioral genetics fully into the economics of fertility, therefore, is vital for two key reasons, which underpin our main contributions. First, it offers us the missing link, so far absent in the social science literatures, between fertility phenotypes and the environment in which such phenotypes are observed. Leveraging unprecedented advances in molecular genetics, we measure genotypes (genetic endowments) and model their relationship jointly with socioeconomic influences on behavior. This informs us on the extent to which genes mediate the impact of the environment

[1] Malthus's (1798) theory of population change was highly influential to understand the relationship between fertility and economic growth. But — as noted by Becker (1981, p. 2) — economists hardly noticed the family and fertility issues prior to the 1950s.

[2] A comprehensive coverage of all these strands of the literature is beyond the scope of this paper.

[3] See, among others, the important reviews by Montgomery and Trussell (1986), Hotz et al. (1997), Balbo et al. (2013), and Esping-Andersen and Billari (2015).

on fertility decisions or, equivalently, how the environment modifies genetic influences. Our measure of genetic endowment is given by specific genetic variants, which predict fertility phenotypes and are summarized by a linear index known as polygenic score (PGS). PGSs are obtained from genome-wide association studies using the UK Biobank, which is one of the largest publicly available genomic resources in the world. To provide a comprehensive picture of women's decisions over their reproductive life cycle, for the first time, we analyze four outcomes (i.e., age at first sexual intercourse, age at first birth, completed family size, and childlessness), for each one of which we construct different polygenic scores.[4]

Second, the integration of sociogenomics into the economics of fertility enables us to break new grounds in understanding how fertility outcomes are shaped by gene-environment interactions, which represent one natural mechanism that goes beyond the nature-nurture dichotomy. There are currently a few different hypotheses to understand this interplay, most of which have been rarely tested and never applied to fertility.[5] A standard problem faced by this line of research is an inconsistent measurement of environmental exposure which is likely to pick up nonexogenous shocks at the individual level, such as parental socioeconomic status, parenting quality, or stressful life events. To minimize this problem, we move away from individually based environmental influences and focus instead on measures at a more aggregate geographic level that are likely to be relevant to female fertility. One of such measures is given by the diffusion of the birth control pill in the local area where women grew up. By lowering the costs of engaging in long-term career investments, access to the pill has been shown to be an important trigger of gender-egalitarian norms (Goldin and Katz, 2002).[6] With women born between the late 1930s and the late 1960s, the UK Biobank offers a sample with abundant cross-cohort variation in such norms.

Our perspective on the role of gene-environment interactions is thus in the spirit of a social control mechanism, according to which genetic influences are mediated by structural constraints, societal contexts, norms, and — more broadly — culture (Shanahan and Hofer, 2005; Guiso et al., 2006; Bisin and Verdier, 2011; Boelmann et al., 2021). Social norms, such as those shaped by the pill revolution, are expected to channel genetic influences into fertility behavior, either magnifying or mitigating their impacts on outcomes.[7] Evolving more rapidly than genes, culture can create novel environments that expose genes to new selective pressures and adaptations (Richerson et al., 2010),

[4]For the sake of space concerns, we do not analyze teenage fertility. Preliminary work suggests that essentially all our benchmark findings carry through this additional outcome. This analysis and its results are left for another paper.

[5]For up-to-date summaries, see Conley and Fletcher (2017) and Mills and Tropf (2020).

[6]Female emancipation may have been facilitated by forces other than the pill, including the legalization of abortion (Myers, 2017) and the rise of feminism (Goldin and Katz, 2002). Demand-side factors, such as equal pay and sex discrimination legislations, may have also played a crucial role in promoting mothers' labor force participation and empowerment (Coles and Francesconi, 2019). Our focus on the contraceptive pill is not meant to favor one factor over another, and it is in part constrained by data availability. Whenever possible, we check for robustness using other proxies.

[7]Two alternative approaches of looking at gene-environment interactions have been proposed. One is given by the contextual triggering (or diathesis-stress) model, which hinges on the idea that a predisposition for a trait lies dormant until triggered by an environmental stressor, e.g., the death of a spouse (Caspi et al., 2002; Domingue et al. 2017). The other is offered by the differential susceptibility model, according to which plasticity is highly heterogeneous, with some individuals being genetically more susceptible to both positive and negative environments and others remaining resilient across all situations (Belsky and Pluess, 2009).

which may enhance or inhibit female reproductive fitness. The association of genetic influences on fertility outcomes can be viewed as an index of selective success of genetic predispositions, such as the innate capabilities to produce and care for offspring, sexual desire, fecundity, and attractiveness. To the extent that social norms tolerant of (or encouraging) pill usage lead to an increase in successful selection of women having sex earlier and postponing births, the influence exerted by genes is likely to be higher in areas and times with greater pill diffusion. An environment that fosters the utilization of the contraceptive pill will then enable women's genetic potential to unfold.[8]

Using data from the UK Biobank, we show that each of the outcome-specific polygenic scores plays a substantial and statistically significant role in explaining the four fertility outcomes under study, contributing to 20–40% of the observed total variance across traits. Women with larger genetic endowments have higher ages at first sex and first birth, have more children, and face lower chances of being childless by age 45 or more. Our measure of the external environmental risk is as important in affecting fertility phenotypes as the polygenic scores are. A greater exposure to the contraceptive pill is significantly associated with an anticipation of sexual debut, a postponement of motherhood, a smaller family size, and a greater chance of childlessness.

Importantly, our gene-environment interaction estimates suggest that the environment plays a key role in the manifestation of genetic influences on fertility. Both the anticipation of sexual initiation and the postponement of motherhood led by the diffusion of the pill are reinforced by gene-environment interactions. Conversely, the reduction in family size and the rise in childlessness associated with greater pill usage are attenuated by the way in which the polygenic scores interact with the rise of gender-egalitarian norms. Upholding the social control mechanism with which we interpret the gene-environment interplay, these estimates indicate that genetic influences on fertility are most relevant when social norms allow a broad range of life-course alternatives. Several sensitivity exercises confirm these baseline results.

Focusing on a subsample of sisters and controlling for family fixed effects lead to similar conclusions. Genetic endowments continue to express themselves more powerfully when the environment is conducive to female empowerment, even if both genes and environment (not only external but also internal to the family of origin) shared by sisters are more homogenous than in the general population. Mother-daughter comparisons are performed on a much smaller sample with potential issues of statistical power. Despite this, they reveal evidence that is consistent with the notion of "genetic nurture" (Kong et al., 2018), according to which fertility outcomes are shaped by intra-family environmental conditions as well as parental genotypes that are not genetically transmitted from parents to daughters. Genetic nurturing is thus one possible way in which gene-environment interactions operate within the family, expressing themselves through internal socialization and intrafamily cultural transmission.

Finally, counterfactual simulations provide clear evidence that the impact of the external environment on the evolution of outcomes is substantial in a world populated by women who are genetically identical. Conversely, in a gender egalitarian world where all women face the maximum exposure

[8]For similar insights, albeit in different contexts, see Turkheimer et al. (2003), Rimfeld et al. (2018), and Engzell and Tropf (2019).

to pill usage, we observe only small differences in outcomes across cohorts but large cross-sectional (within-cohort) variation driven by different genetic endowments. Not only do gene-environment interactions have impacts on fertility outcomes that differ by the intensity of women's genetic susceptibilities, but they also express themselves more sharply when the environmental conditions offer women a broad range of opportunities, which complement their genetic predisposition.

Related Literature — Most of the economic literature focuses on either completed family size, age at first birth, or teenage and nonmarital childbearing (Montgomery and Trussell, 1986; Hotz et al., 1997). More recently, childlessness has also drawn attention (e.g., Baudin et al., 2015). The original child quantity-quality theory elaborated by Becker (1960, 1981) and Becker and Lewis (1973) has been applied to interpret the historical fertility transition, which took place in most of Europe and North America between the late eighteenth and early twentieth centuries (Guinnane, 2011), as well as more recent demographic trends (Bailey and Hershbein, 2018). This same framework has been modified to investigate the economic consequences of fertility at the micro level with a wide variety of data and research designs (e.g., Mincer and Polachek, 1974; Weiss and Gronau, 1981; Moffitt, 1984; Wolpin, 1984; Hotz and Miller, 1988; Eckstein and Wolpin, 1989; Heckman and Walker, 1990; Bronars and Grogger, 1995; Francesconi, 2002; Lundborg et al., 2017; Adda et al., 2017). None of these studies, however, embeds sociogenomic considerations in their investigations. Our work contributes to this strand by incorporating genetic measures directly into the analysis.

Another important literature stems from the demographic and sociological research. A distinctive feature of this strand is its emphasis on the role played by the development of individualistic lifestyle orientations, which may explain the fertility decline after the post-World War II baby boom (Hakim, 2000; Lesthaeghe, 2010), and the emergence of gender-egalitarian family norms, which may explain a possible reversal of trends in fertility patterns in more recent years in some advanced economies (Myrskylä et al., 2009; Goldscheider et al., 2015; Esping-Andersen and Billari, 2015). Our work speaks to these contributions by recognizing the importance of social norms in shaping fertility behavior and, in particular, the interplay between genes and environmental influences. We use the diffusion of the contraceptive pill as our main measure of norms concerning women's role in the family and the labor market, and reflecting female empowerment.

A final strand of research relevant to ours is represented by the burgeoning social science literature that uses polygenic scores to identify gene-environment interactions. Most of this literature focuses on education, wealth, and health, and uses polygenic scores for either educational attainment, smoking, or body mass index. Some of the most significant studies include Barcellos et al. (2018, 2021), Trejo et al. (2018), Wedow et al. (2018), Bierut et al. (2018), Barth et al. (2020), Breinholt and Conley (2020), Papageorge and Thom (2020), Houmark et al. (2020), Ronda et al. (2020), Biroli and Zwyssig (2021), and Bolyard and Savelyev (2021). For instance, Barcellos et al. (2018) show that genes moderate the effect of education on health, whereby education-driven improvements in health (e.g., healthy body size) are larger for individuals with high genetic predisposition to obesity. Confirming for Denmark the results found for the US by Papageorge and Thom (2020), Houmark et al. (2020) demonstrate that the educational returns to genetic endowments are attenuated by

childhood disadvantage: children who experience childhood disadvantage are thus unable to realize their full educational potential. Trejo et al. (2018) find that genetic associations with college completion are moderated by school-level socioeconomic status. Barth et al. (2020) document that the PGS for educational attainment is more strongly related to wealth among individuals who have greater autonomy over their financial decisions, whereby those with lower educational genetic endowments may benefit from outsourcing their investment decisions.[9] We contribute to this literature by expanding the analysis to fertility outcomes and looking for the first time at how fertility related genetic endowments interact with cultural norms.

The remainder of the paper is organised as follows. Section 2 introduces basic concepts of molecular genetics and the main genetic quantities used in our study. Section 3 presents the data and descriptive analysis, while Section 4 explains our conceptual framework and empirical specifications. Section 5 discusses the baseline results on the relationship linking polygenic scores, environmental measures, and their interactions to each of the fertility outcomes. It also reports the evidence from several sensitivity checks. Section 6 explores the role played by gene-environment interactions looking at the extent of heterogeneity across the distribution of genetic susceptibilities, and across the distribution of exposure to environmental risks. It also presents and discusses the results from sister fixed effects models and mother-daughter comparisons, and evaluates some stark counterfactual simulations. Section 7 concludes.

2. Genetic Data: Basic Concepts and Limitations

We start by discussing some basic notions of molecular genetics needed to understand the construction of the measures of genetic endowment used in this paper.[10] We then outline some of the advantages and challenges posed by this method and summarize a few critical issues of interpretation.

2.1 The Human Genome

The human genome is composed of approximately 3 billion nucleotide pairs spread out over 23 chromosomes containing the instructions for synthesizing proteins and how to build and regulate all the biological functions in the body. For about 99% of locations in the genome, there is no variation across individuals. At the remaining genetic addresses (less than 1%), the nucleotide pair may differ across individuals. Diversity is the result of molecular differences in the genetic make-up transmitted via meiosis from parents to offspring.

Recent advances in molecular genetics and the big data revolution have made it possible and relatively inexpensive to measure millions of such genetic variants (Conley and Fletcher, 2017). The most common type of variation is known as single nucleotide polymorphism (SNP). SNPs occur normally throughout an individual's DNA and, on average, once in every 300 nucleotides. This

[9] Black et al. (2020) find instead little evidence of gene-environment interactions to explain the intergenerational transmission of wealth. This work, however, uses a sample of adoptees and no genomic data. Using a subsample of adoptees in the UK Biobank and polygenic scores to proxy genetic endowments, Cheesman et al. (2020) find strong gene-environment influences on education. See also Plomin (2014).

[10] Related discussions are presented in Beauchamp et al. (2011), Benjamin et al. (2012), and Barth et al. (2020).

means there are roughly 10 million SNPs in the human genome. Variation at a particular SNP is measured by a count variable indicating how many copies of a particular base pair molecule an individual possesses at that genetic location. Since individuals inherit two copies of each SNP, one from each parent, there are three possible outcomes at each address: that is, an individual can have either zero, one, or two copies of a specific variant.

An extensive body of current scientific research aims to identify which of these variants are associated with behavioral outcomes or specific diseases. Genome-wide association studies (GWAS) provide one tool for estimating such associations. Under the GWAS methodology, researchers scan the entire genome for SNPs that are associated with a particular phenotype (trait or outcome). Most GWAS find that phenotypes are generally regulated by multiple SNPs (Okbay et al., 2016; Barban et al., 2016; Lee et al., 2018; Linnér et al., 2019). To simplify the statistical analysis and improve predictive power, GWAS results are aggregated into polygenic scores, which are linear combinations of the individual SNP count variables, weighted by their GWAS regression coefficients.

Before describing how polygenic scores are constructed, we discuss another parameter, heritability, which has been often estimated by economists typically using samples of twins rather than genetic data (e.g., Taubman, 1976; Behrman and Taubman 1989; Sacerdote, 2007; Fagereng et al., 2021).[11] This should allow us to appreciate the genomic innovation in general, and the use of polygenic scores in particular, more fully.

2.2 Heritability

As posited by most studies in this literature (see Mills et al., [2020] for a review), we assume that a fertility phenotype, Y, is produced by a linear and additive combination of genetic factors, G, and environmental inputs, E, i.e., $Y = G + E$. As it is commonly assumed that G and E are uncorrelated across the population, the population variance of Y, σ_Y^2, can be written as: $\sigma_Y^2 = \sigma_G^2 + \sigma_E^2$. Heritability is then defined as

$$h^2 = \frac{\sigma_G^2}{\sigma_Y^2},$$

that is, h^2 measures the fraction of the variation in the outcome that is attributable to additive genetic differences in the population. Heritability provides a useful summary description of the genetic architecture of a phenotype, especially when different populations (or different subgroups within a given population) are examined. Clearly, h^2 is a population statistic which depends on both genetic and nongenetic factors. Different populations may have different values of h^2, even in presence of identical G, because of changes in the nongenetic components. Until recently, estimation of h^2 has relied on closely related individuals, typically twins or siblings. A concern about such an approach is that the expected relatedness between twins might be strongly correlated to similarity in environmental factors (Goldberger, 1978).

Recent developments in molecular genetics, however, have allowed the estimation of the vari-

[11] See also the discussions by Goldberger (1979) and Manski (2011).

ance explained by all SNPs identified in a GWAS among genetically unrelated individuals, called SNP-based heritability, h^2_{SNP}. Intuitively, this new estimation method boils down to contrasting the phenotypic similarity between unrelated individuals to their SNP-derived genetic similarity. Following Yang et al. (2017), we shall estimate h^2_{SNP} using a mixed linear model, known as genomic-relatedness-matrix restricted maximum likelihood (GREML), on a subsample of the genetic data described in the next section. This is a subsample of unrelated individuals with only independent SNPs (approximately 250,000 of the 800,000 genotyped). A key assumption of this approach is that, among genetically unrelated individuals, environmental factors are uncorrelated with differences in the degree of genetic relatedness. As a result, estimates of h^2_{SNP} are typically smaller than the heritability coefficients found from twins studies (e.g., Benjamin et al., 2012; Yang et al., 2017; Mills et al., 2020). Since there is more random variation in the realized degree of genome sharing relative to the expected degree when genetic relatedness declines (Hill and Weir, 2011), environmental confounding is less likely to affect the estimation of h^2_{SNP}, which is based on realized relatedness among individuals whose expected relatedness is negligible.[12]

2.3 Polygenic Scores

While twins studies (and, more generally, all the existing studies of heritability) provide an estimate of how much genetic factors collectively matter to explain the variation in a given outcome or trait, they do not reveal which specific SNPs are relevant to that phenotype. This is precisely what GWAS provide.

As mentioned earlier, GWAS regression coefficients are usually aggregated into polygenic scores. A polygenic score (PGS) is constructed through a linear combination of several (possibly thousands of) SNPs across the genome, weighting them by the strength of their association with a given phenotype. A polygenic score is thus a single quantitative index that can be interpreted as a measure of an individual's genetic propensity toward the phenotype relative to the population. Formally, a score for individual i is given by

$$\text{PGS}_i = \sum_{j=1}^{J} w_j a_{ij}, \tag{1}$$

where $a_{ij} \in \{0, 1, 2\}$ is the allele count for each SNP j, J is the total number of observable SNPs in the data, and w_j is a weight. A standard choice of weights is to use the association coefficients derived from a GWAS. The most inclusive criterion is to include all the SNP associations from a GWAS, weighting their contribution with their effect size. Because alleles are not randomly observed across genetic addresses, their occurrence varies according to a block structure, called linkage disequilibrium (Nei and Li, 1973). For this reason, PGSs are often calculated using only SNPs that are independent

[12] More details on GREML estimation are in Yang et al. (2011), which provides information on the publicly available genome-wide complex trait analysis software used for this exercise. See also Benjamin et al. (2012). Such estimates can still be biased if the environment shared by related individuals is more similar than that shared among unrelated individuals (Young et al., 2018). In addition, since h^2_{SNP} is based on genotyped and imputed genetic variants, it underestimates the heritability contributed by rare or poorly tagged variants.

from each other.[13] Such independent SNPs are then used to calculate (1), avoiding the potential bias due to oversampling DNA regions that are highly genotyped. The range of possible values that a PGS can take will depend on the total number of SNPs J included in (1), and converges to a normal distribution if the number of independent SNPs is sufficiently high. A common practise is to standardize the score, by subtracting its mean and dividing it by its standard deviation.

Following (1), PGSs can be calculated as linear combinations of single allele scores derived from existing GWAS studies. However, if the weights used for the calculation are derived from the same study, the estimates are biased due to overfitting (Wray et al., 2013). To avoid this problem in our analysis, we adopt a 10-fold cross-validation procedure. This requires us to partition the full sample in 10 random groups, estimate genome-wide association coefficients for nine-tenths of the sample, and then use them as weights, the w_j's in (1), to calculate the PGS on the excluded one-tenth. The same procedure is iterated for each of the other nine partitions until the whole sample is covered. All the polygenic scores we construct are computed using the PRSice-2 procedure on a set of independent genotyped SNPs. This procedure has been shown to have strong out-of-sample prediction and perform better than other existing methods (Choi and O'Reilly, 2019).[14]

We construct three new scores for our four outcomes, that is, age at first sexual intercourse, age at first birth, and completed fertility (which is used also for childlessness). These scores are based on results from the most recent fertility GWAS, including Barban et al. (2016), Mathieson et al. (2020), and Mills et al. (2021). Evidence presented in these studies implicates mechanisms related to reproductive health, puberty timing, and evolutionary fitness in the biological pathways that link the scores to fertility phenotypes. For example, Mills et al. (2021) find that the timing of age at first sexual intercourse and age at first birth are mainly driven by the genetics of reproductive biology and externalizing behaviour. This is supported by biological follow-up that isolates key genes related to follicle stimulating hormones, implantation, infertility, morphogenesis, and binding.

Finally, we construct PGSs for two other phenotypes, which are used in the analysis. One is for educational attainment, and is based on the most recent GWAS by Lee et al. (2018). The other is for risk tolerance and risky behaviors, and is based on the work by Linnér et al. (2019).

2.4 Opportunities, Limitations, and Interpretation

Constructing polygenic scores with precise genomic information gives us an opportunity to step outside the boundaries imposed by the nature versus nurture debate, which most of the heritability research based on twins or siblings data has been constrained to. We shall return to this point in Section 4.

There are also other specific advantages regarding the use of polygenic scores over single genetic variants in our context. We emphasize three. First, since fertility traits and outcomes are known to have some of the highest degrees of polygenicity of any complex phenotype reflecting the influence of many different genes and consistent with a genetic architecture that may be influenced by negative

[13] To select independent SNPs, we use a procedure called clumping, which prevents that SNPs be highly correlated in linkage disequilibrium terms (Choi et al., 2020).

[14] See <https://www.prsice.info/>. We refer to the study by Choi and O'Reilly (2019) for details.

selection (Visscher et al., 2008; O'Connor et al., 2019; Mathieson et al., 2020), polygenic scores acknowledge that each individual falls on a continuum of genetic predisposition which results from small contributions of multiple genetic variants. By contrast, single variants can only be weak predictors of any given fertility phenotype. Second, the utilization of a PGS permits us to be agnostic about the precise biological processes underlying its corresponding phenotype, even when this is likely to be determined by 'distal' genetic influences scattered across the entire genome (Belsky and Israel, 2014).

Third, PGSs — whose use begins to spread among social scientists (e.g., Barcellos et al., 2018; Papageorge and Thom, 2020; Barth et al., 2020; Papageorge et al., 2019; Houmark et al., 2020) — enable us to credibly explore the interplay between genes and environment when the genetic impact on a phenotype is modifiable by environmental influences or, vice versa, when the environmental effect changes with genetic predisposition (e.g., Caspi et al., 2003; Eley et al., 2004; Pezawas et al. 2005; Rosenquist et al., 2015). While twins studies can only speculate on the role played by gene-environment ($G \times E$) interactions as suggested by Almond et al. (2018), analyses based on a single candidate $G \times E$ approach are problematic as they have been shown to have very low statistical power and extremely high false discovery rates (Duncan and Keller, 2011; Benjamin et al., 2012). Using the entire distribution of genetic predisposition, the approach based on polygenic scores reduces the extent of these shortcomings.

We also must draw attention to three important issues related to the use of fertility PGSs.[15] First, polygenic scores can explain only a small fraction of the observed variation in fertility outcomes, especially when compared with heritability estimates based on twins data (Tropf et al., 2017). This problem, also referred to as the 'missing heritability' puzzle (e.g., Manolio et al., 2009; Zuk et al., 2012), is due to measurement error and is often linked to limited power to detect genetic variants with small association sizes in genome-wide association studies, genetic variation that is not captured by SNP-level differences, and failure to account for non-additive genetic effects. The presence of measurement error is likely to lead to downward biased estimates of G and $G \times E$, something we shall come back to in Section 5. Although heritability estimates from twins studies are likely to be upward biased as they control only for shared environmental factors (Felson, 2014), one cannot draw definitive conclusions about the relative importance of genes versus environment in generating fertility outcomes even with the use of polygenic scores. Instead, we reiterate the opportunity that PGSs offer to estimate the size and directions of $G \times E$ effects (e.g., differences in gene-fertility outcome gradients by environmental intensity), and explicitly identify the variants involved in such interactions.

A second issue is that the PGSs we use have been constructed for samples of individuals of European ancestry. Genetic similarity has been shown to be correlated with ancestral origin and geographic proximity. Allele frequency differences between individuals in a given GWAS due to systematic ancestry differences, or population stratification, may lead to spurious associations (e.g., Price et al., 2006; Beauchamp et al., 2011). This is why we restrict our focus on an ethnically

[15] See also the discussions in Papageorge and Thom (2020) and Barth et al. (2020).

homogeneous population of individuals of European ancestry, and cannot generalize our findings to individuals of non-European ancestry (e.g., Campbell et al., 2005; Martin et al., 2017). To adjust for population stratification directly, our analysis will also include a series of principal components of the genetic data, which explicitly account for ancestry differences, minimize the influence of spurious associations, and maximise power.

Finally, using polygenic scores poses new challenges in terms of economic interpretation. A PGS is a linear index of the genetic variants that are predictive of a given phenotype. As discussed in the Introduction, we interpret each of our scores as measuring a subset of the genetic predispositions, or susceptibilities, to a specific fertility outcome, such as the natural capability to produce and care for offspring, sexual desire and competence, and attractiveness.

Existing pathway and gene expression analyses indicate that the genes most heavily weighted in all our scores are implicated in the development of hormones, proteins, and morphogenesis related to diverse aspects of reproductive biology across the life course, including the development of the follicle stimulating hormone (which facilitates the growth of ovarian follicles), placental growth, and infertility (Mathieson et al., 2020; Mills et al., 2021). Although this suggests that reproductive processes are involved, we lack a comprehensive understanding of the biological mechanisms at play. Furthermore, because of the high degree of polygenicity of the outcomes, our interpretation overlaps with, but cannot exhaust, the full and multifarious range of capabilities associated with fecundity, reproductive abilities, and fertility decisions. For instance, we do not expect our PGSs to pin down genetic material that is relevant to fertility phenotypes exclusively, as they may also pick up genetic variants shared by other cues (e.g., risk tolerance and facility with acquiring new skills) or traits that affect other related behaviors, including the ability and willingness to invest in children and parenting skills.[16]

That said, examining how genetic endowments (albeit imperfectly measured) interact with the environment in which women were born and grew up allows us to open up the "black box" of permanent unobserved heterogeneity (e.g., imperfectly controlled biological reproduction) and family fixed effects, which have long characterized the empirical research in family economics (Heckman and Willis, 1976; Rosenzweig and Wolpin, 1980 and 2000; Hotz, Klerman, and Willis, 1997). In particular, gene-environment interactions give us an insight into how environmental factors shape outcome differences due to genetic predispositions or, put differently, how fertility susceptibility changes across birth cohorts (which are likely to face different social and cultural norms about the family) and across geographic and socioeconomic environments (which may reflect different gender attitudes and technologies). As discussed in the Introduction, we view $G \times E$ effects through the lens of a social control pathway, whereby genes express themselves more easily in an environment in which social norms and economic conditions allow a broad range of life-course alternatives.

[16]To assess the importance of this possible overlap, in the next section we shall provide evidence of the genetic correlation among all the polygenic scores under analysis.

3. Data

3.1 The UK Biobank

The UK Biobank is a population-based prospective study initiated by the UK National Health Service (NHS).[17] Between 2006 and 2010, invitations were mailed to 9.2 million NHS registered individuals aged 40–69 (born between 1934 and 1971), who lived up to 25 miles from one of 22 study assessment centers distributed throughout the UK (Allen et al., 2012). A sample of 502,506 individuals agreed to participate (implying a 5.5% response rate),[18] of which 273,384 were women. As part of the survey, study participants went through an assessment that included: a self-completed touch-screen questionnaire, a computer-assisted interview, the collection of physical and functional measures, and the collection of blood, urine, and saliva samples. All physical and medical measures (e.g., anthropometrics and blood pressure) were gathered by trained nurses or healthcare practitioners.

Every participant in the study was genotyped. This makes the UK Biobank one of the largest publicly available genetic resources in the world and a primary source of study for genetic discovery and drug development.[19] Although the survey has limited information on respondents' socioeconomic background and most of the information on family environment is available retrospectively, it provides geo-coordinates for both place of births and current residence at interview, with one kilometer grid resolution.[20] We assign all respondents with available geographical coordinates to their region and local authority district (LAD) both at birth and at the time of interview.[21] As shown in Figure A1, the geographic distribution of the UK Biobank respondents' area of birth is considerably widespread across the country.

In the analysis, we exclude respondents from Northern Ireland (who represent 2% of the whole sample) and respondents born outside the British Isles (7.8%). We also exclude the cohorts born in 1934–37 and 1970–71, since only few respondents were born in those years (a total of 1.3% of the original sample), making cross-cohort comparisons difficult. Furthermore, as mentioned in the previous section, we restrict our attention to individuals with European ancestry to adjust for population stratification.[22] Last, we focus only on female respondents and we drop respondents for which the area of birth is missing. All these restrictions lead to a final sample of 168,757 unrelated

[17] For a detailed description of the data as well as access and governance issues, see Sudlow et al. (2015).

[18] Other 122 respondents withdrew their participation after data collection.

[19] Genotyping was performed using the Affymetrix UK BiLEVE Axiom array on an initial 50,000 participants; the remaining participants were genotyped using the Affymetrix UK Biobank Axiom® array, which genotyped approximately 850,000 genetic variants. The two arrays are extremely similar, with over 95% common content. The genetic data were later imputed using two different reference panels: the Haplotype Reference Consortium panel and the UK10K together with the 1000 Genomes panel, leading to a dataset with 73,355,667 different genetic variants. The genetic sample consists of 488,265 participants, 264,706 of which are women.

[20] The grid coordinate data are provided in the British National Grid (i.e., OSBS 1936) projection. OSGB1936 is the Ordnance Survey National Grid geographic reference system used in Great Britain.

[21] Our definition of LAD is based on the 2018 Census boundaries, which leads to 380 LADs. The data are available from the Office for National Statistics (England and Wales), the National Records of Scotland, and the Northern Ireland Statistics and Research Agency.

[22] Our selection is based on self-identification, whereby we retain only individuals who self-identify as white British or other white. We perform an additional check using genomic data where we exclude the small group of individuals who self-identify as white Europeans but have genetic ancestry that deviates from European origins by more than five standard deviations from the first two principal components. The results found using this alternative subsample are identical to those shown below and are thus not reported.

women used to construct our polygenic scores. The analysis will also leverage sizeable subsamples of sisters and mother/daughter pairs.[23]

3.2 Fertility Outcomes

The UK Biobank main questionnaire collects fertility histories, where women are asked to provide information on their ages at first sexual intercourse, first live birth, and on the total number of children they ever had. Based on this information, we construct the following four fertility outcomes: age at first sexual intercourse (A1S); age at first birth (A1B); completed fertility (CF); and childlessness (CLN). To minimize right censoring issues for CF and CLN, the sample is restricted to women who are aged 45 or more at interview.

As Britain went through profound socioeconomic transformations since World War II, so did the reproductive behavior of the women in the UK Biobank sample (Wringley and Schofield, 1989; Hobcraft, 1996; Woods, 2000; Garrett et al., 2001; Berrington, 2004; Guinanne, 2011). Figure 1 displays the time trends in all fertility outcomes by birth cohort. First births have been postponed by almost three and a half years (from 24.3 up to 27.5 years), whereas age at first sexual intercourse has been brought forward by three years (from 20.9 down to 17.8 years). On average, therefore, women born in the late 1960s were sexually active for almost 10 years before giving birth. Conversely, women born in the late 1930s and early 1940s had their first child about three years after their sexual debut. Moreover, as women in the early birth cohorts had on average 2 or more children, those in more recent cohorts had fewer than 1.5 children.[24] Finally, the fraction of women who remained childless more than doubled, from 12% among women born at the start of the UK Biobank to almost 30% among those born in the early 1960s.

Table 1 shows descriptive statistics for all four outcomes. Besides the summary statistics for our measures of genetic endowments and environmental risk which we describe next, the table also reports additional information on the respondents' early life conditions, such as their self-reported birth weight (in kilograms), smoking status of their mothers during pregnancy, and whether they were breastfed. These measures have been used to proxy individual socioeconomic background and shown to be correlated to later outcomes (Black et al., 2007; Bharadwaj et al., 2014; Fitzsimons and Vera-Hernández, 2016; Almond et al., 2018). We shall use them in a series of robustness checks.

3.3 Measures of Genetic Assessment

Before describing our polygenic scores, we present heritability estimates which can be related to the evidence found with twins studies. For each of the outcomes, Appendix Figure A2 shows the SNP-based heritability estimates and the 95% confidence interval around them by birth cohort. These are obtained using the GREML model mentioned in Section 2 and computed as 5-year moving averages. Since the short-run variation in genetic composition is expected to be limited, changes in h^2_{SNP} over the sample period are likely to reflect changes in the environment (e.g., differential

[23] In the main analysis sample, we randomly select only one individual per family to avoid issues of overfitting.

[24] Although some women could have had more children after the last available interview, this will not change the general picture given in Figure 1.

assortative mating and internal migration), and may also give an indication of the role played by gene-environment interactions. For age at first sex, we observe stable heritability across all birth cohorts, ranging between 0.10 and 0.18. This is also the case for completed fertility, with only a slight increase to about 0.10 for the late 1950 birth cohorts. There is instead a clear upward trend in h^2_{SNP} for A1B, which doubled from approximately 0.12 to 0.24 across the cohorts in the sample. Heritability in childlessness displays a similar increasing trend, although the estimates are lower and can only be obtained for fewer cohorts.[25] Taken together, these patterns reveal cross-cohort variation in heritability, which we shall come back to below. This in turn suggests the importance of factors other than genetic traits in explaining changes in outcomes, especially for A1B and CLN, over the 30-year span of the data.

Moving on to the polygenic scores, Figure 2 plots the kernel-smoothed density of the three score variables in our study, as well as the PGSs for educational attainment and risky behavior. Values of each score have been demeaned and rescaled to measure standard deviations relative to the corresponding means. The plots indicate that each distribution is approximately normally distributed and symmetric. In fact, for every trait (including education and risky behavior), we formally cannot reject the null hypothesis that the score is normally distributed.

Polygenic scores for different outcomes are likely to pick up similar genetic material, given non-negligible phenotypic overlaps. This is reflected by the genetic correlations shown in Appendix Table A1. The table reports also the correlations with the scores for education and risky behavior, as well as the cross-correlations of all outcomes. Age of sexual initiation and age at first birth have a phenotypic correlation of about 0.33 and share the largest genetic overlap in the sample, 0.39. Completed fertility is negatively correlated with the other three outcomes, especially childlessness (–0.73), and its PGS is also inversely correlated to the scores of both A1S and A1B. The genetic correlation of education and age at first birth is high (0.31), but the genetic predisposition to risky behavior is uncorrelated with any of the other polygenic scores, including that of A1S.

We now go back to the evidence revealed by the heritability analysis that the role of genetic endowments may have changed across cohorts. In Appendix Table A2, we explore this issue by testing the stability of each polygenic score over time, regressing outcome specific PGSs on birth cohorts (grouped in five-year bands) while controlling for local authority district of birth and the first 10 principal components of the full matrix of SNP data to account for population stratification. For A1S, 13 (out of 15) cohort group comparisons show statistically different associations with the polygenic score. This is also true for nine and eight of the comparisons for A1B and CF, respectively, and for five and 11 of the comparisons for the education and risky behavior PGS regressions, respectively. Such results, which are consistent with those presented by Beauchamp (2016) and Kong et al. (2018), confirm the cross-cohort variation in genetic endowments found earlier with the heritability analysis.

Another check of the heterogeneity in genetic predisposition by cohort is to estimate the degree

[25] As mentioned in Section 2, twins studies find higher heritability estimates. For example, h^2 is estimated to be approximately 0.25–0.35 for A1B (Tropf et al., 2015) and 0.1–0.3 for CF (Kohler et al., 1999).

to which the association of the polygenic scores with the fertility outcomes changes across cohorts.[26] The marginal effects of the polygenic scores by cohort of birth, α_3 in (2), are summarized in Figure 3. There is evidence that the relationship between PGS and outcome varies substantially across cohorts.[27] For instance, the marginal effect of one standard deviation increase in the score for age at first birth is just less than 2.5 years for women born in 1939 but almost 3.5 years for those born 30 years later. This suggests a greater role played by the genetic endowments relevant to this outcome for more recent cohorts. We observed the opposite trend for age at first sexual intercourse, with the marginal impact declining from about 2.2 years for the earlier cohorts to 1.8 years for more recent cohorts. Although the association is relatively stable in the case of completed fertility, we find a strong and increasing (in absolute value) negative correlation for childlessness throughout the sample period. These results bolster the heritability estimates and suggest that there is variation in genetic composition across cohorts. This variation is unlikely to be driven by actual changes in the genetic endowment of the women in the sample (Courtiol et al., 2016). Instead, it may in part reflect the presence of interactions between our polygenic scores and the changing physical and socioeconomic environment faced by women from different birth cohorts.

3.4 Measures of Environmental Influence

Our key measure is given by the change in availability of the modern oral contraception (the "contraceptive pill"), which has been recognized as an indicator of gender-equal social norms and one of the main channels leading to women's economic freedom and female empowerment (Goldin and Katz 2002; Bailey 2006). An extensive literature has established the role played by gender norms in explaining women's emancipation (e.g., Fernández, 2011; Alesina et al., 2013). Thus, we take the diffusion of the pill and the gender norms associated with it as the broad measure of environment relevant to female fertility decisions.[28]

The contraceptive pill was first introduced in the UK in 1961. Initially accessible only to married women, its availability was extended to everyone in England and Wales in 1967, in 1968 in Scotland, and in 1969 in Northern Ireland. Women in the UK Biobank are asked retrospectively whether they

[26] Specifically, we estimate

$$Y_{iac} = \alpha_0 + \alpha_1 G^Y_{iac} + \sum_{c=1938}^{1968} \alpha_{2c} \mathbb{I}_{iac} + \sum_{c=1938}^{1968} \alpha_{3c} \left(\mathbb{I}_{iac} \times G^Y_{iac} \right) + \sum_{k=1}^{10} \gamma_k PC_{k,iac} + \varphi_a + \varsigma_{iac}, \tag{2}$$

where Y_{iac} is the fertility outcome for individual i born in cohort c in local authority district (LAD) a; $\mathbb{I}_{iac}$ is an indicator function taking value one if woman i was born in cohort c and LAD a, and zero otherwise; G^Y_{iac} is the outcome specific PGS calculated from the entire sample with the 10-fold cross-validation process described in subsection 2.3; φ_a indicates LAD fixed effects; PC refers to the principal components of the full matrix of SNP data, of which we use the first ten to control for population stratification; and ς_{iac} is an idiosyncratic error term. Standard errors are robust and clustered at the LAD level.

[27] In general, the variation in the marginal effect estimates is larger for the youngest cohorts for which sample sizes are relatively smaller.

[28] For Britain, Murphy (1993) argues that the diffusion of the oral contraceptive pill is the single most important determinant of fertility change during the 1960s and 1970s. As discussed in the Introduction, other supply-side factors (such as the legalization of abortion and the advent of feminism) or demand-side factors (such as the introduction of sex discrimination legislation) might have played a role. Dates on abortions, however, are not available in the UK Biobank. In addition, the first UK Equal Pay Act was introduced in 1970 and the first Sex Discrimination Act in 1975, arguably much later than most of the changes already observable in the 1960s.

have ever used the pill and at which age they began taking it. For each one of the ten data partitions used to calculate the polygenic scores and for each LAD in the data, we compute the proportion of childless women aged 18 or above who ever took the pill using those in the other nine partitions. We repeat the same calculations separately for each of the remaining partitions and end up with our final measure when all partitions are exhausted.[29] We then link this variable to each woman when she was 18-years old.

Figure 4 shows this measure by calendar year aggregated up to broad geographic regions. As expected, contraceptive pill usage varies substantially across cohorts. Women born in the earlier cohorts of the UK Biobank did not have access to the pill until they were in their thirties, while those born in the most recent cohorts faced an environment in which approximately 80% of the 18 years old in their area used the pill. Women born between 1950 and 1965, who represent a large fraction of those observed in the UK Biobank, faced the sharpest variation in pill diffusion as its usage rose from less than 10% to more than 70%. The data reveal also some (albeit smaller) regional variation, with London having pioneered pill usage especially during the 1970s, and Scotland lagging 5–8 percentage points behind.[30]

We should reiterate we do not rely on measures of environment defined at the individual level, e.g., parental education or maternal occupational status during the respondents' childhood. In part, this is due to the fact that the UK Biobank does not collect such information. The use of these alternative proxies may be problematic nonetheless, for they can be strongly related to the same unobserved processes that give rise to the individual fertility outcomes we study, over and above the genetic endowment accounted for by the polygenic scores. Our measure of E instead reflects changes in institutions, policies and markets that are largely external to the family of origin. Internal (family) environments, which may strengthen or weaken external changes, will come into the picture when we estimate sister fixed effects models and mother-daughter comparisons.

For robustness, we also use three other measures of E, which capture potentially different aspects of female empowerment. One is given by the rise in female educational attainment, which many indicate as an important contributor to the reduced fertility observed over a considerable part of the twentieth century (e.g., Currie and Moretti, 2003; McCrary and Royer, 2011; Eckstein et al., 2019). Women in the UK Biobank are asked to report the age at which they completed their full time education. We use this information to construct the proportion of women leaving post-compulsory education by local authority district and year of birth. This is then linked to each woman in the sample when she was 16 years old.[31] Appendix Figure A3 shows the evolution of this alternative

[29] An alternative approach is to construct E over the whole sample rather than relying on the 10-fold cross-validation process just illustrated. Although this alternative procedure may minimize inferential issues due to the double cross validation of both E and G, it delivers a measure of pill adoption that may suffer from a reflection problem (Manski, 1993). In subsection 5.2, where we report on our robustness analysis, we mention the results found with this different measure.

[30] We also constructed an alternative measure of E, which is not stratified by local councils and relies exclusively on the time variation in pill usage. While this measure follows the temporal introduction of the contraceptive pill more closely, it fails to replicate the spatial granularity of its adoption. In our main analysis, therefore, we use the measure of E described in the text. Repeating it with this more aggregate counterpart, however, leads to virtually identical results, which are not presented for the sake of brevity.

[31] An educational reform implemented in England and Wales in 1972 raised the school leaving age from 15 to 16

measure of E. We observe a smooth increasing trend in this proportion over the pre-1957 cohorts, with substantial differences across broad geographic areas. For the post-1957 cohorts affected by the 1972 reform, there is a steep rise in the proportion leaving education after age 16 and a much smaller cross-sectional geographic variation.

The other two measures reflect the change in economic and occupational conditions identified by the variation in regional GDP per capita and the female share of employment in services. The sustained growth in female labor force participation, strongly associated with the rise of the service economy and the diffusion of labor-saving appliances, has been seen as another driver of the changes that have transformed the family from World War II (WWII) onwards (e.g., Greenwood et al., 2005; Goldin, 2006; Buera and Kaboski 2012). We use the statistical series published in the Rosés-Wolf database (Rosés and Wolf, 2018), which contains regional (but not LAD level) demographic and economic indicators from 1900 to 2010. We focus on regional GDP per capita and employment rates in services, and link both measures to each woman in the UK Biobank through her birth year and region of birth when she was 18 years old. Panel A of Appendix Figure A4 shows the trends in income per capita, while panel B displays the trends in female employment shares, both aggregated up to broad geographic regions. For both measures, we observe substantial changes across time and space. People from historically poor regions (e.g., the East Midlands, Wales, and Scotland) nearly doubled their income over the 32 birth years in the sample. Londoners experienced a 50% rise over the same time period and, at any point in time, their income was twice as large as that of residents in the poorer regions. The employment share in the service sector also went through considerable change. Most regions, except London, experienced a decline from the pre-war years to the mid 1950s followed by a steady increase until the end of the sample period, with cross-regional differences remaining large over time, even into the 21st century.

We conclude by looking at how G varies with our main measure of E. Appendix Figure A5 shows the scatter plot of each of the three polygenic scores (for A1S, A1B, and CF) and pill diffusion. In all three cases, the PGS is essentially the same across the entire distribution of E, indicating that there is substantial randomness in the way genes are distributed across the population. This is not in line with a notion based on group-level genetic differences (Herrnstein and Murray [1994]; see the counterargument by Heckman [1995]). It is also hard to reconcile with the hypothesis that social stratification leaves visible marks in geographic arrangements of common allele frequencies, as argued by Abdellaoui et al. (2019). Despite this evidence, in the analysis below we will check if our estimates are robust when we restrict attention to the subsample of "stayers", i.e., women who did not move from the area where they were born.

4. Conceptual Setup and Research Design

To guide the empirical analysis and the interpretation of the results, we provide a simple framework borrowing from Becker and Lewis (1973) and Becker (1991). Suppose each woman (or household)

years (Barcellos et al., 2018). In our analysis, we shall account for this reform by including an indicator variable that takes value one for the affected cohorts born after September 1957, and zero otherwise.

maximizes a utility function that depends on genetic fitness, f, and a single aggregate commodity Z, denoted $U(f, Z)$. Genetic fitness is the relative contribution of a genotype to the next generation's distribution of genotypes, and here is a scalar but broadly defined to include the natural capability to produce and care for offspring, sexual competence, reproductive susceptibility, quantity and quality of children, and the cost of producing child quality. Each woman decides whether her fitness is better enhanced by her own offspring or by her relatives' (siblings, cousins, etc.) offspring. This implies childlessness can be an outcome (Hirshleifer, 1977).

As in Becker (1976), genetic fitness is a commodity produced by households using a variety of inputs, including their own time, innate skills and dexterity, purchased goods, and the physical and social environment. To simplify, we assume $f = f(X; G, E)$, where G is the woman's genetic endowment (proxied by polygenic scores), E is her environment (proxied by the pill diffusion), and X is a composite good. Fitness does not have a market price, since it is not purchased directly, but has a shadow price, π, which is defined as the value of the goods used in changing fitness by one unit. If p is the cost of Z and I full income, the budget constraint of a household is given by $\pi f + pZ = I$.[32] Maximizing utility subject to the budget constraint gives the following marginal utility (MU) condition:

$$\frac{\partial U/\partial f}{\partial U/\partial Z} = \frac{MU_f}{MU_Z} = \frac{\pi}{p}.$$

The demand for fitness, D_f, will then depend on the relative price of genetic fitness and full income and will have the usual income and substitution effects. A reduction in the relative price of fitness will increase the demand for fitness and reduce the demand for other commodities, holding real income constant. The decrease in π/p, for instance, may be driven by an environment more conducive to the enhancement of fitness, better genetic endowments, greater skills, or more time in producing fitness. Note that, since π depends on G and E, we can re-write D_f as a reduced form demand function, $D_f(G, E, p, I)$. Linearizing this expression and conditioning out p and I, which we do not observe in the data, will lead to the following reduced form conditional demand (Pollak, 1969):

$$f = a_g G + a_E E + \xi, \tag{3}$$

where ξ captures other components (including price and income), some of which could be correlated to both genetic predisposition and environmental factors.[33] Some of such components can be subject to genetic influences and interact with the environment either reinforcing or attenuating its impact on f (Kohler et al., 1999; Turkheimer et al., 2003; Rimfeld et al., 2018). For instance, sexual desire may be inhibited or stimulated depending on the socioeconomic milieu in which individuals grow

[32] The close relation between fitness and goods can be made transparent by writing the production function for fitness as $f = \kappa XGE$, where κ is a parameter that captures efficiency in producing fitness. In this case, $\pi = \partial(pX)/\partial f = p/\kappa GE$.

[33] Including education, as a predictor of income, and time trends, as price shifters, does not change our main results. See subsection 5.2 for more details.

up. Similarly, women who inherit genetic endowments that are conducive to relatively greater fitness may have grown up in families with more favorable environments for child development (Cunha and Heckman, 2007; Cunha et al., 2010; Currie, 2011; Almond et al., 2018; Houmark et al. 2020). We capture this possible relationship with a gene-environment interaction term, $G \times E$, so that the conditional demand for fitness (3) becomes

$$f = b_1 G + b_2 E + b_3 (G \times E) + \nu. \tag{4}$$

where ν is an unobserved term. Reiterating the arguments anticipated in the Introduction, we interpret $G \times E$ as a social control mechanism, whereby genetic influences captured by G are mediated by prevailing social norms, embedded in E. As pill usage becomes more widespread, genes may magnify or moderate its impact on fitness. Expression (4) goes beyond the tradeoff between nature and nurture, for it explicitly recognizes that environmental influences and genetic components interact to produce fitness.

In our empirical analysis, f is proxied with the four fertility outcomes described in the previous section, and expression (4) is operationalized with the following regression model:

$$Y_{ia} = \theta_0 + \theta_1 G^Y_{ia} + \theta_2 E^Y_{ia} + \theta_3 \left(G^Y_{ia} \times E^Y_{ia} \right) + \mathbf{X}'_{ia} \Theta + \delta_a + \varepsilon_{ia}, \tag{5}$$

where Y_{ia} is the outcome for woman i born in local authority district (LAD) a, G^Y_{ia} is the outcome specific polygenic score for i from area a, E^Y_{ia} is the outcome specific contraceptive pill utilization rate which woman i was exposed to in LAD a, δ_a represents district fixed effects; $\mathbf{X}$ is a vector of controls, which include the first 10 principal components of the full matrix of SNP data to control for population stratification, their interactions with E to account for the possible interplay between ancestral commonality and pill usage absorbing any further stratification bias from the $G \times E$ term (Keller, 2014), and a WWII indicator (taking value 1 if i was born between 1939 and 1945, and 0 otherwise);[34] and ε_{ia} is an individual specific idiosyncratic error term. Robust standard errors are clustered at the district level, where the variation in E occurs. In robustness analysis, for each outcome, we include all the fertility scores as well as the PGSs for educational attainment and risky behaviors. We also perform sensitivity assessments on E, using female educational attainment rates by LAD, regional GDP per capita, and regional employment rate in services rather than pill usage. The estimates from all these exercises are discussed below, but we anticipate that our main results are unchanged.

[34] Other controls are proxies of early life conditions (such as being breastfed and birth weight), which are not available for all the women in the sample. To maximize sample size, they are not included in the benchmark analysis. Their inclusion, however, does not alter our main findings, as we document in subsection 5.2.

5. Empirical Results

5.1 Baseline Estimates

Table 2 summarizes our main results. Each column reports the estimates from equation (5) for a different fertility outcome. All the estimates in the table are statistically significant even at the tighter p-value threshold of 0.005, which Benjamin et al. (2018) argue should be the standard of evidence for claims of new discoveries.

Looking at G, a one standard deviation increase in each outcome-specific PGS is associated with: (i) a postponement of both age at first sex and age at first birth by 2.1 and 2.7 years, respectively; (ii) an increase in the total number of children born to a woman by 0.66; and (iii) a lower chance of being childless by 14 percentage points. Genetic penetrance, therefore, plays a key role in explaining the observed variation in all outcomes, both statistically and substantively.[35] This is upheld by the incremental R^2 figures (bottom row in Table 2), according to which genes account for not less than 17% of the variation in childlessness, 32% of the variation in the age at first sexual intercourse and completed fertility, and 42% for age at first birth.[36]

Turning to the association of E with Y, we find that a 10 percentage point increase in pill exposure (which corresponds to a modest change when comparing women born in the 1950s to those born in the 1960s) leads to: (i) 0.26 year reduction in A1S; (ii) 0.36 year postponement of A1B; (iii) a compression in CF by 0.07 children; and (iv) a 2.2 percentage point increase in the probability of CLN.[37] These associations are quantitatively large, and suggests that our measure of environmental risk is equally important in affecting fertility fitness as much as the polygenic scores.

Before examining the role played by gene-environment interactions, we mention the results found when such interactions are excluded. The θ_1 and θ_2 estimates from this exercise reported in Appendix Table A3 are close to those just discussed, both in sign and magnitude. The evidence is that E and G make complementary contributions to age at first birth, while having offsetting influences on the remaining three outcomes.

An indirect way of ascertaining the importance of genes and their interactions with the environment is given by the comparison (in percent change) of the marginal effect of E with respect to a model in which all genetic influences are shut down, i.e., both θ_1 and θ_3 in (5) are set to zero. These

[35] As mentioned in Section 2, all PGSs are measured with error, which may yield estimates of θ_1 (and θ_3) that are biased downward in absolute value. Addressing this issue by either meta-analyzing GWAS results from multiple studies (e.g., Dudridge, 2016; Becker et al., 2021) or using instrumental variables regressions on PGSs from independent GWAS samples (e.g., DiPrete et al. [2018]; van Kippersluis et al., [2021]; see also Gillen et al. [2019]) would be likely to improve the predictive power of G and $G \times E$ and increase their impacts on Y. Although interesting, we leave this extension for future research.

[36] Since these statistics are computed with respect to a model that does not include genetics as a comparator (i.e., $\theta_1 = \theta_3 = 0$), they also account for the contribution of $G \times E$. The corresponding figures without gene-environment interactions are comparable to those presented in Table 2, and so are the estimates. These results, which are shown in Appendix Table A3, are discussed below.

[37] To compare our E–Y results to some of the existing estimates of the impact of the legalization of the pill available for the United States, we ought to redefine outcomes appropriately. Discussing the results from this analysis in detail goes beyond the scope of the paper. Nevertheless, for completeness, we report the estimates from this exercise in the Online Appendix Table A4. In general, we find greater associations of the pill diffusion (our measure of E) with the likelihood of having a child by age 23 (a proxy for the probability of being married by age 23, used by Goldin and Katz [2002]), the likelihood of having a child before age 22 (used by Bailey [2006] and Myers [2017]), or the number of children ever born (Bailey, 2006).

results are reported in the next-to-last row of Table 2. When G and $G \times E$ are excluded from the analysis, the contribution of E is underestimated by 6.2% for A1S, 2.6% for A1B, 8.1% for CF, and 6.3% for CLN.

A direct way of assessing the interplay between genes and environment is to focus on the $G \times E$ effect as captured by θ_3 in Table 2. We find evidence of statistically significant gene-environment interactions for all outcomes. In particular, the delay in age at first sex induced by the genetic predisposition captured by the polygenic score is moderated by gene-environment interactions. Conversely, $G \times E$ effects amplify the delay in the age at first birth, the increase in the number of children, and the reduction in the probability of childlessness implied by genes alone. Seen from the perspective of the environmental influences, the estimates reveal a strong presence of environmentally mediated risks. Both the anticipation of sexual debut and the postponement of motherhood led by the popularization of the pill are reinforced by gene-environment interactions. The reduction in family size and the rise in childlessness associated with greater pill usage are, instead, weakened by the way our PGSs interact with the diffusion of the contraceptive pill.

Within the framework developed in Section 4, these findings are consistent with a reduction in the relative price of fitness, in particular with a decrease in the shadow price π, which may be driven by an environment conducive to fitness advancement. This means that genetic influences on fertility may become more important when social norms and economic conditions allow a broad range of life-course alternatives. Pre-pill cohorts, whose early lives were affected by traditional gender norms and relatively poorer economic conditions, may exhibit low levels of genetic influence. Later cohorts, which went through the pill revolution, reflect stronger genetic effects.[38]

In Section 6, we will explore these interactions in greater detail. Ours are the first results to lend explicit support to the existence of environmentally mediated genetic effects on fertility-related behaviors, which have been posited by standard economic theories of fertility (Becker, 1960 and 1991) and suggested, but not properly documented, by other social scientists (e.g., Udry, 1996; Kohler et al., 1999). They also confirm the burgeoning view that the sharp distinction between nature and nurture is inadequate (Gluckman and Hanson, 2005; Rutter, 2006; Heckman, 2006 and 2007; Cunha and Heckman, 2007; Kong et al., 2018; Barcellos et al., 2018; Houmark et al., 2020).

5.2 Robustness Checks

We present results from several sensitivity exercises, which are reported in the Online Appendix. First, we re-estimate (5) including proxies of early life conditions in $\mathbf{X}$. These comprise the respondent's birth weight and indicators of smoking status of the respondent's mother during pregnancy and whether the respondent was breastfed. Such measures have been shown to be correlated to pre- and post-natal parental investments and predictive of both early and later child achievement (Almond et al., 2018). Information on these controls is available at most for 55% of the samples

[38] Interestingly, in plant physiology and botany, a parallel of the manifestation of this interaction is called seed dormancy, which is an evolutionary adaptation that prevents seeds from germinating during unsuitable ecological conditions (Bewley, 1997; Finch-Savage and Leubner-Metzgerthat, 2006). Seed dormancy, typically, affects large populations of plants dispersed over vast geographic areas, mirroring the social control mechanism implied by the cross-cohort diffusion of the contraceptive pill technology in our case.

used to estimate our benchmark specifications. Despite the loss in sample size, the estimates in Appendix Table A5 which account for early life controls are remarkably similar to those reported in Table 2, except for the gene-environment effect on completed fertility that is no longer statistically significant.

Second, we extend the age window over which pill exposure is measured. Specifically, the exposure used is widened from age 18 to ages 18–30. With this extension, we try to lessen measurement error problems that could be generated by the original single-year measure. Doing so, however, may lead us to detect changes in the environment that did not influence the women in the sample while they were growing up. As the estimates in Appendix Table A6 reveal, this exercise does not modify our main results. Repeating the analysis with exposure defined between ages 13–18 leads to the same estimates.

Third, we check whether all fertility outcomes are driven by a secular aggregate time trend common to all women, rather than by time (and geographic) variation in pill diffusion. We do this by adding 5-year birth cohort fixed effects as additional controls in (5). The estimates in Appendix Table A7 are virtually identical to those reported in Table 2.[39]

Fourth, we consider two other genetic influences that may be relevant to fertility choices in addition to those already accounted for in our baseline analysis. One of them is the PGS for educational attainment based on the work by Lee et al. (2018), and the other is the PGS for risky behaviors developed by Linnér et al. (2019). The results in Appendix Table A8 corroborate the baseline estimates reported in Table 2.[40] This suggests that our main findings do not merely reflect genetically endowed academic abilities or degrees of risk tolerance and adventurousness, despite the non-negligible genetic overlaps shown in Appendix Table A1 and discussed in Section 3. To account even more comprehensively for the genetic correlations across all polygenic scores, we repeat the analysis including all PGSs in (5) for each outcome. The estimates in Appendix Table A9 reiterate the basic message of the benchmark results.

Fifth, we perform a series of tests with the three alternative measures of E described in Section 3. The results shown in Appendix Table A10 indicate that using education expansion (rather then pill diffusion) as a measure of environment does not affect the overall results. The only change is that the gene-environment effect on completed fertility is now statistically indistinguishable from zero. Using regional GDP per capita confirms the benchmark estimates, but leads to a statistically insignificant $G \times E$ effect on A1S and attenuated impacts of G on A1B and childlessness. As evidenced by the greater standard errors in Appendix Table A11, these results are likely to be driven by the higher level of geographic aggregation implied by this measure of E. They could also be driven by the fact that GDP and pill exposure may capture different dimensions of female empowerment. Finally, Appendix Table A12 confirms our main estimates when E is proxied by the regional employment share in services, except for the statistically insignificant $G \times E$ effect on A1S and G impact on CLN.

[39] As mentioned in Section 4, besides time trends, which can be seen as predictors of price shifters, we also included education, as a proxy for income. All the results are similar to those shown in Appendix Table A7 and thus not reported for brevity.

[40] Interestingly, the PGS for educational attainment is not statistically significantly correlated to childlessness, while the PGS for risky behaviors has an effect on A1S and A1B, but no effect on the other two outcomes.

Again, the higher level of aggregation and the different facets of female emancipation summarized by this measure of E may be responsible for these departures from the baseline results.[41]

Sixth, standard errors in Table 2 are clustered at the district level. While this clustering follows the measure of E, it may not be appropriate for the polygenic scores, which vary at the individual level, although these are also known to have a broader geographic structure that reflects population stratification and its possible interactions with pill usage. We already account for these latter aspects directly with the inclusion of the first 10 principal components of the genetic data and their interactions with E, but other clustering levels might be considered. We therefore re-estimated (5) clustering either at the individual level or at the broader regional level for all outcomes. The inference resulting from both of these clustering alternatives is the same as that reported in Table 2. We also find similar results when we use two clusters only, i.e., one for Scotland and the other for England and Wales combined given that the pill was introduced in separate years for these two groups of countries.

Seventh, we re-fit (5) on the sample of stayers, i.e., women whose local authority district of residence at the time of interview is the same as that observed at birth. In spite of the smaller sample size, all the estimates in Appendix Table A13 confirm the baseline results (except the $G \times E$ effect on total number of children which becomes insignificant), indicating that the potential selection bias due to migration is likely to be limited.

6. Understanding the Interplay Between Genes and Environment

The results in the previous section reveal the importance of genetic influences on fertility outcomes. They also provide clear evidence of the presence of environmentally mediated genetic effects, with gene-environment interactions either attenuating or reinforcing the direct genetic associations or the direct environmental effects. This interplay seems to emerge in particular when environmental conditions expand the range of socioeconomic opportunities for women, which we relate to female emancipation. The already-mentioned strong similarity of these $G \times E$ interactions with seed dormancy adaptations in plant biology is interesting and worth of further research in the future.

In what follows, we deepen the implications of these insights and ask whether the impact of this interplay changes across the distributions of polygenic scores and environmental risk exposure. The existence of such nonlinearities provides further evidence against the nature-versus-nurture dichotomy, even without the presence of gene-environment interactions. Moreover, to control for shared environmental risks, we estimate family-based models. Although the influence of genotypes is always reduced in these models, they can help us to uncover the presence of genetic nurture and assess the extent of $G \times E$ effects even when the variation in G or the differences in G and E are more restricted than in the general population. We conclude this exploration with a series of simulations that enable us to understand better the implications of our results.

[41] As mentioned in subsection 3.4, we also experimented with an additional measure of pill adoption constructed using the whole sample rather than a 10-fold cross-validation procedure. We detect no inferential problem with the baseline measure of E and find that the estimates obtained with this alternative measure are statistically indistinguishable from those shown in Table 2. Such results, therefore, are not presented.

6.1 Nonlinearities in Genetic Endowments and Environmental Risks

We modify the benchmark specification (5) and, for each fertility outcome, estimate

$$Y_{ia} = \phi_0 + \sum_{d=1}^{10} \phi_{1(d)} G^Y_{ia(d)} + \phi_2 E^Y_{ia} + \sum_{d=1}^{10} \phi_{3(d)} \left(G^Y_{ia(d)} \times E^Y_{ia} \right) + \mathbf{X}'_{ia}\Phi + \delta_a + \upsilon_{ia}, \tag{6}$$

where d refers to the deciles of the G (PGS) distribution. The results on the ϕ_1 and ϕ_3 estimates are displayed in Figure 5. As expected by construction of the polygenic scores, the higher the decile of G, the larger (in absolute value) the association with the outcome, with childlessness being the only phenotype with a negative $G-Y$ gradient.

There is evidence of ample heterogeneity in gene-outcome gradients. Women with median genetic predispositions to delay their first sexual intercourse (motherhood) postpone sexual debut (first birth) by almost 2.5 (4) years, while those whose predispositions are in the top decile postpone the first intercourse (motherhood) by about 7 (9) years. Women whose endowment falls in the interquartile range of the susceptibility to have children are on average expected to have one child extra (from 0.6 to 1.4) compared to those in the bottom decile, while those in the top decile are predicted to have up to 2.3 additional children. Finally, the likelihood of being childless by age 45 is reduced by 50% if women's scores are in the top decile of the PGS distribution.

Although gene-environment interactions are quantitatively smaller, they also display a good deal of heterogeneity. As genetic predispositions induce a higher age at first sex, gene-environment interactions reduce this association more for those in the top deciles of the PGS distribution. This gives evidence of nonlinear $G \times E$ effects which, to a small degree, mitigate the direct impact of G. For age at first birth, instead, the postponement driven by G is reinforced by gene-environment interactions which gain strength as the value of the polygenic score increases. There is reinforcement also for childlessness, but only among women in the top half of the PGS distribution, that is, those with stronger genetic susceptibilities not to have children at all. Finally, for completed fertility, gene-environment interactions magnify the negative impact of G, but only moderately across the whole score distribution.

Repeating the analysis but stratifying the estimates by decile of E while keeping G continuous leads to specular results, which emphasise that gene-environment interactions are heterogeneous across the E distribution. We report the results in Appendix Table A14. The 3-year first birth postponement among women from areas in the top deciles of pill usage is magnified by another 10-month delay through the interplay between pill diffusion and genetic predispositions to delay motherhood. Similarly, $G \times E$ effects amplify the reduction in A1S driven by female empowerment: for women who had much exposure to pill use (i.e., those in the top half of the distribution of pill utilization), gene-environment interactions magnify the 3-year reduction in age at first sexual intercourse by another 4 months. $G \times E$ effects for childlessness, instead, moderate the direct impacts of E across the distribution of pill usage. For instance, although women in the top decile of the pill exposure distribution face a 17 percentage point increase in the likelihood of being childless compared

to women in the bottom decile, their gene-environment interactions reduce that association by half.[42]

Taken together, these and the previous estimates uphold our benchmark results. Reproductive fitness is powerfully shaped by both genetic predispositions and societal environmental risks. On the one hand, the external environment, which embeds the reduction in the relative price of fitness, affects fertility outcomes in different ways depending on individuals' genotypes; on the other hand, genes affect outcomes differently depending on women's circumstances. The diffusion of the contraceptive pill, a market manifestation of the decline in π/p, seems to have complemented women's genetic predispositions to delay their first childbirth and remain childless and, at the same time, to reinforce their susceptibilities to anticipate sexual debut and lower the total number of children they give birth to. These findings, in turn, suggest that genetic influences are most important when female fertility decisions are taken in environments in which social norms empower women, allowing them to access a relatively broad range of life-course alternatives. However relevant genetic influences can be to human reproductive behavior, there is nothing genetically deterministic and inevitable about fertility phenotypes. Evolving more rapidly than genes, social norms, technology, and culture may create novel environments that expose genes to new selective pressures (Richerson et al., 2010), which enhance female fertility fitness.

6.2 Family-Based Models

The results above demonstrate that our PGSs identify genetic susceptibilities that heighten women's fitness. They also document the interplay between genetic factors and environments which improves female reproductive outcomes. As the environment becomes more conducive to female emancipation, so are genetic predispositions to fertility capable to exert their influences more fully. So far, our measures of environmental risk have focused on aggregate external forces, which women were exposed to at birth or while growing up, and which have been shown to be relevant to female fertility.

Fertility decisions, however, could also be shaped by the environment women face within their own family of origin. This internal environment may be affected by physical risk factors (e.g., parental resources, parenting styles, housing, and neighborhoods) as well as parental genotypes that are not genetically transmitted from parents to offspring, but might yet affect child phenotypes, a phenomenon known as "genetic nurture" (Wolf et al., 1998; Kong et al., 2018; Cawley et al. 2020; Houmark et al., 2020). This speaks to the economic literature on parents' human and nonhuman capital investment in children, where parents compensate or reinforce child endowments depending both on the technological properties of the production function that links parental inputs to child outcomes and on parental preferences, such as their aversion to inequity (Behrman et al., 1982; Becker and Tomes, 1986; Cunha and Heckman, 2007; Cunha et al., 2010).

Family-based models can be used to account for shared internal environmental risks. The richness of the UK Biobank data enables us to perform two different exercises. In the first, we estimate sister

[42]We also explored the possibility of heterogeneous gene-environment responses by intensity of the educational attainment and risk tolerance polygenic scores. Although we find little heterogeneity across the distribution of each of the two scores, the findings illustrated above are confirmed. For the sake of space concerns, however, they are not reported but can be obtained from the authors.

fixed effects (SFE) models, which have been extensively used in economic research on fertility (e.g., Geronimus and Korenman, 1992; Rosenzweig and Wolpin, 1995). By comparing siblings who grew up in similar (internal and external) environments, these models allow us to assess the importance of gene-environment interactions, even when the influence of both genetic and social transmissions is heavily restrained (Belsky et al., 2018) or when parents shape their offspring's outcomes by responding to their genetic endowment differentials (Behrman et al., 1982; ; Del Bono et al., 2012; Fletcher et al., 2020). In the second exercise, we look at mother-daughter comparisons, analyzing intergenerational correlations in outcomes while controlling for polygenic scores and re-estimating our baseline specification after controlling for maternal genotypes. In this case, genetic endowments continue to be similar as in the previous exercise (although in a way that differs from the SFE model), but the environmental risks are likely to be substantially heightened.[43]

Sister Comparisons — The first exercise is performed on a subsample of approximately 12,000 biological sisters, only one of whom is included in the baseline sample to limit issues of overfitting. For each woman i in family j born (or resident) in area a, we estimate:

$$Y_{ija} = \rho_0 + \rho_1 G^Y_{ija} + \rho_2 E^Y_{ija} + \rho_3 \left(G^Y_{ija} \times E^Y_{ija}\right) + \mathbf{X}'_{ija}\Psi + \delta_a + \mu_j + \zeta_{ija}, \tag{7}$$

where μ_j measures the time-invariant family unobserved components shared among sisters.[44] The SFE estimates are summarized in Table 3. The genetic influences on outcomes go in the same direction as those found in the benchmark analysis and remain statistically significant, although they are quantitatively more modest. The drop in genetic penetrance is as small as 0.5% in the case of childlessness and as large as 20% in the case of age at first birth. That ρ_1 is smaller than θ_1 is not surprising, since on average one-half of the SNPs inherited from parents are shared among sisters.

Conversely, the estimates of the environmental effect are of the same sign and statistical significance as those in Table 2, and about 15-25% higher (in absolute value) in the cases of A1B and A1S, despite the high observed correlation in E in this subsample. However similar the external environments faced by sisters might be, small differences in such environments are associated with important differences in outcomes: women who grew up in milieux with greater pill exposure have their first sexual intercourse earlier, postpone their first birth further, are more likely to be childless, and end up with fewer children. This reaffirms the key role played by our measure of environmental risk.

Finally, the estimated gene-environment interactions are quantitatively similar to the benchmark case, except for completed fertility where $G \times E$ becomes statistically indistinguishable from zero, perhaps reflecting the decline in genetic penetrance. The interactions nonetheless continue to amplify the increase in the age at first birth and the reduction in the age at sexual debut associated with a greater pill usage (although only at the 10% level of significance in this latter case), while they

[43] In this case, in fact, the correlation in E for mothers and daughters is 0.229 (s.e.=0.048), whereas in the sisters' subsample the same correlation is 0.739 (s.e.=0.042).

[44] Despite the fact that, in every comparison, the older sister is typically (but not always) exposed to a lower pill diffusion, including 5-year birth cohort fixed effects in $\mathbf{X}$ does not alter the results displayed in Table 3. For the sake of brevity, therefore, these estimates are not shown.

attenuate the decline in childlessness induced by the pill diffusion. As in the full sample, therefore, genetic susceptibilities to fertility tend to express themselves more distinctly when environmental risks (e.g., societal cultural norms) are more favorable to female emancipation, even in this highly homogeneous sample of sisters.

Mother-Daughter Comparisons — In the UK Biobank subsample of about 1,800 mother-daughter dyads, daughters were born on average around 1965 and mothers in 1942. This analysis therefore leverages considerable variation in the external environment but, as in the previous exercise, keeps the variation in genetic influences limited. We look at mother-daughter comparisons from two standpoints. One is the traditional perspective of intergenerational correlations; the other explores a possible channel of influence through genetic nurture and re-estimates the baseline specification (5) for daughters controlling for maternal genotypes.

The results are in Table 4, which also reports the level estimates found on the same subsample of daughters used in this specific analysis (columns (a), (d), and (g)). We obviously cannot analyze childlessness intergenerationally. From the level regressions we detect no statistically significant $G\times E$ effect. This null result may be partly driven by the substantially smaller statistical power of the mother-daughter subsample compared to the full sample used in the benchmark analysis. On average, for all three outcomes, standard errors around the $G\times E$ estimates are 30 times greater in the mother-daughter subsample, whereas the size of the benchmark sample is at least 80, 100, and 220 times bigger for A1S, A1B, and CF, respectively.

Moving to the estimates on outcome transmission reported in columns (c), (f), and (i), four findings are worth stressing. First, genetic penetrance is as high as for entire sample. This is remarkable not only because of the smaller mother-daughter subsample, but also because mothers and daughters share half of their genetic material. Second, the estimates of the external environmental influence retain statistical significance for A1B and CF, and only the impact of E on A1S is significant at the 10% level. This is also remarkable, as it emphasizes that mothers and daughters in the UK Biobank have a considerably different exposure to pill usage. Third, the $G\times E$ effects are negligible across all outcomes, confirming the above mentioned level results. Fourth, the intergenerational correlation of fertility outcomes is always quantitatively small and statistically indistinguishable from zero.[45]

Taken together, these results provide a few new insights. They suggest that the strength of the intergenerational correlations in fertility phenotypes has been probably overstated in previous demographic research (e.g., Murphy and Wang, 2001; Kolk et al., 2014). Of course, this must be interpreted against the likely low power of the small mother-daughter subsamples. But in as much as maternal outcomes reflect both environmental and genetic effects, their predictive power is weakened once environment and genes are explicitly taken into account. In fact, the dominant role now played by genes gives an indication of the importance of genetic nurture, which can work through intrafamilial socialization processes, parenting habits, and role models (Bisin and Verdier, 2011; Alesina and Giuliano, 2014; Doepke and Zilibotti, 2017). This may even substitute for the

[45]Notice, however, that the raw intergenerational correlation is sizeable and significant for all three outcomes, going from 0.20 (s.e.=0.039) for CF to 0.30 (s.e.=0.035) and 0.52 (s.e.=0.058) for A1S and A1B, respectively.

influence of gene-environment interactions, which become less relevant than in the whole population.

Differently from the abundant literature on the intergenerational transmission of reproductive outcomes, there is no work on the relationship between maternal genotypes and daughters' fertility outcomes. Besides the fact that our PGSs measure biological factors that enhance reproductive fitness, the score-outcome relationship could also reflect correlations between genetic factors and environments that promote fertility fitness. For example, some of the genetic components driving the score for A1B might affect parents' parenting skills that encourage a delay in the timing of motherhood, even if those components do not affect the daughter's predisposition to become a mother. Since the genotypes of individuals are necessarily correlated with the genotypes of their birth parents, this form of genetic nurture could generate a relationship between a woman's score and her fitness that works purely through intrafamily environmental factors, e.g., internal socialization, family norms, role models, parenting habits, and cultural transmission.

To account for such factors, we re-estimate the baseline specification for daughters controlling for parental genotypes, since the genotype of a child is randomly assigned conditional on parental genes (Papageorge and Thom, 2020). The results in columns (b), (e), and (h) of Table 4 show that the estimates of the daughter's own polygenic score, environment, and $G \times E$ are very close to those found when we control for maternal outcomes. Except for completed fertility, the other two outcomes are associated with maternal genotypes (although the correlation with A1B is only significant at the 10% level), with mother's and daughter's PGSs affecting the outcomes in the same direction. A one standard deviation increase in the maternal genotype implies an additional delay of 3 months for the age at first birth and a further postponement of 5 months for the age at sexual debut.[46] These results confirm that genetic nurturing effects are likely. They also suggest that, within families, genetic nurturing may be the process that replaces gene-environment interactions. Operating through these internal-to-the-family environmental channels, nontransmitted alleles have an added effect on reproductive fitness.

6.3 Simulations

We perform two sets of counterfactual experiments using the baseline estimates reported in Table 2. In both exercises, for the sake of conciseness, we only focus on two outcomes, age at sexual initiation and age at first birth.

Equal E or Equal G? — This first experiment is designed to assess the importance of E and G in a world where all women face either the same external environment or the same genetic endowment. In the first case, we simulate the outcomes using the estimated baseline model parameters but force E to be at the maximum value observed in each local authority of birth over the entire period. Albeit unrealistic, this scenario is meant to inform us of the impact of real-life policy interventions that aim to achieve a socially progressive milieu for all women (represented in our case by full access to the oral contraceptive pill). In the second scenario, we repeat the exercise but now assign to each

[46] Repeating this analysis also for CLN, we find that a one standard deviation increase in G^{mother} is significantly associated with a 13 percentage point increase in the probability that the daughter is childless by age 45.

woman the minimum (or maximum) observed value of G for A1S (A1B) by birth cohort.[47] Although even less realistic than the previous exercise, this scenario reveals the role played by pill exposure over and beyond the impact of genetic predisposition, when this is kept equal for all women of a given cohort as if they were all twin sisters. In both exercises, gene-environment interactions are left to operate freely.

The results are reported in Figure 6, which also shows the trends in A1S and A1B observed in the data (as reported in Figure 1) as well as the trends predicted using the baseline estimates. We stress three findings. First, despite its simplicity, our benchmark model with just one polygenic score and one proxy for environmental risks does extremely well in fitting the data. Second, exposure to full access to the pill leads to a large and immediate reduction in A1S of about 2 years for women in the early cohorts of the UK Biobank, and an equally large and immediate increase of 3 years in A1B for early cohorts again (see the dashed lines corresponding to 'E max'). Since genes evolve less rapidly than the environment over time and the environment changes little in the experiment throughout the sample period, both changes are also permanent, i.e., they affect early cohorts as much as they do later cohorts. In the case of age at sexual debut, the further reduction for the post-1945 cohorts is due to the switching-off of the WWII dummy variable, while the slow increase from the cohorts born after 1950 is driven by the interaction between G and E.

Third, in a world where all women of a given cohort are monozygotic twin sisters and their genetic endowment is assumed to be at the extremes of the polygenic score distributions, the impacts on A1S and A1B are massive, implying at least a 4-year anticipation of the first sexual intercourse and a 6-year postponement of the first birth (see the lines corresponding to 'G min' and 'G max', respectively in the top and bottom panels of Figure 6). Moving away from the extremes of the PGS distributions will clearly lead to smaller effects. What matters to us, however, is not the size of the impact in this case. A key feature of this simulation is that the environment continues to influence fertility behavior substantially, even when G exerts the same maximal impact for all the women of a given cohort. This is due to differences in environmental risk over time. For all the cohorts of women born after the end of WWII and up to the end of the period, i.e., those increasingly exposed to the diffusion of the pill, A1S declines and A1B rises by about 2 and 3 years, respectively, and the size of these changes is independent of the assumed value of G.

G×E Effects by Strength of Genetic Predisposition — Do gene-environment interactions vary depending on the intensity of genetic penetrance? To address this question, we first re-estimate the baseline model (5) after splitting the sample into two groups of women based on their polygenic scores, one above and the other below the median. We then simulate each outcome for the two groups, labelled 'high G' and 'low G', and plot them in Figure 7 with a thick continuous line and a thick dotted line, respectively. The figure also shows the results from another counterfactual experiment in which the $G \times E$ effects of this model are set to zero.

[47] We focus on the bottom PGS for A1S, because this allows us to see the impacts of E and G on Y go in the same direction. More precisely, we assign the values observed in the bottom (top) percentile of the A1S (A1B) polygenic score for each birth cohort. Repeating the same exercise at higher (lower) percentiles yields smaller level changes, but the results are otherwise identical to those discussed in the paper and are thus not reported.

We highlight three results. First, the impact of gene-environment interactions is imperceptible for women born before 1950, but it gets stronger over time, especially among women born after 1960. This confirms the importance of environmentally mediated genetic effects, which emerge precisely when the environmental conditions expand the range of opportunities for women through the diffusion of the oral contraceptive pill. Second, $G \times E$ effects attenuate the differences in A1S and magnify those in A1B between women with high and low genetic susceptibilities. This supports the baseline estimates of Table 2.

Third, the inclusion of gene-environment interactions clearly illustrates that the anticipation of sexual debut and the postponement of motherhood are greatest among women with higher polygenic scores. In the pre-pill period, anticipating sexual initiation is likely to be most costly for these women, especially if early sex heightens the risk of an undesired pregnancy. But the advent of an almost infallible contraceptive method enables the same women to engage in early sexual activities while avoiding the penalty of abstinence. Interestingly, such stark $G \times E$ effects emerge only when the external environment permits women's genetic predisposition to express itself.

7. Conclusion

Using unique data from the UK Biobank, we study how genes (G) and environment (E) are linked to female fertility. Rather than removing genetic endowments, through twinning for instance, we measure them using one of the largest genomic resources in the world and exploiting the latest developments in molecular genetics. They are summarized by a linear index called polygenic score (PGS). The socioeconomic environment is captured by the diffusion of the oral contraceptive pill in the area where women grew up, an indicator of the advent of gender-egalitarian norms and women's empowerment. We analyze four fertility outcomes over the entire female reproductive cycle.

Six new results are worth stressing. First, both G and E are strong predictors of all four traits. In particular, greater genetic susceptibilities (i.e., higher values of the polygenic scores) are associated with an increase in the ages at first sexual intercourse and first birth and in the total number of children, and with a decline in the likelihood of childlessness. Similarly, women who grew up in areas with greater pill exposure tend to anticipate their sexual debut, postpone motherhood, have fewer children overall, and a greater probability of not having children at all.

Second, these results are broadly robust across several sensitivity exercises. For example, they emerge also if E is proxied by female educational attainment rather than pill usage. And they persist when, in addition to the PGSs specific to our outcomes, we control for the polygenic scores for educational attainment and risky behaviors, which may pick up other unobservables relevant to our fertility phenotypes.

Third, a key contribution of our analysis is the identification of gene-environment interactions which, albeit acknowledged by several scholars (e.g., Heckman and Mosso, 2014; Fagereng et al., 2021), have never been quantified before for women's fertility. Both the anticipation of sexual debut and the postponement of motherhood led by the diffusion of the pill are magnified by gene-environment interactions; conversely, the decline in family size and the rise in childlessness associated

with female emancipation are attenuated by $G \times E$ effects. This interplay becomes more important when social norms and economic conditions enable women to expand the set of their life-course opportunities, possibly reflecting a decline in the relative price of fertility fitness. In more egalitarian environments that empower women, genes express themselves more fully.

Fourth, there is evidence of heterogeneity of this interplay. For instance, as genetic predispositions induce a higher age at first sex, gene-environment interactions mitigate this association more for women in the top of the PGS distribution than elsewhere. For the same group of women, gene-environment interactions also reinforce the postponement of motherhood driven by G.

Fifth, most of the previous results emerge also when we restrict attention to the subsample of sisters. Genetic predispositions continue to express themselves more sharply when the environment is conducive to women's empowerment, even if both genes and environment shared by sisters are more homogenous than in the general population. Mother-daughter comparisons reveal that the strength of the intergenerational correlations in fertility outcomes is not as strong as originally thought, once genetic endowments are controlled for. The same comparisons reveal also that, within families, one possible way in which $G \times E$ effects operate is through genetic nurturing, i.e., the link between reproductive fitness and genotypes that are not genetically transmitted from parents to daughters, which can express itself through internal socialization and intrafamily cultural transmission.

Sixth, counterfactual simulations illustrate that, even in a world populated by women who are genetically identical, the impact of E on the evolution of outcomes is substantial. Instead, when all women face equal environmental exposure to pill usage, we would observe only small differences in outcomes across cohorts but large cross-sectional variation driven by different genetic endowments. Not only do gene-environment interactions moderate the differences in age at sexual initiation and reinforce those in age at first motherhood between women with high and low genetic propensities, but they also express themselves more strongly when the environmental conditions offer women a broad range of opportunities, which complement their genetic predispositions.

Several new areas for future economic research on fertility seem to be promising. One is to extend the analysis to male fertility (Dudel and Klüsener, forthcoming). This would allow us to explore gender differences and strengthen our understanding of how gene-environment interactions operate. Another direction is to examine a wider range of external environmental risks which might affect women's emancipation and career plans differently from the contraceptive pill, such as the availability of in vitro fertilization treatments (Lundborg et al., 2017) or the changing role of gender stereotypes (Bertrand, 2020). An alternative is to open up the analysis to other contextual aspects of the internal family environment, such as childhood family structure. Although this would require a credible identification strategy (Dettmer et al., 2020), it would enrich the interpretation of gene-environment interactions beyond social control mechanisms and make it possible to consider alternative explanations based on social compensation or contextual triggering (Domingue et al., 2017; Harden et al., 2019). Differences in early family structure may also be associated with epigenetic differences between sisters, as well as between unrelated women (Heijmans et al., 2008). Combining polygenic scores with information on DNA methylation could lead to completely new insights on the

gene-environment interplay.

Furthermore, new research may help us to refine our knowledge of intrafamily responses, identifying not only the effects of $G\times E$ and genetic nurturing, but also the extent to which parents reinforce or compensate for genetic differences among their children in reproductive fitness, as they appear to do in other life domains (Behrman et al. 1982; Del Bono et al., 2012; Fletcher et al., 2020). This would link up to another important strand of research investigating how assortative mating, possibly along genetic markers as well, could affect fertility outcomes and other family behaviors (Eika et al., 2019; Barban et al., 2021). Finally, expanding the analysis to populations of non-European descent will allow us to ascertain the generalizability of our results to other ethnic groups (Peterson et al., 2019).

While some of these new avenues of research could be examined with the UK Biobank, others may have to be addressed with additional data sources, including Add Health, ALSPAC, and Understanding Society. Moving this research agenda forward will crucially depend on the availability of detailed socioeconomic information with high-quality genomic data on large populations. As heralded by Schultz (1973) fifty years ago at the start of the modern economic analysis of fertility, a new dialogue between data and theory at the intersection of different scientific disciplines is needed. As then, this new dialogue has just begun and its research opportunities are abundant.

References

Abdellaoui, Abdel, David Hugh-Jones, Loic Yengo, Kathryn E. Kemper, Michel G. Nivard, Laura Veul, et al., and Peter M. Visscher. 2019. "Genetic Correlates of Social Stratification in Great Britain." *Nature Human Behaviour*, 3 (December): 1332–1342.

Adda, Jérôme, Christian Dustmann, and Katrien Stevens. 2017. "The Career Costs of Children." *Journal of Political Economy*, 125(2): 293–337.

Alesina, Alberto, and Paola Giuliano. 2014. "Family Ties." In: Philippe Aghion and Steven N. Durlauf (eds.), *Handbook of Economic Growth*, volume 2A. Amsterdam: Elsevier, pp. 177–215.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn. 2013. "On the Origins of Gender Roles: Women and the Plough." *Quarterly Journal of Economics*, 128(2): 469–530.

Allen, Naomi, Cathie Sudlow, Paul Downey, Tim Peakman, John Danesh, Paul Elliott, et. al., and Rory Collins. 2012. "UK Biobank: Current Status and What It Means for Epidemiology." *Health Policy and Technology*, 1(3): 123–126.

Almond, Douglas, Janet Currie, and Valentina Duque. 2018. "Childhood Circumstances and Adult Outcomes: Act II." *Journal of Economic Literature*, 56(4): 1360–1446.

Bailey, Martha J. 2006. "More Power to the Pill: The Impact of Contraceptive Freedom on Women's Life Cycle Labor Supply." *Quarterly Journal of Economics*, 121(1): 289–320.

Bailey, Martha J. and Brad J. Hershbein. 2018. "US Fertility Rates and Childbearing, 1800 to 2010." In: Louis P. Cain, Price V. Fishback, and Paul W. Rhode (eds.), *The Oxford Handbook of American Economic History*, volume 1. New York: Oxford University Press.

Balbo, Nicoletta, Francesco C. Billari, and Melinda Mills. 2013. "Fertility in Advanced Societies: A Review of Research." *European Journal of Population*, 29(1): 1–38.

Barban Nicola, Elisabetta De Cao, Sonia Oreffice, and Climent Quintana-Domeque. 2021. "The Effect of Education on Spousal Education: A Genetic Approach." *Labour Economics*, forthcoming.

Barban, Nicola, Rick Jansen, Ronald de Vlaming, Ahmad Vaez, Jornt J. Mandemakers, Felix C. Tropf, et. al., and Melinda C. Mills. 2016. "Genome-Wide Analysis Identifies 12 Loci Influencing Human Reproductive Behavior." *Nature Genetics*, 48(12): 1462–1472.

Barcellos, Silvia H., Leandro S. Carvalho, and Patrick Turley. 2018. "Education Can Reduce Health Differences Related to Genetic Risk of Obesity." *Proceedings of the National Academy of Sciences*, 115(42), E9765–E9772.

Barcellos, Silvia H., Leandro S. Carvalho, and Patrick Turley. 2021. "The Effect of Education on the Relationship between Genetics, Early-Life Disadvantages, and Later-Life SES." NBER Working Paper No. 28750, May.

Barro, Robert J., and Garey S. Becker. 1989. "Fertility Choice in a Model of Economic Growth." *Econometrica*, 57(2): 481–501.

Barth, Daniel, Nicholas W. Papageorge, and Kevin Thom. 2020. "Genetic Endowments and Wealth Inequality." *Journal of Political Economy*, 128(4): 1474–1522.

Baudin, Thomas, David de la Croix, and Paula E. Gobbi. 2015. "Fertility and Childlessness in the United States." *American Economic Review*, 105(6): 1852–1882.

Beauchamp, Jonathan P. 2016. "Genetic Evidence for Natural Selection in Humans in the Contemporary United States." *Proceedings of the National Academy of Sciences*, 113(28): 7774–7779.

Beauchamp, Jonathan P., David Cesarini, Magnus Johannesson, Matthijs J. H. M. van der Loos, Philipp D. Koellinger, Patrick J. F. Groenen, et al., and Nicholas A. Christakis. 2011. "Molecular Genetics and Economics." *Journal of Economic Perspectives*, 25(4): 57–82.

Becker, Gary S. 1960. "An Economic Analysis of Fertility." In: George B. Roberts (ed.), *Demographic and Economic Change in Developed Countries* for the National Bureau of Economic Research. New York, NY: Columbia University Press, pp. 209–240.

Becker, Gary S. 1964. *Human Capital*. Chicago, IL: University of Chicago Press.

Becker, Gary S. 1965. "A Theory of the Allocation of Time." *Economic Journal*, 75(299): 493–517.

Becker, Gary S. 1973. "A Theory of Marriage: Part I." *Journal of Political Economy*, 81(4): 813–846.

Becker, Gary S. 1976. "Altruism, Egoism, and Genetic Fitness: Economics and Sociobiology." *Journal of Economic Literature*, 14(3): 817–826.

Becker, Gary S. 1981. *A Treatise on the Family*. Cambridge, MA: Harvard University Press, and 1991 (enlarged edition).

Becker, Gary S., and H. Gregg Lewis. 1973. "On the Interaction between the Quantity and Quality of Children." *Journal of Political Economy*, 81(2, pt. 2): S279–S288.

Becker, Gary S., Kevin M. Murphy, and Robert F. Tamura. 1990. "Human Capital, Fertility, and Economic Growth." *Journal of Political Economy*, 98(5, pt. 2): S12–S37.

Becker, Gary S., and Nigel Tomes. 1986. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics*, 4(3, pt. 2): S1–S39.

Becker, Joel, Casper A.P. Burik, Grant Goldman, Nancy Wang, Hariharan Jayashankar, Michael Bennett, et al., and Aysu Okbay. 2021. "Resource Profile and User Guide of the Polygenic Index Repository." bioRxiv, 443158.

Behrman, Jere R., Robert A. Pollak, and Paul Taubman. 1982. "Parental Preferences and Provision for Progeny." *Journal of Political Economy*, 90(1): 52–73.

Behrman, Jere R., and Paul Taubman. 1989. "Is Schooling 'Mostly in the Genes'? Nature-Nurture Decomposition Using Data on Relatives." *Journal of Political Economy*, 97(6): 1425–1446.

Belsky Daniel W., Benjamin W. Domingue, Robbee Wedow, Louise Arseneault, Jason D. Boardman, Avshalom Caspi, et al., and Kathleen Mullan Harris. 2018. "Genetic Analysis of Social-Class Mobility in Five Longitudinal Studies." *Proceedings of the National Academy of Sciences*, 115(31): E7275–E7284.

Belsky, Daniel W., and Salomon Israel. 2014. "Integrating Genetics and Social Science: Genetic Risk Scores." *Biodemography and Social Biology*, 60(2): 137–155.

Belsky, Jay, and Michael Pluess. 2009. "Beyond Diathesis Stress: Differential Susceptibility to Environmental Influences." *Psychological Bulletin*, 135(6): 885–908.

Benjamin, Daniel J., David Cesarini, Matthijs J. H. M. van der Loos, Christopher T. Dawes, Philipp D. Koellinger, Patrik K. E. Magnusson, et al., and d Peter M. Visscher. 2012. "The Genetic Architecture of Economic and Political Preferences." *Proceedings of the National Academy of Sciences*, 109(21):8026–8031.

Benjamin, Daniel J., James O. Berger, Magnus Johannesson, Brian A. Nosek, E.-J. Wagenmakers, Richard Berk, et al., and Valen E. Johnson. 2018. "Redefine Statistical Significance." *Nature Human Behavior*, 2(January): 6–10.

Ben-Porath, Yoram. 1973. "Economic Analysis of Fertility in Israel: Point and Counterpoint." *Journal of Political Economy*, 81(2, pt. 2): S202–S233.

Berrington, Ann. 2004. "Perpetual Postponers? Women's, Men's and Couple's Fertility Intentions and Subsequent Fertility Behaviour." Population Trends, 117 (Autumn): 9–19.

Bertrand, Marianne. 2020. "Gender in the Twenty-First Century." *American Economic Association Papers and Proceedings*, 110(5): 1–24.

Bewley, J. Derek. 1997. "Seed Germination and Dormancy." *The Plant Cell*, 9(7): 1055–1066.

Bharadwaj, Prashant, Julian V. Johnsen, and Katrine V. Løken. 2014. "Smoking Bans, Maternal Smoking and Birth Outcomes." *Journal of Public Economics*, 115 (July): 72–93.

Bierut, Laura, Pietro Biroli, Titus J. Galama, and Kevin Thom. 2018. "Childhood Socioeconomic Status Moderates Genetic Predisposition for Peak Smoking." bioRxiv, 336834.

Biroli, Pietro, and Laura Zwyssig. 2021. "Moral Hazard Heterogeneity: Genes and Health Insurance Influence Smoking after a Health Shock." bioRxiv, 434163.

Bisin, Alberto, and Thierry Verdier. 2011. "The Economics of Cultural Transmission and Socialization." In: Jess Benhabib, Alberto Bisin, and Matthew O. Jackson (eds.), *Handbook of Social Economics*, volume 1A. Amsterdam: Elsevier, pp. 339–416.

Black, Sandra E., Paul J. Devereux, Petter Lundborg, and Kaveh Majlesi. 2020. "Poor Little Rich Kids? The Role of Nature versus Nurture in Wealth and Other Economic Outcomes and Behaviours." *Review of Economic Studies*, 87(4): 1683–1725.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes, 2007. "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes." *Quarterly Journal of Economics*, 122(1): 409–439.

Boelmann, Barbara, Anna Raute, and Uta Schönberg. 2021. "Wind of Change? Cultural Determinants of Maternal Labor Supply." Unpublished Manuscript.

Bolyard, Atticus, and Peter Savelyev. 2021. "Understanding the Educational Attainment Polygenic Score and its Interactions with SES in Determining Health in Young Adulthood." HCEO Working Paper No. 2021-26.

Breinholt, Asta, and Dalton Conley. 2020. "Child-Driven Parenting: Differential Early Childhood Investment by Offspring Genotype." NBER Working Paper No. 28217, December.

Browning, Martin. 1992. "Children and Household Economic Behavior." *Journal of Economic Literature*, 30(3): 1434–1475.

Bronars, Stephen G., and Jeffrey Grogger. 1995. "The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment." *American Economic Review*, 84(5): 1141–1156.

Buera, Francisco J., and Joseph P. Kaboski. 2012. "The Rise of the Service Economy." *American Economic Review*, 102(6): 2540–2569.

Campbell, Catarina D., Elizabeth L. Ogburn, Kathryn L. Lunetta, Helen N. Lyon, Matthew L. Freedman, Leif C. Groop, et al., and Joel N. Hirschhorn. 2005. "Demonstrating Stratification in a European American Population." *Nature Genetics*, 37(8): 868–872.

Caspi, Avshalom, Joseph McClay, Terrie E. Moffitt, Jonathan Mill, Judy Martin, Ian W. Craig, Alan Taylor, and Richie Poulton. 2002. "Role of Genotype in the Cycle of Violence in Maltreated Children." *Science* 297(5582): 851–854.

Caspi, Avsholom, Karen Sugden, Terrie E. Moffitt, Alan Taylor, et al., and Richie Poulton. 2003. "Influence of Life Stress on Depression: Moderation by a Polymorphism in the 5-HTT Gene." *Science*, 301(5631): 386–389.

Cawley, John, Euna Han, Jiyoon Kim, and Edward C. Norton. 2020. "Sibling Correlation in Educational Attainment: A Test of Genetic Nurture". NBER Working Paper No. 27336.

Cheesman, Rosa, Avina Hunjan, Jonathan R.I. Coleman, Yasmin Ahmadzadeh, Robert Plomin, Tom A. McAdams, Thalia C. Eley, and Gerome Breen. 2020. "Comparison of Adopted and Nonadopted Individuals Reveals Gene–Environment Interplay for Education in the UK Biobank." *Psychological Science*, 31(5): 582–591.

Choi, Shing Wan, Timothy Shin-Heng Mak, and Paul F. O'Reilly. 2020. "Tutorial: A Guide to Performing Polygenic Risk Score Analyses." *Nature Protocols*, 15 (July): 2759–2772.

Choi, Shing Wan, and Paul F. O'Reilly. 2019. "PRSice-2: Polygenic Risk Score Software for Biobank-Scale Data." *GigaScience*, 8(7): 1–6.

Coles, Melvyn G., and Marco Francesconi. 2019. "Equilibrium Search with Multiple Attributes and the Impact of Equal Opportunities for Women." *Journal of Political Economy*, 127(1): 138–162.

Conley, Dalton, and Jason M. Fletcher. 2017. *Genome Factor. What the Social Genomics Revolution Reveals About Ourselves, Our History and the Future.* Princeton, NJ: Princeton University Press.

Courtiol, Alexandre, Felix C. Tropf, and Melinda C. Mills. 2016. "When Genes and Environment Disagree: Making Sense of Trends in Recent Human Evolution." *Proceedings of the National Academy of Sciences*, 113(28): 7693–7695.

Cunha, Flavio, and James J. Heckman. 2007. "The Technology of Skill Formation." *American Economic Review Papers and Proceedings*, 97(2): 31–47.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica*, 78(3): 883–931.

Currie, Janet. 2011. "Inequality at Birth: Some Causes and Consequences." *American Economic Review*, 101(3): 1–22.

Currie, Janet, and Enrico Moretti. 2003. "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings." *Quarterly Journal of Economics*, 118(4): 1495–1532.

de la Croix, David, and Matthias Doepke. 2003. "Inequality and Growth: Why Differential Fertility Matters." *American Economic Review*, 93(4): 1091–1113.

Del Bono, Emilia, John Ermisch, and Marco Francesconi. 2012. "Intrafamily Resource Allocations: A Dynamic Structural Model of Birth Weight." *Journal of Labor Economics*, 30(3): 657–706.

Dettmer, Amanda M., James J. Heckman, Juan Pantano, Victor Ronda, and Stephen J. Suomi. 2020. "Intergenerational Effects of Early-Life Advantage: Lessons from a Primate Study." NBER Working Paper No. 27737.

DiPrete, Thomas A., Casper A.P. Burik, and Philipp D. Koellinger. 2018. "Genetic Instrumental Variable Regression: Explaining Socioeconomic and Health Outcomes in Nonexperimental Data." *Proceedings of the National Academy of Sciences*, 115(22): E4970–E4979.

Doepke, Matthias, and Fabrizio Zilibotti. 2017. "Parenting With Style: Altruism and Paternalism in Intergenerational Preference Transmission." *Econometrica*, 85(5): 1331–1371.

Domingue, Benjamin W., Hexuan Liu, Aysu Okbay, Daniel W. Belsky. 2017. "Genetic Heterogeneity in Depressive Symptoms Following the Death of a Spouse: Polygenic Score Analysis of the U.S. Health and Retirement Study." *American Journal of Psychiatry*, 174(10): 963–970.

Dudel, Christian, and Sebastian Klüsener. Forthcoming. "Male-Female Fertility Differentials Across 17 High-Income Countries: Insights from a New Data Resource." *European Journal of Population.*

Dudbridge, Frank. 2016. "Polygenic Epidemiology." *Genetic Epidemiology*, 40(4): 268–272.

Duncan, Laramie E., and Matthew C. Keller. 2011. "A Critical Review of the First 10 Years of Candidate Gene-by-Environment Interaction Research in Psychiatry." *American Journal of Psychiatry*, 168(10): 1041–1049.

Eckstein, Zvi, Michael Keane, and Osnat Lifshitz. 2019. "Career and Family Decisions: Cohorts Born 1935–1975." *Econometrica*, 87(1): 217–253.

Eckstein, Zvi, and Kenneth I. Wolpin. 1989. "Dynamic Labour Force Participation of Married Women and Endogenous Work Experience." *Review of Economic Studies*, 56(3): 375-–390.

Eika, Lasse, Magne Mogstad, and Basit Zafar. 2019. "Educational Assortative Mating and Household Income Inequality." *Journal of Political Economy*, 127(6): 2795–2835.

Eley, T.C., K Sugden, A Corsico, A M Gregory, P Sham, P McGuffin, R Plomin, and I W Craig. 2004. "Gene–Environment Interaction Analysis of Serotonin System Markers with Adolescent Depression." *Molecular Psychiatry*, 9: 908–915.

Engzell, Per, and Felix C. Tropf. 2019. "Heritability of Education Rises with Intergenerational Mobility." *Proceedings of the National Academy of Sciences*, 116(51): 25386–25388.

Esping-Andersen, Gøsta, and Francesco C. Billari. 2015. "Re-theorizing Family Demographics." *Population and Development Review*, 41(1): 1–31.

Fagereng, Andreas, Magne Mogstad, Marte Rønning. 2021. "Why Do Wealthy Parents Have Wealthy Children?" *Journal of Political Economy*, 129(3): 703–756.

Felson, Jacob. 2014. "What Can We Learn from Twin Studies? A Comprehensive Evaluation of the Equal Environments Assumption." *Social Science Research*, 43(1): 184–199.

Fernández, Raquel. 2011. "Does Culture Matter?" In: Jess Benhabib, Alberto Bisin, and Matthew O. Jackson (eds.) *Handbook of Social Economics*, Amsterdam: Elsevier, pp. 481–510.

Fernández, Raquel and Alessandra Fogli. 2009. "Culture: An Empirical Investigation of Beliefs, Work and Fertility." *American Economic Journal: Macroeconomics*, 1(1): 146–177.

Finch-Savage, William E., and Gerhard Leubner-Metzger. 2006. "Seed Dormancy and the Control of Germination." *New Phytologist*, 171(3): 501–523.

Fitzsimons, Emla, and Marcos Vera-Hernández. 2016. "Breast Feeding and the Weekend Effect: An Observational Study." *BMJ Open*, 6(7).

Fletcher, Jason M., Yuchang Wu, Zijie Zhao, and Qiongshi Lu. 2020. "The Production of Within-Family Inequality: Insights and Implications of Integrating Genetic Data." bioRxiv, 137778.

Francesconi, Marco. 2002. "A Joint Dynamic Model of Fertility and Work of Married Women." *Journal of Labor Economics*, 20(2): 336–380.

Galor, Oded, and David N. Weil, 2000. "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond." *American Economic Review*, 90(4): 806–828.

Garrett, Eilidh, Alice Reid, Kevin Schürer, and Simon Szreter. 2001. *Changing Family Size in England and Wales: Place, Class and Demography, 1891–1911.* Cambridge, UK: Cambridge University Press.

Geronimus Arline T., and Sanders Korenman. 1992. "The Socioeconomic Consequences of Teen Childbearing Reconsidered." *Quarterly Journal of Economics*, 107(4): 1187–1214.

Gillen, Ben, Erik Snowberg, and Leeat Yariv. 2019. "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study." *Journal of Political Economy*, 127(4): 1826–1863.

Gluckman, Peter D., and Mark Hanson. 2005. *The Fetal Matrix: Evolution, Development, and Disease.* Cambridge: Cambridge University Press.

Goldberger, Arthur S. 1978. "The Genetic Determination of Income: Comment." *American Economic Review*, 68(5): 960–969.

Goldberger, Arthur S. 1979. "Heritability." *Economica*, 46(184): 327-347.

Goldin, Claudia. 2006. "The Quiet Revolution That Transformed Women's Employment, Education, and Family." *American Economic Review Papers and Proceedings*, 96(2): 1–21.

Goldin, Claudia, and Lawrence F. Katz. 2002. "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions." *Journal of Political Economy*, 110(4): 730–770.

Goldscheider, Frances, Eva Bernhardt, and Trude Lappergård. 2015. "The Gender Revolution: A Framework for Understanding Changing Family and Demographic Behavior." *Population and Development Review*, 41(2): 207–239.

Greenwood, Jeremy, Ananth Seshadri, and Mehmet Yorukoglu. 2005. "Engines of Liberation." *Review of Economic Studies*, 72(1): 109–33.

Guinnane, Timothy W. 2011. "The Historical Fertility Transition: A Guide for Economists." *Journal of Economic Literature*, 49(3): 589–614.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2006. "Does Culture Affect Economic Outcomes?" *Journal of Economic Perspectives*, 20(2): 23–48.

Hakim, Catherine. 2000. *Work-Life-style Choices in the 21st Century: Preference Theory.* New York: Oxford University Press.

Harden, K. Paige, Benjamin W. Domingue, Daniel W. Belsky, Jason D. Boardman, Robert Crosnoe, Margherita Malanchini, et al., and Kathleen Mullan Harris. 2019. "Genetic Associations with Mathematics Tracking and Persistence in Secondary School." *Science of Learning*, 5(1): 1–8.

Heckman, James J. 1995. "Lessons from the Bell Curve." *Journal of Political Economy*, 103(5): 1091–1120.

Heckman, James J. 2006. "Skill Formation and the Economics of Investing in Disadvantaged Children." *Science*, 312(5782): 1900–1902.

Heckman, James J. 2007. "The Economics, Technology and Neuroscience of Human Capability Formation." *Proceedings of the National Academy of Sciences*, 104(33): 13250–13255.

Heckman, James J., and Stefano Mosso. 2014. The Economics of Human Development and Social Mobility. *Annual Review of Economics*, 6, 689–773.

Heckman, James J., and James R. Walker. 1990. "The Relationship Between Wages and Income and the Timing and Spacing of Births: Evidence from Swedish Longitudinal Data."*Econometrica*, 58(6):1411–1441.

Heckman, James J., and Robert J. Willis. 1976. "Estimation of a Stochastic Model of Reproduction: An Econometric Approach." In: Nestor E. Terleckyj (ed.), *Household Production and Consumption.* NBER: Columbia University Press, pp. 99–146.

Heijmans, Bastiaan T., Elmar W. Tobi, Aryeh D. Stein, Hein Putter, Gerard J. Blauw, Ezra S. Susser, Eline Slagboom, and L. H. Lumey. 2008. "Persistent Epigenetic Differences Associated with Prenatal Exposure to Famine in Humans." *Proceedings of the National Academy of Sciences*, 105(44): 17046–17049.

Herrnstein, Richard J., and Murray, Charles. 1994. *The Bell Curve: Intelligence and Class Structure in American Life.* New York, NY: Free Press.

Hill, W. G., and B. S. Weir. 2011. "Variation in Actual Relationship as a Consequence of Mendelian Sampling and Linkage." *Genetics Research* 93(1): 47–64.

Hirshleifer, Jack. 1977. "Economics from a Biological Viewpoint." *Journal of Law and Economics*, 20(1): 1–52.

Hobcraft, John. 1996. "Fertility in England and Wales: a Fifty-Year Perspective." *Population Studies*, 50(3): 485—524.

Hotz, V. Joseph, Jacob Alex Klerman, and Robert J. Willis. 1997. "The Economics of Fertility in Developed Countries." In: Mark R. Rosenzweig and Oded Stark (eds.), *Handbook of Population and Family Economics.* New York, NY: Elsevier, pp. 275–347.

Hotz, V. Joseph. and Robert A. Miller. 1988. "An Empirical Analysis of Life Cycle Fertility and Female Labor Supply." *Econometrica*, 56(1): 91–118.

Houmark, Mikkel Aagaard, Victor Ronda, and Michael Rosholm. 2020. "The Nurture of Nature and the Nature of Nurture: How Genes and Investments Interact in the Formation of Skills." IZA Discussion Papers, No. 13780.

Keller, Matthew C. 2014. "Gene × Environment Interaction Studies Have Not Properly Controlled for Potential Confounders: The Problem and the (Simple) Solution." *Biological Psychiatry*, 75(1): 18–24.

van Kippersluis, Hans, Pietro Biroli, Titus J. Galama, Stephanie von Hinke, S. Fleur W. Meddens, Dilnoza Muslimova, Rita Pereira, and Cornelius A. Rietveld. 2021. "Stop Meta-Analyzing, Start Instrumenting: Maximizing the Predictive Power of Polygenic Scores." bioRxiv, 439157.

Kohler, Hans-Peter, Joseph L. Rogers, and Kaare Christensen. 1999. "Is Fertility Behavior in Our Genes? Findings from a Danish Twin Study." *Population and Development Review*, 25(2): 253–288.

Kolk Martin, Daniel Cownden, and Magnus Enquist. 2014. "Correlations in Fertility across Generations: Can Low Fertility Persist?" *Proceedings of the Royal Society B*, 281(1779): 1–7.

Kong, Agustine, Gudmar Thorleifsson, Michael L. Frigge, Bjarni J. Vilhjalmsson, Alexander I. Young, Thorgeir E. Thorgeirsson, et al., and Kari Stefansson. 2018. "The Nature of Nurture: Effects of Parental Genotypes." *Science*, 359(6374): 424–428.

Lee, James J., Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzian, Meghan Zacher, et al., and David Cesarini. 2018. "Gene Discovery and Polygenic Prediction from a 1.1-Million-Person GWAS of Educational Attainment". *Nature Genetics*, 50(8): 1112–1121.

Lesthaeghe, Ron. 2010. "The Unfolding Story of the Second Demographic Transition." *Population and Development Review*, 36(2): 211–251.

Linnér, Richard K., Pietro Biroli, Edward Kong, S. Fleur W. Meddens, Robbee Wedow, Mark A. Fontana, et al., and Jonathan P. Beauchamp. 2019. "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.

Lundborg, Petter, Erik Plug, and Astrid Würtz Rasmussen. 2017. "Can Women Have Children and a Career? IV Evidence from IVF Treatments." *American Economic Review*, 107(6): 1611–1637.

Malthus, Thomas R. 1798. "An Essay on the Principle of Population." Cambridge: Cambridge University Press.

Manolio, Teri A., Francis S. Collins, Nancy J. Cox, David B. Goldstein, et al., Peter M. Visscher. 2009. "Finding the Missing Heritability of Complex Diseases." *Nature*, 461(7265): 747–753.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60(3): 531–542.

Manski, Charles F. 2011. "Genes, Eyeglasses, and Social Policy." *Journal of Economic Perspectives*, 25(4): 83–94.

Martin, Alicia R., Christopher R. Gignoux, Raymond K. Walters, Genevieve L. Wojcik, Benjamin M. Neale, Simon Gravel, et al., and Eimear E. Kenny. 2017. "Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations." *American Journal of Human Genetics*, 100(4): 635–649.

Mathieson, Iain, Felix R. Day, Nicola Barban, Felix C. Tropf, David M. Brazel, et al., and John R.B. Perry. 2020. "Genome-Wide Analysis Identifies Genetic Effects on Reproductive Success and Ongoing Natural Selection at the FADS Locus." bioRxiv, 104455.

McCrary, Justin, and Heather Royer. 2011. "The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth." *American Economic Review*, 101(1): 158–195.

Mills, Melinda C., Nicola Barban, and Felix C. Tropf. 2020 *An Introduction to Statistical Genetic Data Analysis*. Cambridge, MA: MIT Press.

Mills, Melinda C., and Felix C. Tropf. 2020. "Sociology, Genetics, and the Coming of Age of Sociogenomics." *Annual Review of Sociology*, Vol. 46: 553–581.

Mills, Melinda C., Felix C. Tropf, David M. Brazel, Natalie van Zuydam, Ahmad Vaez, et al., and Felix R. Day. 2021. "Identification of 370 Loci for Age at Onset of Sexual and Reproductive Behaviour, Highlighting Common Aetiology with Reproductive Biology, Externalizing Behaviour and Longevity." *Nature Human Behaviour*, `https://doi.org/10.1038/s41562-021-01135-3`.

Mincer, Jacob, and Solomon Polachek. 1974. "Family Investments in Human Capital: Earnings of Women." *Journal of Political Economy*, 82(2, pt. 2): 76–110.

Moffitt, Robert A. 1984. Profiles of fertility, labour supply and wages of married women: A complete life-cycle model. *Review of Economic Studies*, 51(2), 263–278.

Montgomery, Mark, and James Trussell. 1986. "Models of Marital Status and Childbearing." In: Orley C. Ashenfelter and Richard Layard (eds.), *Handbook of Labor Economics*, volume 1. North-Holland: Amsterdam, pp. 205–271.

Murphy, Michael. 1993. "The Contraceptive Pill and Women's Employment as Factors in Fertility Change in Britain 1963-1980: A Challenge to the Conventional View." *Population Studies*, 47(2): 221–243.

Murphy, Michael, and Duolao Wang. 2001. "Family-Level Continuities in Childbearing in Low-Fertility Societies." *European Journal of Population*, 17(1): 75–96.

Myers, C. K. 2017. "The Power of Abortion Policy: Reexamining the Effects of Young Women's Access to Reproductive Control." *Journal of Political Economy*, 125(6), 2178–2224.

Myrskylä, Mikko, Hans-Peter Kohler, and Francesco C. Billari. 2009. "Advances in Development Reverse Fertility Declines." *Nature*, 460(7256): 741–743.

Nei, Masatoshi, and Wen-Hsiung Li. 1973. "Linkage Disequilibrium in Subdivided Populations." *Genetics*, 75(1): 213–219.

O'Connor, Luke J., Armin P. Schoech, Farhad Hormozdiari, Steven Gazal, Nick Patterson, Alkes L.Price. 2019. "Extreme Polygenicity of Complex Traits Is Explained by Negative Selection." *American Journal of Human Genetics*, 105(3): 456–476.

Okbay, Aysu, Jonathan P. Beauchamp, Mark A. Fontana, James J. Lee, Tune H. Pers, Cornelius A. Rietveld, et al., and Daniel J. Benjamin. 2016. "Genome-Wide Association Study Identifies 74 Loci Associated with Educational Attainment." *Nature*, 533(7604): 539–542.

Papageorge, Nicholas W., Victor Ronda, and Yu Zheng. 2019. "The Economic Value of Breaking Bad: Misbehavior, Schooling and the Labor Market." Unpublished Manuscript.

Papageorge, Nicholas W., and Kevin Thom. 2020. "Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Survey." *Journal of the European Economic Association*, 18(3): 1351–1399.

Peterson, Roseann E., Karoline Kuchenbaecker, Raymond K. Walters, Chia-Yen Chen, Alice B. Popejoy, Sathish Periyasamy, et al., and Laramie E. Duncan. 2019. "Genome-Wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations." *Cell*, 179(3): 589–603.

Pezawas, Lukas, Andreas Meyer-Lindenberg, Emily M. Drabant, Beth A. Verchinski, Karen E. Munoz, Bhaskar S. Kolachana, et al., and Daniel R. Weinberger. 2005. "5-HTTLPR Polymorphism Impacts Human Cingulate-Amygdala Interactions: A Genetic Susceptibility Mechanism for Depression." *Nature Neuroscience*, 8: 828–834.

Plomin, Robert. 2014. "Genotype-Environment Correlation in the Era of DNA." *Behavior Genetics*, 44(6): 629–638.

Pollak, Robert A. 1969. "Conditional Demand Functions and Consumption Theory." *Quarterly Journal of Economics*, 83(1): 60–78.

Price, Alkes L., Nick J. Patterson, Robert M. Plenge, Michael E. Weinblatt, Nancy A. Shadick, and David Reich. 2006. "Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies." *Nature Genetics*, 38(8): 904–909.

Richerson, Peter J., Robert Boyd, and Joseph Henrich. 2010. "Gene-Culture Coevolution in the Age of Genomics." *Proceedings of the National Academy of Sciences*, 107(Supplement 2): 8985–8992.

Rimfeld, Kaili, Eva Krapohl, Maciej Trzaskowski, Jonathan R.I. Coleman, Saskia Selzam, Philip S. Dale, Tonu Esko, Andres Metspalu, and Robert Plomin. 2018. "Genetic Influence on Social Outcomes during and after the Soviet Era in Estonia." *Nature Human Behaviour*, 2(4): 269–275.

Ronda, Victor, Esben Agerbo, Dorthe Bleses, Preben Bo Mortensen, Anders Børglum, David M. Hougaard, et al., and Michael Rosholm. 2020. "Family Disadvantage, Gender and the Returns to Genetic Human Capital." IZA DP No. 13441, July.

Rosenquist, James Niels, Steven F. Lehrer, A. James O'Malley, Alan M. Zaslavsky, Jordan W. Smoller, and Nicholas A. Christakis. 2015. "Cohort of Birth Modifies the Association Between FTO Genotype and BMI." *Proceedings of the National Academy of Sciences*, 112(2): 354–359.

Rosenzweig, Mark R., and Kenneth I. Wolpin. 1980. "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment." *Econometrica*, 48(1): 227–240.

Rosenzweig, Mark R., and Kenneth I. Wolpin. 1995. Sisters, Siblings, and Mothers: The Effect of Teen-Age Childbearing on Birth Outcomes in a Dynamic Family Context. *Econometrica*, 63(2): 303–326.

Rosenzweig, Mark R., and Kenneth I. Wolpin. 2000. "Natural 'Natural Experiments' in Economics." *Journal of Economic Literature*, 38(4): 827–874.

Rosés, Joan R., and Nikolaus Wolf. 2018. "Regional Economic Development in Europe, 1900–2010: A Description of the Patterns." CESifo Working Paper, No. 6952.

Rutter Michael. 2006. *Genes and Behavior: Nature–Nurture Interplay Explained.* Oxford: Blackwell.

Sacerdote, Bruce (2007). "How Large Are the Effects from Changes in Family Environment? A Study of Korean American Adoptees." *Quarterly Journal of Economics*, 122(1): 119–157.

Schultz, Theodore W. 1963. *The Economic Value of Education.* New York, NY:Columbia University Press.

Schultz, Theodore W. 1973. "The Value of Children: An Economic Perspective." *Journal of Political Economy*, 81(2, pt. 2): S2–S13.

Shanahan, Michael J., and Scott M. Hofer. 2005. "Social Context in Gene-Environment Interactions: Retrospect and Prospect." *Journal of Gerontology: Series B*, 60(Special Issue 1): 65–76.

Taubman, Paul. 1976. "The Determinants of Earnings: Genetics, Family, and Other Environments; A Study of White Male Twins." *American Economic Review*, 66(5): 858–870.

Trejo, Sam, Daniel W. Belsky, Jason D. Boardman, Jeremy Freese, Kathleen Mullan Harris, Pam Herd, Kamil Sicinski, and Benjamin W. Domingue. 2018. "Schools as Moderators of Genetic Associations with Life Course Attainments: Evidence from the WLS and Add Heath." *Sociological Science*, 5 (August): 513–540.

Tropf, Felix C., Nicola Barban, Melinda C. Mills, Harold Snieder, and Jornt J. Mandemakers. 2015. "Genetic Influence on Age at First Birth of Female Twins Born in the UK, 1919–68." *Population Studies*, 69(2): 129–145.

Tropf, Felix C., S. Hong Lee, Renske M. Verweij, Gert Stulp, et al., and Melinda C. Mills. 2017. "Hidden Heritability Due to Heterogeneity Across Seven Populations." *Nature Human Behavior*, 1(10): 757–765.

Turkheimer, Eric, Andreana Haley, Mary Waldron, Brian D'Onofrio, and Irving Gottesman. 2003. "Socioeconomic Status Modifies Heritability of IQ in Young Children." *Psychological Science*, 14(6): 623–628.

Udry, J. Richard. 1996. "Biosocial Models of Low-Fertility Societies." *Population and Development Review*, 22(Supplement): 325–336.

Visscher, Peter M., William G. Hill, and Naomi R. Wray. 2008. "Heritability in the Genomics Era: Concepts and Misconceptions." *Nature Reviews Genetics*, 9(4): 255–266.

Wedow, Robbee, Meghan Zacher, Brooke M. Huibregtse, Kathleen Mullan Harris, Benjamin W. Domingue, and Jason 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.

Weiss, Yoram, and Reuben Gronau. 1981. "Expected Interruptions in Labour Force Participation and Sex-Related Differences in Earnings Growth." *Review of Economic Studies*, 48(4): 607–619.

Willis, Robert J. 1973. "A New Approach to the Economic Theory of Fertility Behavior." *Journal of Political Economy*, 81(2, pt. 2): S14—S64.

Wolf, Jason, Edmund Brodie, James Cheverud, Allen Moore, and Michael Wade. 1998. "Evolutionary Consequences of Indirect Genetic Effects." *Trends in Ecology and Evolution*, 13(2): 64–69.

Wolpin, Kenneth I. 1984. "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality." *Journal of Political Economy*, 92(5): 852–874.

Woods, Robert. 2000. *The Demography of Victorian England and Wales.* Cambridge, UK: Cambridge University Press.

Wray Naomi R., Jian Yang, Ben J. Hayes, Alkes L. Price, Mike E. Goddard, and Peter M. Visscher. 2013. "Pitfalls of Predicting Complex Traits from SNPs." *Nature Reviews Genetics*, 14(7): 507–515.

Wrigley, Edward A., and Roger S. Schofield. 1989. *The Population History of England 1541–1871: A Reconstruction.* Cambridge, UK: Cambridge University Press.

Yang Jian, S. Hong Lee, Michael E. Goddard, Peter M. Visscher. 2011. "GCTA: A Tool for Genome-wide Complex Trait Analysis." *American Journal of Human Genetics*, 88(1): 76–82.

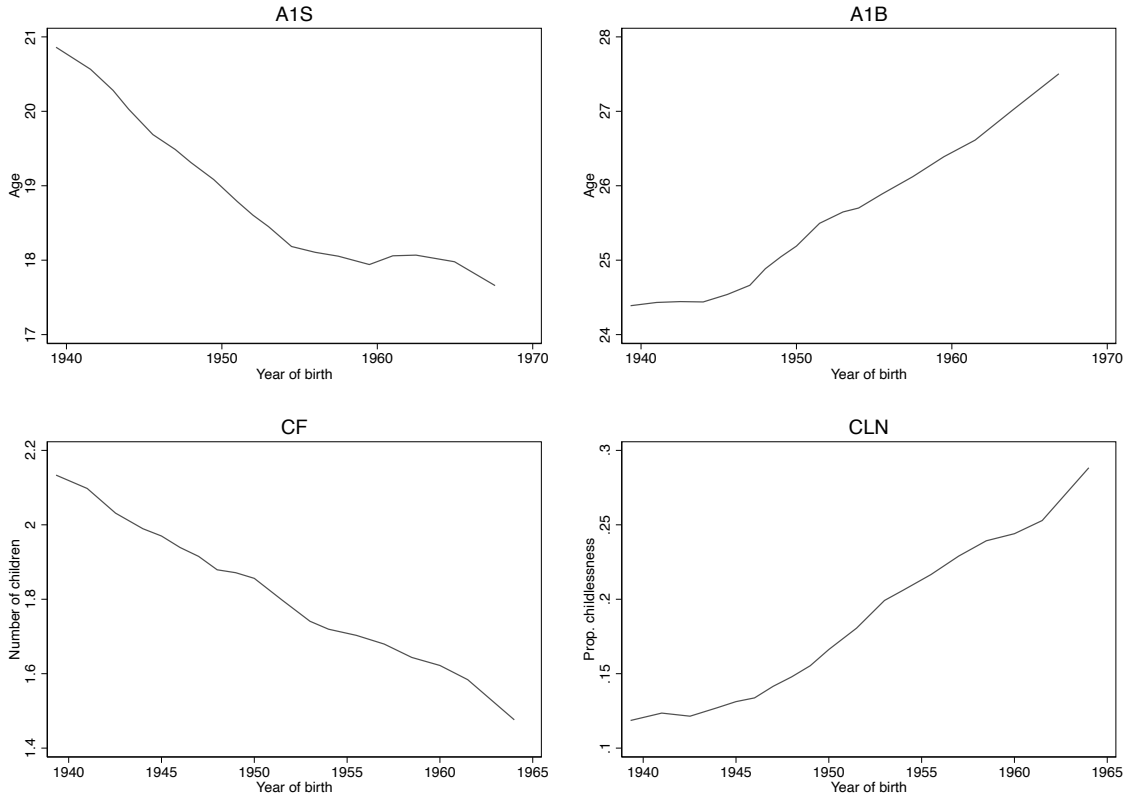
Yang, Jian, Jian Zeng, Michael E. Goddard, Naomi R. Wray, and Peter M. Visscher. 2017. "Concepts, Estimation and Interpretation of SNP-based Heritability". *Nature Genetics*, 49(9): 1304–1310.

Young, Alexander I., Michael L. Frigge, Daniel F. Gudbjartsson, Gudmar Thorleifsson, Gyda Bjornsdottir, Patrick Sulem, et al., and Augustine Kong. 2018. "Relatedness Disequilibrium Regression Estimates Heritability without Environmental Bias." *Nature Genetics*, 50(9): 1304–1310.

Zuk, Or, Eliana Hechter, Shamil R. Sunyaev, and Eric S. Lander. 2012. "The Mystery of Missing Heritability: Genetic Interactions Create Phantom Heritability." *Proceedings of the National Academy of Sciences*, 109(4): 1193–1198.

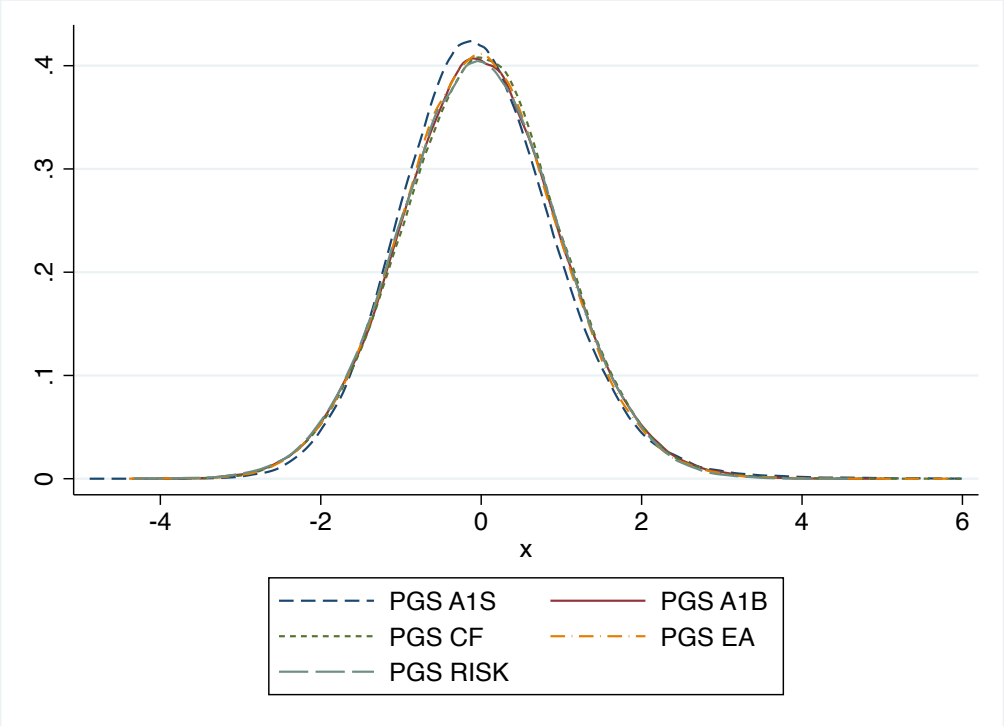
Figures and Tables

Figure 1: Fertility Outcomes by Birth Cohort in the UK Biobank



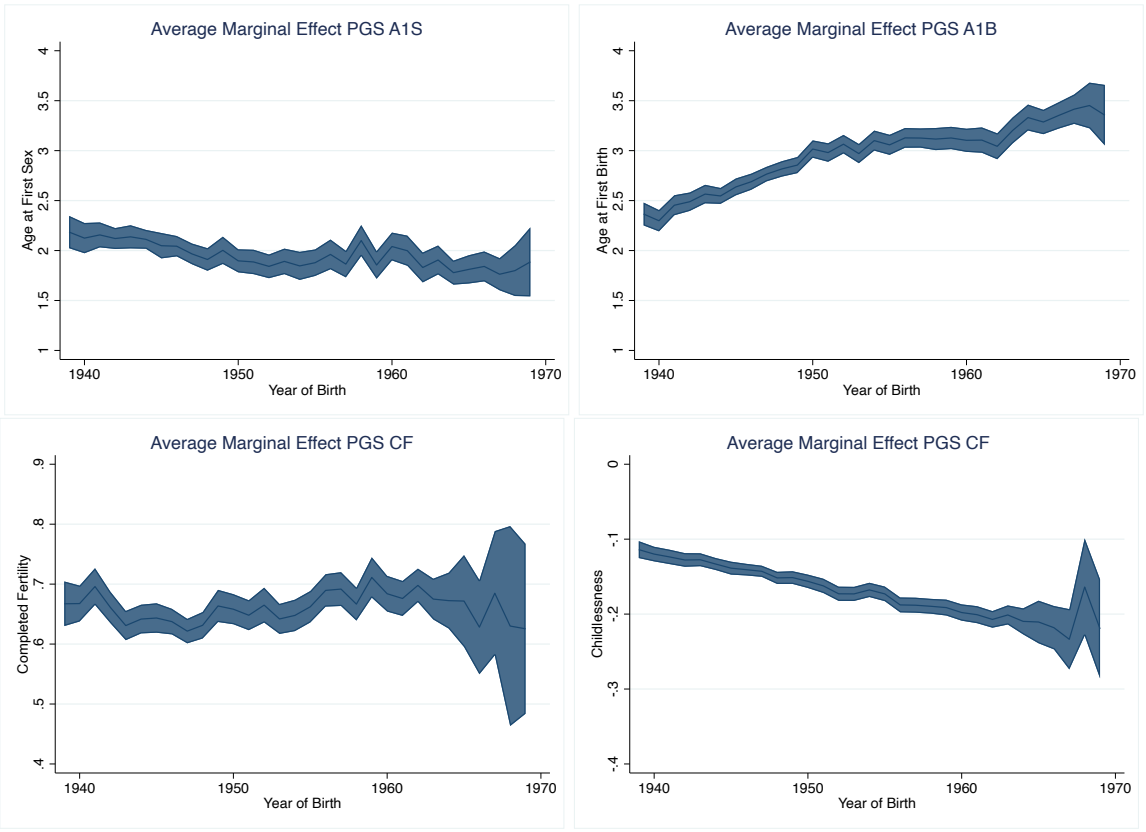
Notes: Each panel displays time trends in the outcomes: age at first sex (A1S), age at first birth (A1B), completed fertility (CF), and childlessness (CLN). Each figure is constructed with the working samples used in estimation. For CF and CLN, we restrict the sample to women aged 45 years old or more at the time of the interview.

Figure 2: Distribution of the Polygenic Scores



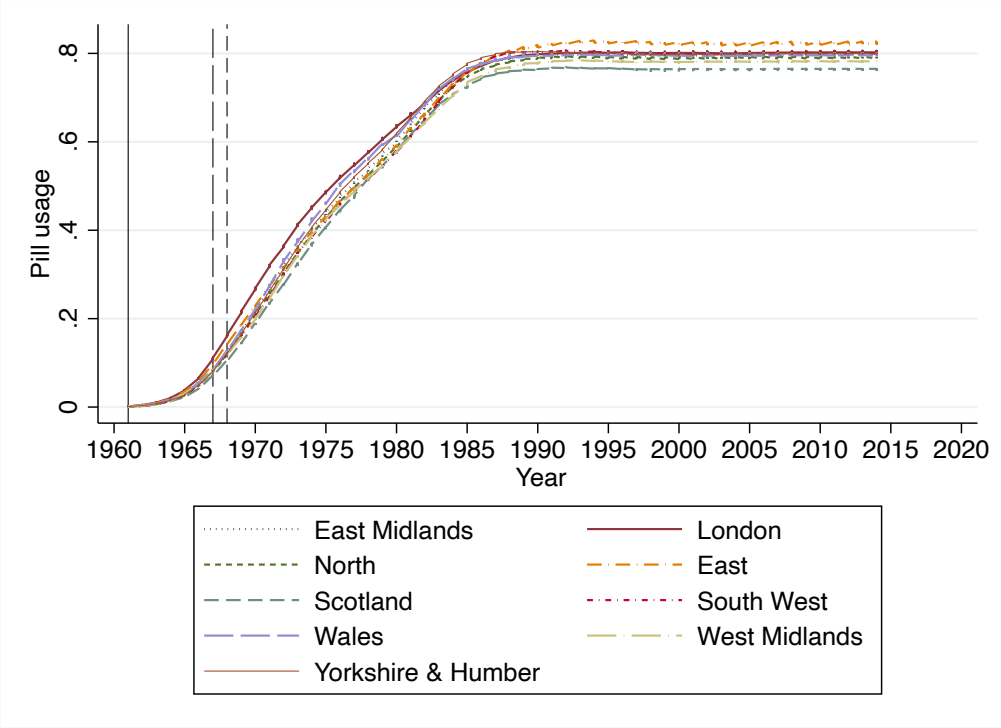
Notes: The figure plots the standardized kernel-smoothed density of the three fertility-related polygenic scores (PGS A1S, PGS A1B, and PGS CF) as well as the polygenic scores for educational attainment and risk tolerance and risky behaviors (PGS EA and PGS RISK, respectively).

Figure 3: Polygenic Scores' Marginal Effects by Outcome and Birth Cohort



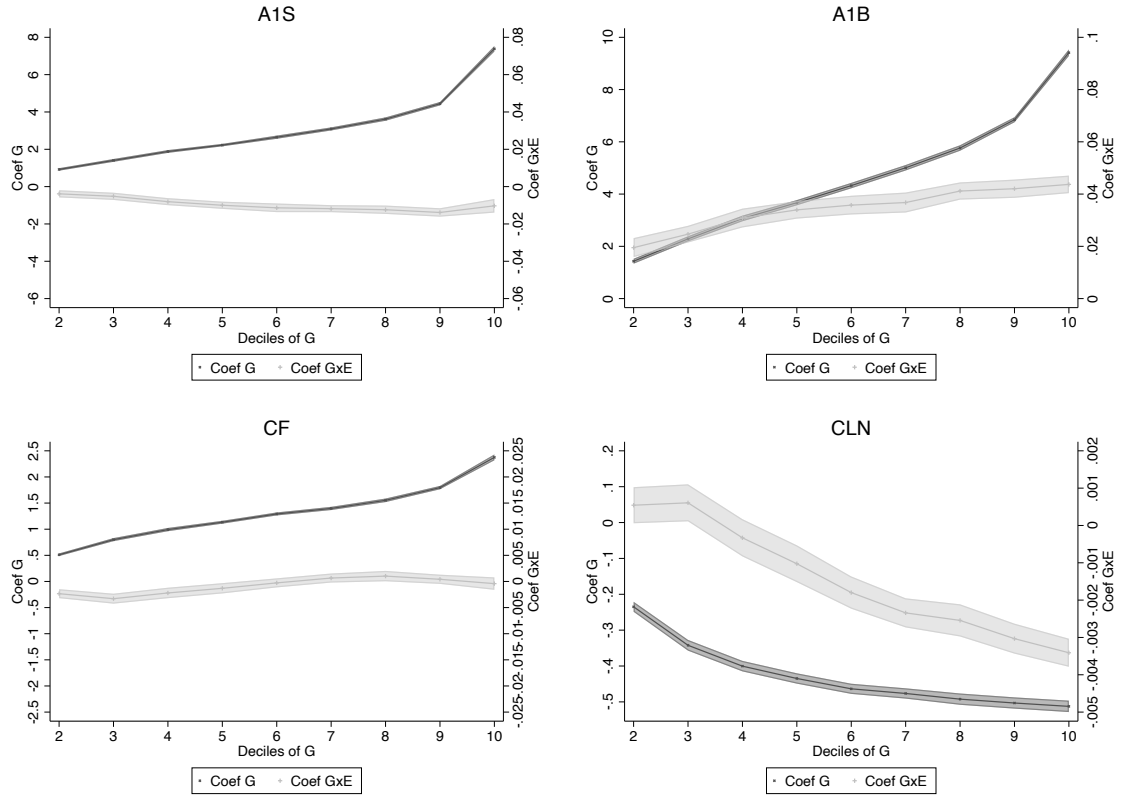
Notes: Each panel shows how the polygenic score variable for each outcome varies by cohort. The marginal effects, α_3, are obtained from equation (2). See the text for more details.

Figure 4: Contraceptive Pill Usage by Calendar Year Across Broad Geographic Regions



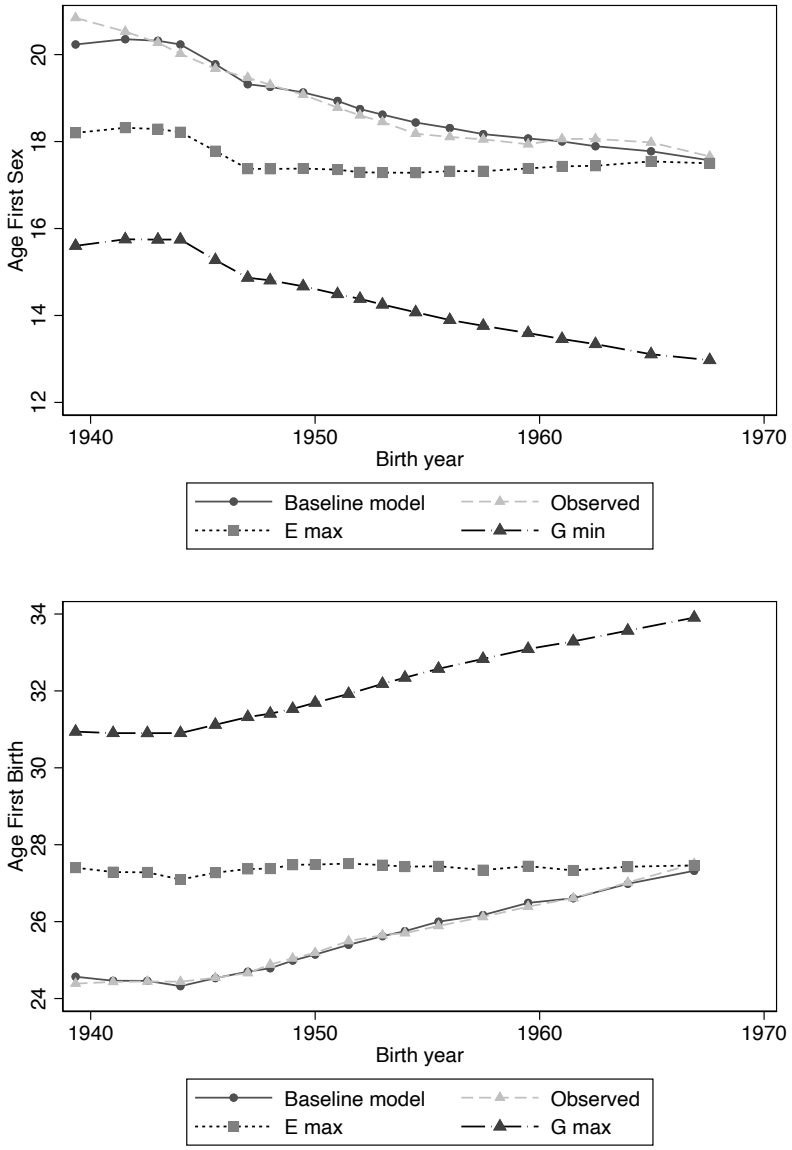
Notes: The figure shows the proportion of childless women aged 18 or more who ever took the birth control pill by broad geographic region of birth and calendar year. This proportion is constructed using data from the UK Biobank (see the text for more details). The pill was first introduced in the UK in 1961, when it was available only to married women (solid vertical line). Its availability was extended to everyone in England and Wales in 1967 (long dashed vertical line), in 1968 in Scotland (short dashed vertical line).

Figure 5: Nonlinear effects of E by deciles of G



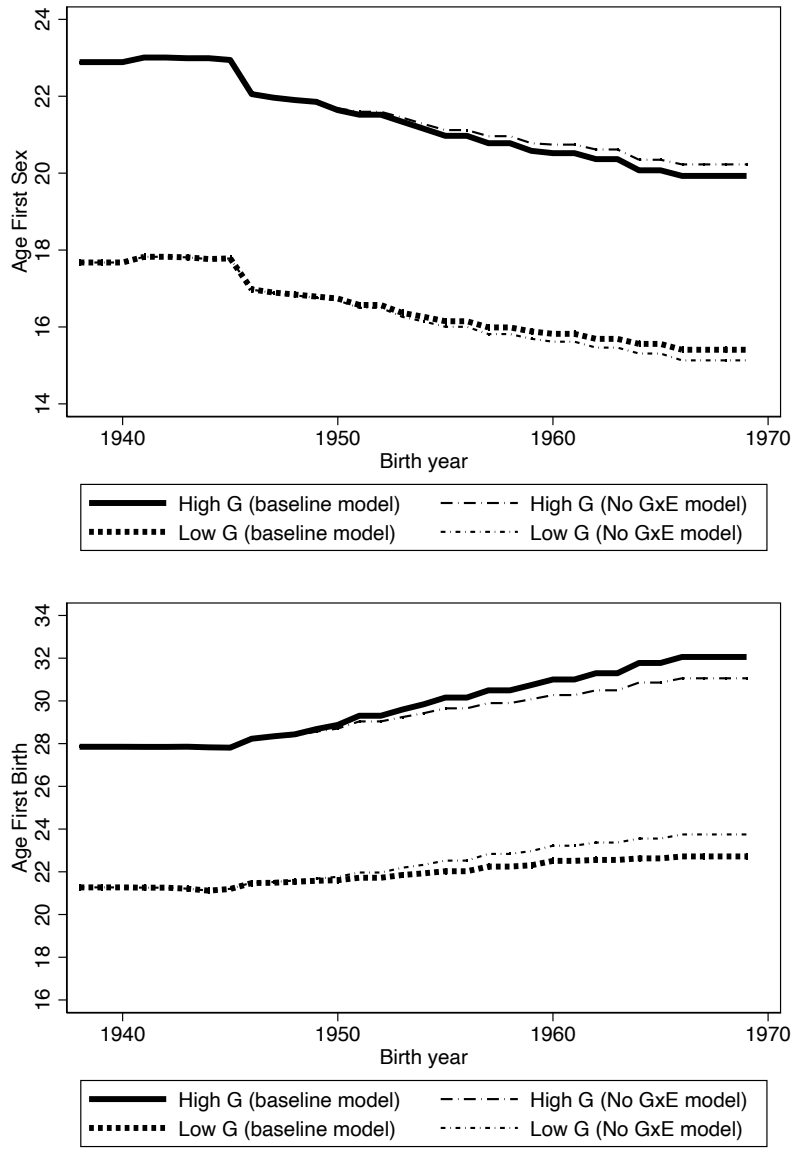
Notes: The figure shows point estimates and 95% confidence intervals for G, ϕ_1, and $G \times E$, ϕ_3, for each outcome separately. The estimates come from fitting model (6), in which each outcome is regressed on its specific polygenic score (G) discretized in deciles, the local authority district level pill usage (E), and their interactions. See the text for details on other controls included in estimation. Standard errors are robust and clustered at the local authority district level.

Figure 6: Counterfactual Simulation (I): Equal E or Equal G?



Notes: Each panel shows the observed outcome and the outcome predicted using our baseline model (5) by birth cohort. In addition, it displays the outcome predicted from the baseline model in which pill usage is forced to take the maximum value observed in each local authority of birth over the entire sample period (labelled 'E max'). It also displays the outcome predicted from the baseline model in which every woman is assigned the bottom (top) percentile of the A1S (A1B) polygenic score for each birth cohort (labelled 'G min' in the top panel, or 'G max' in the bottom panel). See the text for more details.

Figure 7: Counterfactual Simulation (II): Effects by Strength of Genetic Predisposition



Notes: Estimates in each panel are obtained from the fitting the baseline model (5) after splitting the sample into two groups of women based on their polygenic scores, one above and the other below the median. Each outcomes is then simulated for the two groups, labelled 'High G' (thick continuous line) and 'Low G' (thick dotted line), respectively. Both simulations are then replicated after setting to zero the $G \times E$ effects. These lead to the dash-dotted lines 'High G (No $G \times E$ model)' and 'Low G (No $G \times E$ model)'.

Table 1: Summary Statistics

Variables	Mean	Std. Dev.	N
Outcomes			
Age at first sex (A1S)	18.98	3.51	148,286
Age at first birth (A1B)	25.33	4.54	115,229
Completed fertility (CF)	1.84	1.16	155,435
Childlessness (CLN)	0.17	0.38	155,435
Measures of genetic assessment (G)			
PGS A1S	-0.03	0.98	168,724
PGS A1B	-0.01	0.99	168,757
PGS CF	0.01	0.99	168,754
PGS Educational attainment	-0.02	0.98	168,757
PGS Risk	-0.03	0.99	168,757
Measures of environmental influence (E)			
Pill exposure	0.24	0.25	168,757
Female education	0.71	0.22	167,117
GDP per capita	10.66	2.38	167,486
Share in services	0.51	0.08	167,486
Early life controls			
Mother smoking during pregnancy	0.30	0.46	146,532
Birthweight (kg)	3.24	0.63	109,409
Breastfed	0.69	0.46	136,488

Source: UK Biobank.
Notes: 'Pill exposure' is measured by the proportion of childless women who used the pill for the first time by age 18 in the local authority district (LAD) of birth of each woman in the sample (derived from the UK Biobank). 'Female education' corresponds to the proportion of women leaving post-compulsory schooling by LAD and year of birth (derived from the UK Biobank). 'GPD per capita' and 'Share in services' are regional variables linked to each woman in the UK Biobank when she was 18 years old. These two latter measures are from the statistical series published in the Rosés-Wolf database (see Rosés and Wolf, 2018). N refers the number of women.

Table 2: Baseline Estimates

Outcome	(a) A1S	(b) A1B	(c) CF	(d) CLN
G (θ_1)	2.1155***	2.6746***	0.6597***	-0.1381***
	(0.0269)	(0.0160)	(0.0069)	(0.0019)
E (θ_2)	-0.0259***	0.0359***	-0.0068***	0.0022***
	(0.0003)	(0.0005)	(0.0002)	(0.0001)
$G \times E$ (θ_3)	-0.0031***	0.0099***	0.0005***	-0.0012***
	(0.0005)	(0.0004)	(0.0001)	(0.0000)
Observations	148,253	115,229	155,432	155,432
R^2	0.403	0.495	0.347	0.201
Mean of Dep. Var.	18.97	25.33	1.840	0.175
SD of De. Var.	3.460	4.543	1.156	0.380
% Change E †	6.147	2.594	8.135	6.327
Incremental R^2 ‡	0.320	0.422	0.320	0.173

Notes: Obtained from the baseline model (5). E is measured by the proportion of women using the pill for the first time in the district of birth at the age at 18. G corresponds to PGS A1S in column (a), PGS A1B in column (b), and PGS CF in columns (c) and (d). Standard errors are robust and clustered at the local authority district (LAD) level. All regressions include: an indicator variable taking value 1 if a woman was born between 1939 and 1945, and 0 otherwise; the first 10 principal components of the full matrix of SNP data; their interactions with E; and LAD fixed effects. 'Observations' is the number of women used in the analysis.

† '% Change E' compares the marginal effect of E (θ_2 in (5)) setting all the other covariates to their mean values to the same coefficient found in a model in which G and $G \times E$ are set to zero ($\theta_1 = \theta_3 = 0$).

‡ 'Incremental R^2' shows the extra value of the R^2 statistic attributable to G and $G \times E$ compared to a model that does not include genes (i.e., $\theta_1 = \theta_3 = 0$).

* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3: Sisters' Fixed Effects Estimates

Outcome	(a) A1S	(b) A1B	(c) CF	(d) CLN
G (ρ_1)	1.7549***	2.1315***	0.5907***	-0.1372***
	(0.1006)	(0.1083)	(0.0245)	(0.0079)
E (ρ_2)	-0.0325***	0.0410***	-0.0064***	0.0015***
	(0.0028)	(0.0051)	(0.0010)	(0.0004)
$G \times E$ (ρ_3)	-0.0047*	0.0111***	-0.0005	-0.0007***
	(0.0027)	(0.0033)	(0.0008)	(0.0003)
Observations	11,777	9,209	12,682	12,682
R^2	0.238	0.275	0.211	0.148
No. of households	6,393	5,758	6,460	6,460
Mean of Dep. Var.	19.02	25.14	1.837	0.182
% Change E †	4.860	5.873	-6.310	-4.716
Incremental R^2 ‡	0.145	0.187	0.143	0.085

Notes: Obtained from the sister fixed effects model (7). Standard errors are robust and clustered at the household level. For the definitions of G and E and a list of the other variables included in the analysis, see the notes to Table 2.

† '% Change E' compares the marginal effect of E (ρ_2 in (7)) setting all the other covariates to their mean values to the same coefficient found in a model in which G and $G \times E$ are set to zero ($\rho_1 = \rho_3 = 0$).

‡ 'Incremental R^2' shows the extra value of the R^2 statistic attributable to G and $G \times E$ compared to a model that does not include genes (i.e., $\rho_1 = \rho_3 = 0$).

* Significant at 10%; ** significant at 5%; *** significant at 1%.

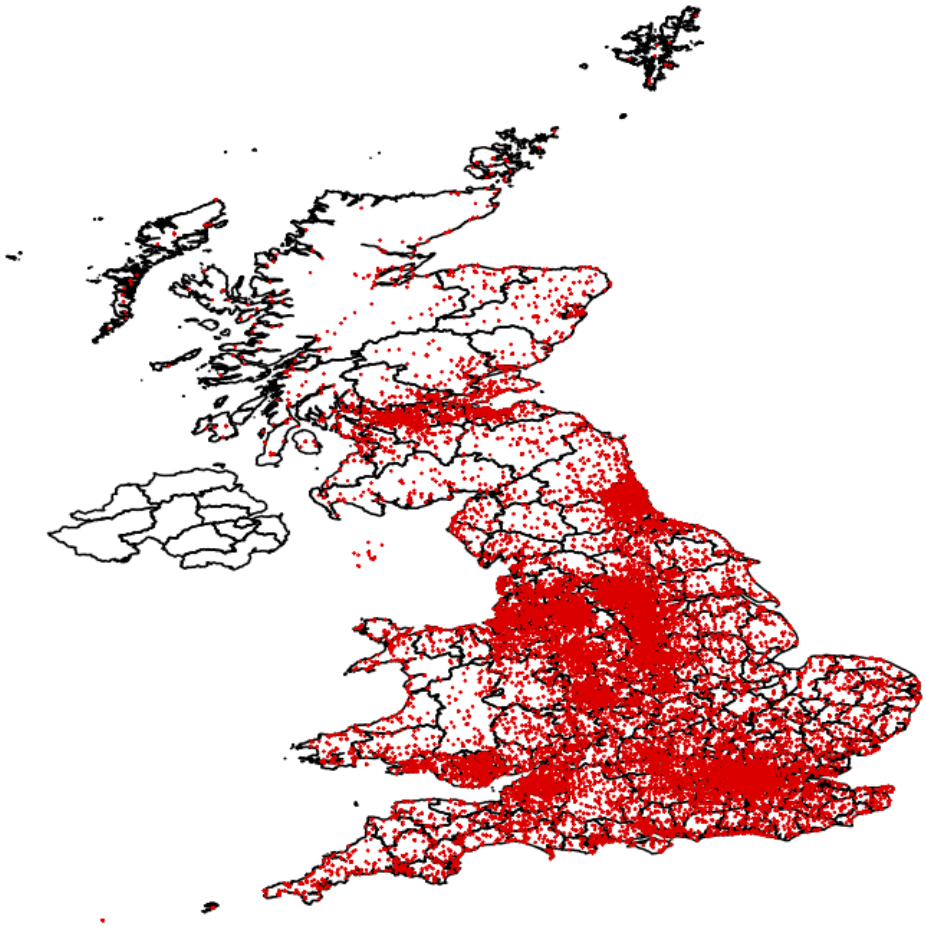
Table 4: Mother-Daughter Comparisons

Outcomes	(a) A1S	(b) A1S	(c) A1S	(d) A1B	(e) A1B	(f) A1B	(g) CF	(h) CF	(i) CF
G	2.9716***	2.8024***	3.0099***	3.1728***	3.0814***	3.4557***	0.8750***	0.8706***	0.9257***
	(1.0799)	(1.0850)	(1.1564)	(0.7607)	(0.7617)	(0.7678)	(0.2200)	(0.2244)	(0.2225)
E	-0.0180*	-0.0218**	-0.0219*	0.0468***	0.0443***	0.0436***	-0.0092**	-0.0092**	-0.0098**
	(0.0101)	(0.0100)	(0.0113)	(0.0137)	(0.0138)	(0.0144)	(0.0042)	(0.0042)	(0.0042)
$G \times E$	-0.0171	-0.0172	-0.0166	-0.0008	-0.0011	-0.0044	-0.0052	-0.0052	-0.0058*
	(0.0153)	(0.0153)	(0.0162)	(0.0110)	(0.0110)	(0.0111)	(0.0035)	(0.0035)	(0.0035)
G^{mother}		0.3834***			0.2465*			0.0062	
		(0.0797)			(0.1474)			(0.0421)	
Y^{mother}			0.0464			-0.0399			-0.0426
			(0.0300)			(0.0508)			(0.0347)
Observations	1,752	1,752	1,460	1,081	1,081	1,028	670	670	670
R^2	0.320	0.329	0.353	0.522	0.524	0.531	0.330	0.330	0.332
Mean of Dep. Var.	17.73	17.73	17.74	26.87	26.87	26.87	1.560	1.560	1.560

Notes: Obtained from the mother-daughter subsample. Standard errors are computed using the Huber-White sandwich estimator. For the definitions of G and E and a list of the other variables included in the analysis, see the notes to Table 2.
* Significant at 10%; ** significant at 5%; *** significant at 1%.

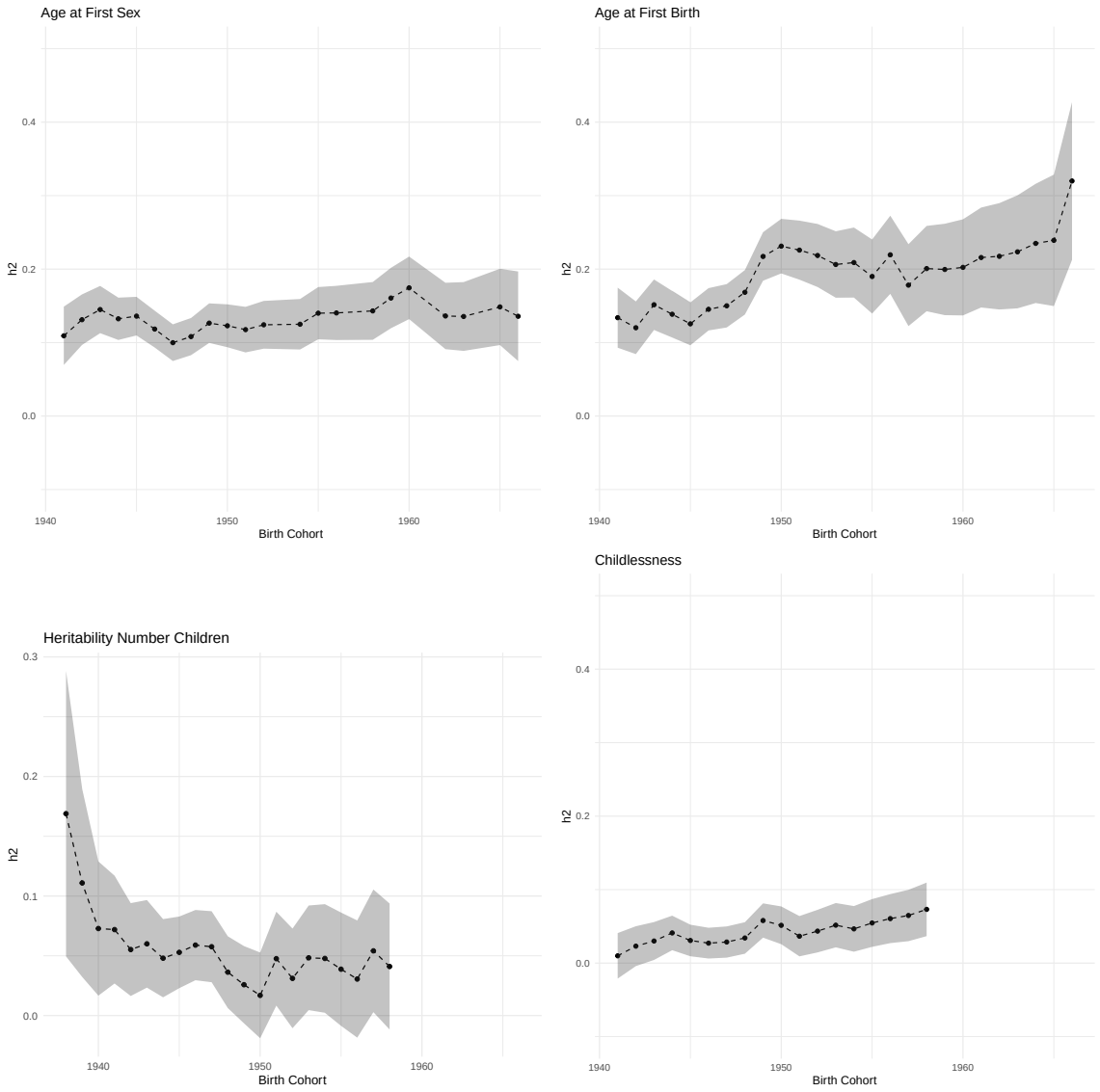
Online Appendix

Figure A1: Birth Location for Each UK Biobank Respondent



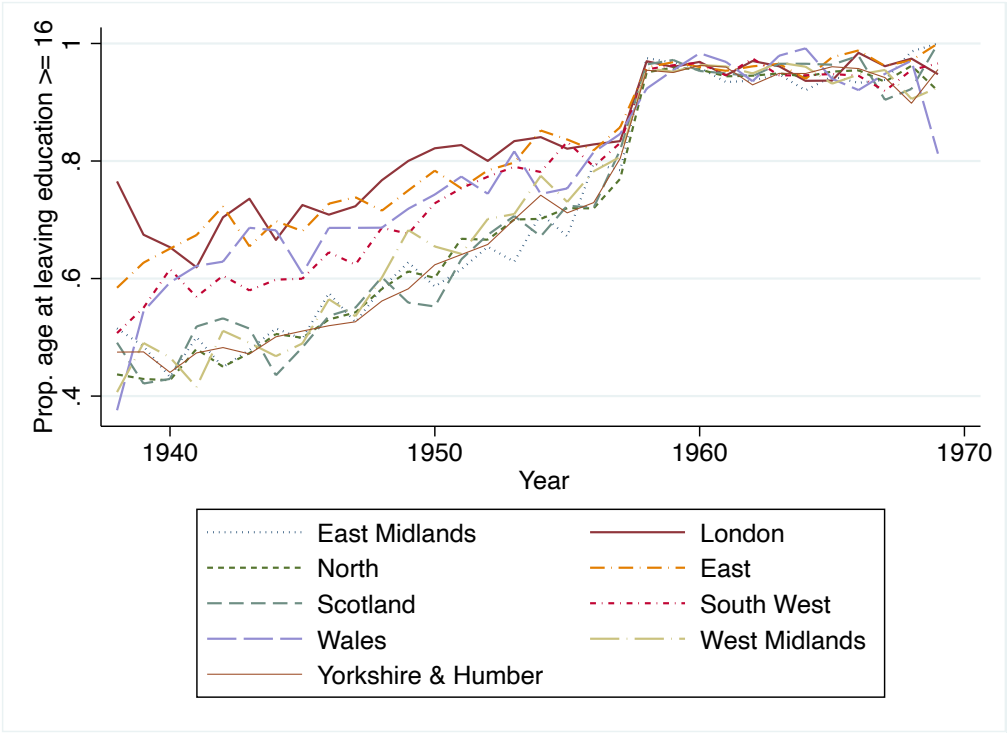
Note: Each dot in the map corresponds to the birth location of a UK Biobank respondent.

Figure A2: Heritability Trends, by Outcome and Birth Cohort



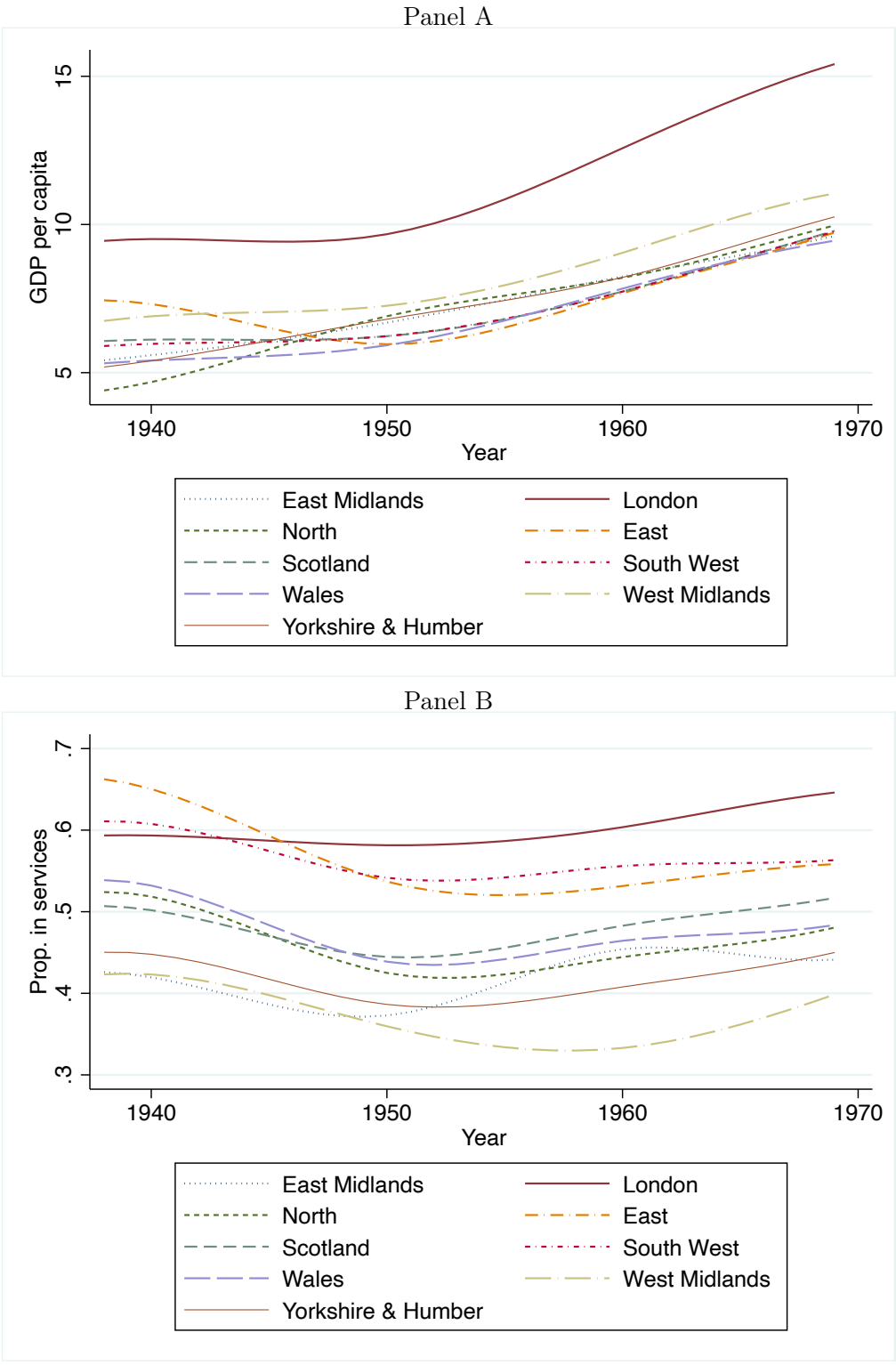
Notes: The figure shows the SNP-based heritability estimates and their 95% confidence interval by birth cohort. Estimates are obtained from a GREML model (see Section 2 in the text) and computed as 5-year moving averages.

Figure A3: Proportion of Women Leaving Post-Compulsory Education by Broad Geographic Region



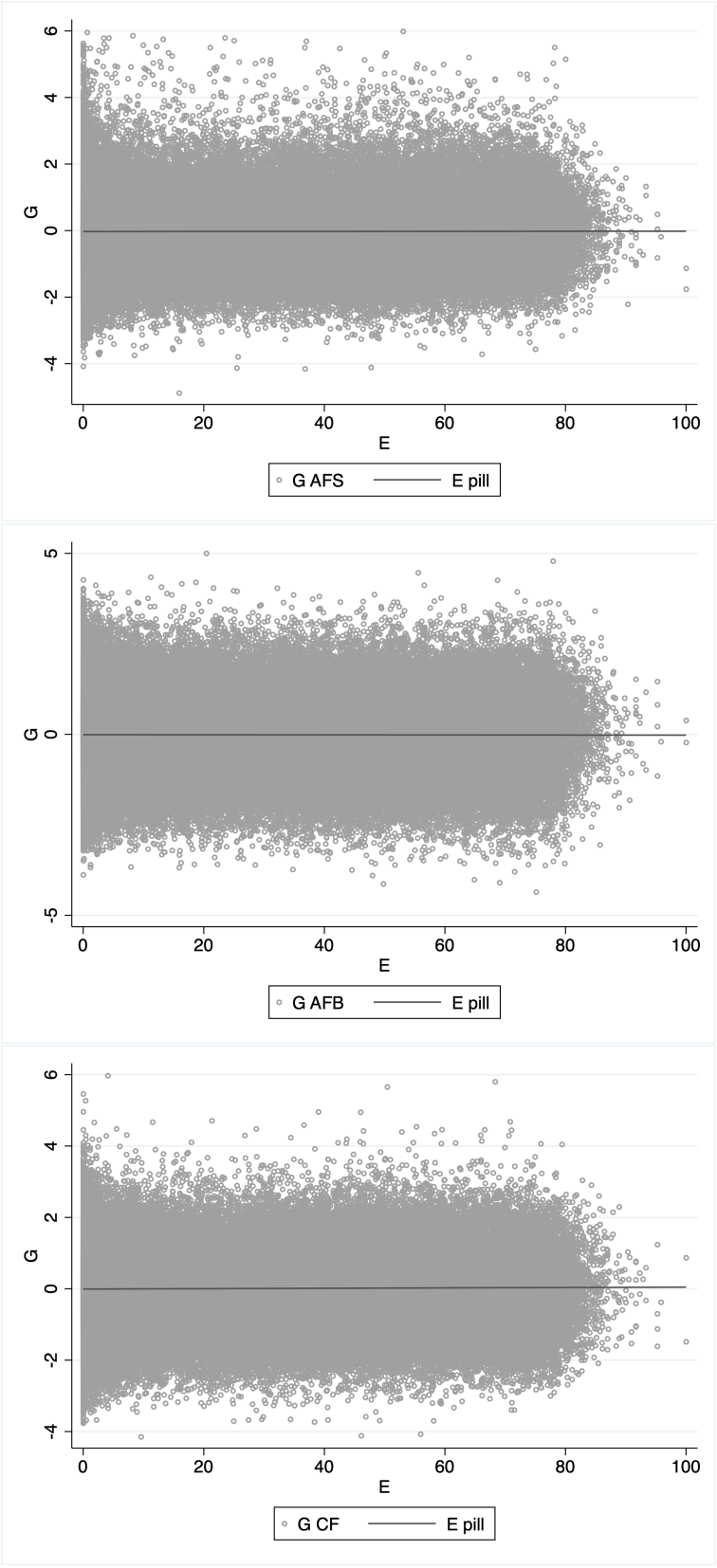
Notes: The figure shows the proportion of women leaving post-compulsory education at age 16 or more, by year and region of birth. The data come from the UK Biobank.

Figure A4: Trends in Regional GDP per Capita and Female Employment Shares in Services



Note: GDP per capita (panel A) and share of female workers employed in services (panel B) come from the statistical series published in the Rosés-Wolf database (Rosés and Wolf, 2018) and are available at a broad regional level from 1900 to 2010. We use splines to interpolate both series.

Figure A5: Correlations between G and E



Notes: The figure shows the correlation plots between each polygenic score (PGS A1S, PGS A1B, and PGS CF), as our measures of G, and pill usage (in percentage points), as our measure of E. Each panel also reports a fitted regression line.

Table A1: Cross-Correlations

Variables	A1S	A1B	CF	CLN	PGS A1S	PGS A1B	PGS CF	PGS EA	PGS RISK
A1S	1.00								
A1B	0.33	1.00							
CF	-0.12	-0.26	1.00						
CLN	0.14	.	-0.73	1.00					
PGS A1S	0.57	0.25	-0.10	0.12	1.00				
PGS A1B	0.23	0.66	-0.11	0.05	0.39	1.00			
PGS CF	-0.12	-0.14	0.56	-0.41	-0.17	-0.23	1.00		
PGS EA	0.07	0.14	-0.04	0.04	0.16	0.31	-0.12	1.00	
PGS RISK	-0.03	-0.02	0.02	-0.01	-0.05	-0.03	0.03	0.02	1.00

Notes: The table reports the cross-correlations of the four fertility outcomes and their polygenic scores (PGS). Note that the PGS for childlessness (CLN) is the same as PGS for completed fertility (CF).

Table A2: Stability of the Polygenic Scores Across Birth Cohorts

	(1) PGS A1S	(2) PGS A1B	(3) PGS CF	(4) PGS EA	(5) PGS RISK
1940–1944 (cohort 2)	-0.0309*	-0.0560***	-0.0076	0.0066	0.0068
	(0.0180)	(0.0157)	(0.0141)	(0.0125)	(0.0136)
1945–1949 (cohort 3)	-0.0574***	-0.0853***	-0.0067	0.0014	0.0054
	(0.0172)	(0.0162)	(0.0136)	(0.0116)	(0.0131)
1950–1954 (cohort 4)	-0.0970***	-0.0738***	-0.0029	-0.0000	0.0227*
	(0.0182)	(0.0156)	(0.0145)	(0.0118)	(0.0134)
1955–1959 (cohort 5)	-0.1056***	-0.0856***	0.0256*	-0.0128	0.0297**
	(0.0189)	(0.0163)	(0.0149)	(0.0120)	(0.0147)
1960–1964 (cohort 6)	-0.0364*	-0.0952***	0.0385**	-0.0353***	0.0170
	(0.0188)	(0.0165)	(0.0161)	(0.0115)	(0.0153)
1965–1969 (cohort 7)	0.0285	-0.0857***	0.0044	-0.0333**	0.0114
	(0.0192)	(0.0176)	(0.0171)	(0.0133)	(0.0154)
Observations	168,724	168,757	168,754	168,757	168,757
R^2	0.081	0.063	0.028	0.118	0.051
Mean of Dep. Var.	-0.026	-0.011	0.006	-0.025	-0.027
p-value cohort2-cohort3	0.000	0.000	0.896	0.468	0.837
p-value test cohort2-cohort4	0.000	0.0144	0.502	0.357	0.026
p-value test cohort2-cohort5	0.000	0.000	0.000	0.013	0.006
p-value test cohort2-cohort6	0.553	0.000	0.000	0.000	0.239
p-value test cohort2-cohort7	0.000	0.009	0.275	0.000	0.676
p-value test cohort3-cohort4	0.000	0.126	0.546	0.838	0.010
p-value test cohort3-cohort5	0.000	0.972	0.000	0.042	0.004
p-value test cohort3-cohort6	0.006	0.226	0.000	0.000	0.168
p-value test cohort3-cohort7	0.000	0.972	0.276	0.000	0.547
p-value test cohort4-cohort5	0.330	0.150	0.00	0.0761	0.401
p-value test cohort4-cohort6	0.000	0.009	0.000	0.000	0.519
p-value test cohort4-cohort7	0.000	0.278	0.487	0.000	0.222
p-value test cohort5-cohort6	0.000	0.275	0.196	0.003	0.132
p-value test cohort5-cohort7	0.000	0.991	0.080	0.030	0.068
p-value test cohort6-cohort7	0.000	0.418	0.004	0.851	0.613

Notes: The table presents estimates from regressions of each polygenic score on 5-years cohort indicators. All regressions include the first 10 principal components of the full matrix of the SNP data and local authority district fixed effects. The table also reports the p-value of the test for the differences between cohort coefficients. Standard errors are robust and clustered at the local authority district level.
* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table A3: Estimates Excluding $G \times E$

Outcome	(a) A1S	(b) A1B	(c) CF	(d) CLN
G	2.0388***	2.9026***	0.6706***	-0.1619***
	(0.0208)	(0.0123)	(0.0052)	(0.0016)
E	-0.0259***	0.0355***	-0.0068***	0.0022***
	(0.0004)	(0.0005)	(0.0002)	(0.0001)
Observations	148,253	115,229	155,432	155,432
R^2	0.403	0.492	0.347	0.197
Mean of Dep. Var.	18.97	25.33	1.840	0.175
Incremental R^2 ‡	0.320	0.422	0.320	0.173

Notes: Obtained from the baseline model (5). Standard errors are robust and clustered at the local authority district (LAD) level. All regressions include: an indicator variable taking value 1 if a woman was born between 1939 and 1945, and 0 otherwise; the first 10 principal components of the full matrix of SNP data; their interactions with E; and LAD fixed effects. 'Observations' is the number of women used in the analysis. See the notes to Table 2 for other details.
‡ 'Incremental R^2' shows the extra value of the R^2 statistic attributable to G compared to a model that does not include genes (i.e., $\theta_1 = \theta_3 = 0$).
* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table A4: Benchmarking Our Estimates Against Existing Studies

	(a) Goldin and Kats (2002) Birth prior to age 23	(b) Bailey (2006, 2009) Birth prior to age 22	(c) Myers (2017) Birth prior to age 19
G	-0.2236***	-0.1934***	-0.0676***
	(0.0019)	(0.0021)	(0.0014)
E	-0.0800***	-0.0663***	-0.0212***
	(0.0029)	(0.0027)	(0.0019)
$G \times E$	0.0384***	0.0356***	0.0079***
	(0.0027)	(0.0026)	(0.0023)
Abortion			0.0101***
			(0.0018)
Observations	115,229	115,229	115,229
R^2	0.267	0.234	0.095
Mean of Dep. Var.	0.283	0.216	0.0554
% Change E †	2.841	2.906	3.552
Incremental R^2 ‡	0.238	0.212	0.0860

Notes: Obtained from the estimation of model (5), in which E is measured by an indicator variable taking value one if the proportion of women using the pill for the first time in the district of birth at the age of 13 is above 30%, and zero otherwise; G is the PGS A1B. 'Abortion' is a dummy variable measuring exposure to the abortion law, which takes value one if a woman is aged 18 or less in 1968, and zero otherwise. Standard errors are robust and clustered at the local authority district (LAD) level. For other details, see the notes to Table 2.

† '% Change E' compares the marginal effect of E (θ_2 in (5)) setting all the other covariates to their mean values to the same coefficient found in a model in which G and $G \times E$ are set to zero ($\theta_1 = \theta_3 = 0$).

‡ 'Incremental R^2' shows the extra value of the R^2 statistic attributable to G and $G \times E$ compared to a model that does not include genes (i.e., $\theta_1 = \theta_3 = 0$).

* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table A5: Estimates with Early Life Controls

Outcome	(a) A1S	(b) A1B	(c) CF	(d) CLN
G	2.1113***	2.6813***	0.6669***	-0.1425***
	(0.0323)	(0.0189)	(0.0095)	(0.0022)
E	-0.0248***	0.0369***	-0.0064***	0.0021***
	(0.0004)	(0.0006)	(0.0002)	(0.0001)
$G \times E$	-0.0027***	0.0083***	0.0003	-0.0011***
	(0.0006)	(0.0005)	(0.0002)	(0.0001)
Birthweight	-0.0901***	0.0368*	0.0157***	-0.0032
	(0.0144)	(0.0210)	(0.0058)	(0.0022)
Maternal smoking	-0.4281***	-0.4500***	0.0339***	-0.0158***
	(0.0184)	(0.0299)	(0.0092)	(0.0026)
Breastfeed	-0.0326	0.1482***	0.0013	0.0054**
	(0.0205)	(0.0297)	(0.0068)	(0.0027)
Observations	81,160	60,142	80,610	80,610
R^2	0.405	0.505	0.350	0.207
Mean of Dep. Var.	18.89	25.67	1.809	0.185
% Change E	7.082	-1.541	4.823	3.900
Incremental R^2	0.317	0.409	0.322	0.178

Notes: Obtained from the estimation of model (5), after controlling for early life conditions. For all other details, see the notes to Table 2.
* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table A6: Estimates with Alternative Measures of Pill Exposure

	(a)	(b)	(c)	(d)
Outcome	A1S	A1B	CF	CLN
G	2.1422***	2.6136***	0.6580***	-0.1338***
	(0.0284)	(0.0168)	(0.0075)	(0.0020)
E	-0.0236***	0.0286***	-0.0051***	0.0017***
	(0.0003)	(0.0004)	(0.0001)	(0.0000)
$G \times E$	-0.0029***	0.0081***	0.0004***	-0.0008***
	(0.0004)	(0.0003)	(0.0001)	(0.0000)
Observations	148,248	115,226	155,427	155,427
R^2	0.405	0.493	0.346	0.201
Mean of Dep. Var.	18.97	25.33	1.840	0.175
% Change E	1.847	3.070	8.494	6.584
Incremental R^2	0.318	0.422	0.320	0.173

Notes: Obtained from the estimation of model (5), in which E is measured by the proportion of women using the pill for the first time in the district of birth between ages 18–30. For all other details, see the notes to Table 2.
* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table A7: Estimates Controlling for Year-of-Birth Fixed Effects

	(a)	(b)	(c)	(d)
Outcome	A1S	A1B	CF	CLN
G	2.1093***	2.6759***	0.6596***	-0.1381***
	(0.0257)	(0.0159)	(0.0069)	(0.0019)
E	-0.0169***	0.0293***	-0.0057***	0.0019***
	(0.0014)	(0.0025)	(0.0005)	(0.0002)
$G \times E$	-0.0031***	0.0098***	0.0005***	-0.0012***
	(0.0005)	(0.0004)	(0.0001)	(0.0000)
Observations	148,253	115,229	155,432	155,432
R^2	0.408	0.496	0.348	0.202
Mean of Dep. Var.	18.97	25.33	1.840	0.175
% Change E	-5.462	-2.214	2.657	-1.000
Incremental R^2	0.318	0.422	0.320	0.173

Notes: Obtained from the estimation of model (5), in which besides the other variables included in **X** we also control for 5-year cohort fixed effects. For all other details, see the notes to Table 2.
* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table A8: Estimates controlling for PGS risk and PGS education

Outcome	(a) A1S	(b) A1B	(c) CF	(d) CLN
G	2.1356***	2.7547***	0.6615***	-0.1382***
	(0.0273)	(0.0161)	(0.0070)	(0.0019)
E	-0.0260***	0.0357***	-0.0068***	0.0022***
	(0.0003)	(0.0005)	(0.0002)	(0.0001)
$G \times E$	-0.0031***	0.0098***	0.0005***	-0.0012***
	(0.0005)	(0.0004)	(0.0001)	(0.0000)
PGS EA	-0.1101***	-0.3158***	0.0218***	-0.0014
	(0.0083)	(0.0110)	(0.0025)	(0.0009)
PGS RISK	0.0241***	0.0532***	0.0035	0.0001
	(0.0073)	(0.0103)	(0.0027)	(0.0009)
Observations	148,253	115,229	155,432	155,432
R^2	0.404	0.499	0.347	0.201
Mean of Dep. Var.	18.97	25.33	1.840	0.175
% Change E	7.254	1.268	7.519	5.876
Incremental R^2	0.315	0.413	0.319	0.171

Notes: Obtained from the estimation of model (5), in which we also include the polygenic scores for educational attainment and for risk tolerance and risky behaviors, PGS EA and PGS RISK, respectively. For all other details, see the notes to Table 2.
* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table A9: Estimates Controlling for All Polygenic Scores

Outcome	(a) A1S	(b) A1B	(c) CF	(d) CLN
G	2.1225***	2.7939***	0.6622***	-0.1399***
	(0.0287)	(0.0181)	(0.0070)	(0.0020)
E	-0.0260***	0.0358***	-0.0068***	0.0022***
	(0.0003)	(0.0005)	(0.0002)	(0.0001)
$G \times E$	-0.0031***	0.0096***	0.0005***	-0.0012***
	(0.0005)	(0.0004)	(0.0001)	(0.0000)
PGS A1S		-0.1119***	-0.0109***	0.0258***
		(0.0146)	(0.0036)	(0.0012)
PGS A1B	0.0064		0.0142***	-0.0315***
	(0.0092)		(0.0038)	(0.0013)
PGS CF	-0.0615***	-0.0220*		
	(0.0094)	(0.0118)		
PGS EA	-0.1154***	-0.3081***	0.0200***	0.0026***
	(0.0084)	(0.0111)	(0.0025)	(0.0009)
PGS RISK	0.0252***	0.0476***	0.0033	0.0005
	(0.0073)	(0.0106)	(0.0027)	(0.0009)
Observations	148,251	115,214	155,399	155,399
R^2	0.405	0.500	0.347	0.208
Mean of Dep. Var.	18.97	25.33	1.840	0.175
% Change E	-0.246	0.322	0.0776	0.508
Incremental R^2	0	0	0	0

Notes: Obtained from the estimation of model (5), in which for each outcome we also include all the polygenic scores, not just the outcome-specific PGS. For all other details, see the notes to Table 2. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table A10: Estimates with an Alternative Measure of E (Female Educational attainment)

Outcome	(a) A1S	(b) A1B	(c) CF	(d) CLN
G	2.1214***	2.2954***	0.6516***	-0.0881***
	(0.0534)	(0.0401)	(0.0172)	(0.0047)
E	-0.0081***	0.0124***	-0.0026***	0.0007***
	(0.0012)	(0.0011)	(0.0003)	(0.0001)
$G \times E$	-0.0012*	0.0086***	0.0003	-0.0011***
	(0.0006)	(0.0006)	(0.0002)	(0.0001)
Observations	146,717	114,310	154,311	154,311
R^2	0.394	0.486	0.343	0.196
Mean of Dep. Var.	18.97	25.29	1.842	0.174
% Change E	21.20	-32.62	-5.718	-8.205
Incremental R^2	0.320	0.421	0.320	0.172

Notes: Obtained from the estimation of model (5), in which E is measured by the proportion of women leaving post-compulsory education at age 16 or above by local authority district and year of birth. All regressions also include an indicator variable taking value one if a woman is born after September 1957, and zero otherwise. For all other details, see the notes to Table 2.
* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table A11: Estimates with an Alternative Measure of E (Regional GDP Per Capita)

Outcome	(a) A1S	(b) A1B	(c) CF	(d) CLN
G	2.5393***	0.5893	0.5356***	0.1131
	(0.3626)	(0.6703)	(0.0723)	(0.0627)
E	-4.4866***	5.8792***	-1.2023***	0.3634***
	(0.5778)	(0.4680)	(0.1252)	(0.0325)
$G \times E$	-0.2117	0.9976**	0.0580*	-0.1183***
	(0.1604)	(0.2976)	(0.0293)	(0.0273)
Observations	147,161	114,398	154,230	154,230
R^2	0.401	0.490	0.345	0.196
Mean of Dep. Var.	18.98	25.33	1.841	0.175
Wildboot p-value G	0.000	0.688	0.000	0.043
Wildboot p-value E	0.000	0.000	0.000	0.000
Wildboot p-value $G \times E$	0.242	0.000	0.207	0.000
% Change E	6.146	1.662	6.129	4.666
Incremental R^2	0.324	0.436	0.322	0.173

Notes: Obtained from the estimation of model (5), in which E is measured by the regional GDP per capita linked to each woman in the UK Biobank when she was 18 years old. For all other details, see the notes to Table 2.
* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table A12: Estimates with an Alternative Measure of E (Regional Share of Female Employment in Services)

Outcome	(a) A1S	(b) A1B	(c) CF	(d) CLN
G	2.0468***	1.7658***	0.5521***	-0.0232
	(0.1519)	(0.3970)	(0.0249)	(0.0368)
E	-11.6427***	15.7420***	-3.1994***	1.0063***
	(0.6643)	(1.1015)	(0.2429)	(0.0686)
$G \times E$	-0.0098	2.2611**	0.2331***	-0.2721***
	(0.3086)	(0.7619)	(0.0466)	(0.0697)
Observations	147,161	114,398	154,230	154,230
R^2	0.3968	0.485	0.343	0.194
Mean of Dep. Var.	18.98	25.33	1.841	0.175
Wildboot p-value G	0.000	0.012	0.000	0.645
Wildboot p-value E	0.000	0.000	0.000	0.000
Wildboot p-value $G \times E$	0.957	0.000	0.023	0.000
% Change E	7.818	7.914	8.197	6.666
Incremental R^2	0.324	0.437	0.322	0.173

Notes: Obtained from the estimation of model (5), in which E is measured by the regional share of female employment in services linked to each woman in the UK Biobank when she was 18 years old. For all other details, see the notes to Table 2.
* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table A13: Estimates on the Subsample of Stayers

Outcome	(a) A1S	(b) A1B	(c) CF	(d) CLN
G	2.0901***	2.5512***	0.6705***	-0.1328***
	(0.0466)	(0.0274)	(0.0139)	(0.0039)
E	-0.0267***	0.0350***	-0.0065***	0.0020***
	(0.0006)	(0.0010)	(0.0003)	(0.0001)
$G \times E$	-0.0050***	0.0099***	0.0001	-0.0010***
	(0.0008)	(0.0006)	(0.0002)	(0.0001)
Observations	46,826	37,722	50,313	50,313
R^2	0.397	0.457	0.352	0.195
Mean of Dep. Var.	18.66	24.24	1.867	0.162
% Change E	4.276	13.69	9.267	7.826
Incremental R^2	0.304	0.406	0.321	0.163

Notes: Obtained from the estimation of model (5) fitted on the subsample of women whose local area of residence at the time of interview has not changed since birth. For all other details, see the notes to Table 2.
* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table A14: Nonlinear Effects of E

Outcome	(a) A1S	(b) A1B	(c) CF	(d) CLN
G	2.2279***	2.4865***	0.6876***	-0.1300***
	(0.0338)	(0.0252)	(0.0091)	(0.0028)
E-2	-0.3867***	0.1376**	-0.0782***	0.0009
	(0.0390)	(0.0627)	(0.0127)	(0.0043)
E-3	-0.7035***	0.2016***	-0.0974***	0.0037
	(0.0428)	(0.0422)	(0.0097)	(0.0033)
E-4	-0.9402***	0.3280***	-0.1620***	0.0118**
	(0.0495)	(0.0559)	(0.0137)	(0.0050)
E-5	-1.2419***	0.5671***	-0.1934***	0.0255***
	(0.0523)	(0.0556)	(0.0141)	(0.0054)
E-6	-1.6311***	1.0048***	-0.2692***	0.0498***
	(0.0516)	(0.0552)	(0.0153)	(0.0053)
E-7	-2.0789***	1.3326***	-0.3516***	0.0781***
	(0.0508)	(0.0608)	(0.0144)	(0.0056)
E-8	-2.3013***	1.7698***	-0.4292***	0.1081***
	(0.0481)	(0.0626)	(0.0154)	(0.0052)
E-9	-2.4227***	2.2590***	-0.5033***	0.1287***
	(0.0522)	(0.0638)	(0.0159)	(0.0061)
E-10	-2.6293***	2.8158***	-0.6177***	0.1669***
	(0.0505)	(0.0580)	(0.0195)	(0.0076)
$G \times E-2$	-0.0720	0.0228	-0.0589***	0.0009
	(0.0588)	(0.0539)	(0.0165)	(0.0057)
$G \times E-3$	-0.1182**	0.1706***	-0.0367***	-0.0072*
	(0.0461)	(0.0319)	(0.0102)	(0.0038)
$G \times E-4$	-0.1868***	0.3234***	-0.0581***	-0.0162***
	(0.0580)	(0.0401)	(0.0101)	(0.0036)
$G \times E-5$	-0.2473***	0.4157***	-0.0174	-0.0279***
	(0.0492)	(0.0386)	(0.0106)	(0.0034)
$G \times E-6$	-0.2468***	0.5351***	-0.0185	-0.0382***
	(0.0466)	(0.0392)	(0.0117)	(0.0036)
$G \times E-7$	-0.3125***	0.5996***	-0.0245**	-0.0476***
	(0.0468)	(0.0376)	(0.0102)	(0.0037)
$G \times E-8$	-0.2055***	0.6284***	0.0154	-0.0604***
	(0.0521)	(0.0347)	(0.0120)	(0.0036)
$G \times E-9$	-0.2547***	0.6839***	0.0036	-0.0733***
	(0.0484)	(0.0403)	(0.0097)	(0.0038)
$G \times E-10$	-0.3384***	0.8386***	-0.0131	-0.0819***
	(0.0426)	(0.0379)	(0.0172)	(0.0060)
Observations	148,253	115,229	155,432	155,432
R^2	0.408	0.496	0.348	0.202
Mean of Dep. Var.	18.97	25.33	1.840	0.175

Notes: Obtained from the estimation of model (5), in which our baseline measure of E (pill diffusion) has been discretized into deciles ($E-1$,..., $E-10$). Each of them has been estimated (base=$E-1$) and interacted with the outcome-specific PGS. For all other details, see the notes to Table 2.
* Significant at 10%; ** significant at 5%; *** significant at 1%.