PhD-FDEF-2021-016
The Faculty of Law, Economics and Finance

DISSERTATION

Defence held on 28/09/2021 in Esch-sur-Alzette

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN SCIENCES ÉCONOMIQUES

by

Giorgia MENTA
Born on 9 November 1992 in Catania (Italy)

ESSAYS ON HUMAN CAPITAL, INEQUALITY, AND INCOME

Dissertation defence committee

Conchita D'Ambrosio, dissertation supervisor
Professor, Université du Luxembourg

Cecilia García-Peñalosa
Professor, Aix-Marseille University
CNRS, EHESS, AMSE

Skerdilajda Zanaj, Chairman
Associate Professor, Université du Luxembourg

Marta Menéndez
Associate Professor, Université Paris-Dauphine
DIAL

Philippe Van Kerm, Vice Chairman
Professor, Université du Luxembourg

Acknowledgments

There is a saying in my hometown that goes: "*Cu nesci, arrinesci*", which roughly translates to "He/She who gets out, makes it". I am not sure I am any close to 'making it', but it is safe to say that in the past 10 years I have got out. Out of my island, first, and country, later. Out of my comfort zone, stumbling – more often than stepping – into new challenges and achievements that I thought beyond me. I wouldn't be where I am today without a fair dose of luck and, more importantly, the help of many great persons along the way. Here comes a clumsy attempt to acknowledge some of their merits.

First of all, I would like to thank my PhD advisor, Conchita D'Ambrosio. Her support and guidance have been fundamental over the past four years, and more. Without her precious encouragement, I would not have pursued a PhD. She has always been present in times of need, giving me access to opportunities and resources, all while leaving me the space to grow as an independent researcher. I am particularly grateful for the unconditional trust she has always placed in my capabilities and for the numerous chances of being involved in projects and activities beyond my PhD research. Her intellectual curiosity and far-reaching research interests have been and continue to be a source of inspiration for my own research.

I would also like to thank Philippe van Kerm, for accepting to be by my side in this journey as part of my thesis supervision committee. I am grateful to Philippe for his graceful dissent and for pushing me to see things from different angles, contributing to making this thesis significantly stronger. I am also greatful to Skerdilajda Zanaj, Cecilia Garcia Peñalosa and Marta Menéndez for kindly agreeing to be part of my PhD Jury.

Special thanks go to Andrew Clark, to whose kind 'honorary' supervision during my PhD I owe much of the researcher I am today. In a place (Luxembourg) and time (a global pandemic) of limited academic interactions, Andrew has been advisor, mentor and the most brilliant source of constructive criticism any young researcher could hope for. I am lucky to gravitate around him and to be his co-author.

I am incredibly grateful to my sponsors, Sonia Bhalotra and Emilia Del Bono, for the opportunity of visiting the Institute of Socio-Economic Research at the University of Essex during my PhD. I fondly remember the conversations on education and genetics with Emilia. Sonia kindly took me under her wing, making me feel at home at ISER and in the Economics Department. Despite her busy schedule, she selflessly found the time to make herself available to me, both academically and personally.

I am grateful to the University of Luxembourg for setting up the optimal conditions for the pursuit of a PhD. I cherish the good times and stimulating discussions with colleagues and friends from the *Doctoral School of Economics and Finance* and the *Department of Behavioural and Cognitive Sciences*, in particular those in the *Economic and Social Well-being* research group: Anthony, Liyou, Sonia, Gemma, Rémi, Vincent, Carlos, Niccolò, and Alex. Special thanks go to all those who have passed through the Original Office 4.222, for the fun moments and for bearing with my many "I have a question..."s, and to those, at the University and LISER alike, that I have the chance of calling friends.

Devo la mia buona salute mentale agli sgoccioli di questi anni di dottorato agli amici, quelli nuovi e quelli di una vita; quelli che ci sono sempre stati e con cui, anche se lontani, la distanza non conta.

Ho avuto la fortuna di avere il sostegno di una famiglia formidabile, a cui devo tutta la creazione di 'human capital' che mi porta a laurearmi per la terza volta. Dedico questo traguardo a mia mamma, che mi ha voluta e amata incondizionatamente; a Nino che, pur non avendomi voluta, mi ha scelta e amata come sua; e ai miei fratelli, Giulia e Totò, i migliori compagni di avventure che avessi potuto desiderare.

Ça ne serait pas une thèse faite au Luxembourg si je ne changeais pas de langue au moins trois fois pendant le même discours. En me conformant à cette tradition, j'en profite pour remercier dans sa langue maternelle la personne grâce à laquelle je suis aujourd'hui capable d'écrire ce morceau en français (et à laquelle je dois infiniment plus que ça). Anthony, cette thèse n'aurait pas été possible sans ton soutien, ta patience et ton amour. Sans ton exemple, chaque jour, de ce que signifie être un bon chercheur et, surtout, une bonne personne. Sans l'intelligence et la gentillesse, l'humour et le bonheur dont tu me couvres depuis qu'on se connait. Pour ça et tous ce que tu me donnes, je te suis reconnaissante.

Contents

Abstract

The thesis is divided in the following four chapters; as they address different research questions with different methods, each can be considered as a stand-alone paper.

Chapter 1 - Maternal Depression and Child Human Capital: A Genetic Instrumental Variable Approach. The Chapter addresses the causal relationship between maternal depression and child human capital using UK cohort data. We exploit the conditionally-exogenous variation in mothers' genomes in an instrumental-variable approach, and describe the conditions under which mother's genetic variants can be used as valid instruments. An additional episode of maternal depression between the child's birth up to age nine reduces both their cognitive and non-cognitive skills by 20 to 45% of a SD throughout adolescence. Our results are robust to a battery of sensitivity tests addressing, among others, concerns about pleiotropy and the maternal transmission of genes to her child.

Chapter 2 - Boys Don't Cry (or Do the Dishes): Family Size and the Housework Gender Gap. *Published in the Journal of Economic Behavior and Organization.* We here use data from the British Cohort Study (BCS) to link family size to age-16 children's contribution to household chores and the adult housework gender gap. Assuming that home production is an increasing function of family size and using an instrument to account for the endogeneity of fertility, we show that larger families have a different effect on boys and girls at age 16: girls in large families are significantly more likely to contribute to housework, with no effect for boys. We then show that childhood family size affects the housework gender gap between the cohort members and their partners at age 34. Women who grew up in larger families are more likely to carry out a greater share of household tasks in adulthood, as compared to women from smaller families. In addition, growing up in a large family makes cohort members more likely to sort into households with a wider housework gender gap as adults. We show that the persistent effect of family size is due to the adoption of behaviours in line with traditional gender roles: a lower likelihood of employment and shorter commutes for women, along with a higher employment probability for their partners.

Chapter 3 - Families at a Loss: The Relationship Between Income Changes and Child Human Capital The main scope of this paper is to assess the relationship and pathways that link gains and losses in family income to the cognitive and non-cognitive development of children. With data from the UK Millennium Cohort Study (MCS), I use a value added model to link distributional changes in family income to children's reading scores and internalising and externalising behaviour trajectories between age 3 and 15. I find that only income losses have a significant negative impact on the non-cognitive development of children and that around one third of the effect operates through channels related to mental health and well-being of mothers. Instead, movements upwards and downwards the income

distribution affect cognitive outcomes symmetrically. I find evidence suggesting that past income losses matter only in conjunction with current losses in explaining residualised reading test scores and that experiencing an income loss predicts the probability of entering the bottom quintile of the distribution of cognitive and non-cognitive skills. The evidence further suggests that the bottom quintile of non-cognitive skills is "stickier" than that of cognitive skills, with income gains having no significant effect in predicting the probability of exiting the bottom of the skills distribution.

Chapter 4 - Income and Wealth Volatility: Evidence from Italy and the U.S. in the Past Two Decades. *Published in the Journal of Economic Inequality.* Income volatility and wealth volatility are central objects of investigation for the literature on income and wealth inequality and dynamics. Here we analyse the two concepts in a comparative perspective for the same individuals in Italy and the U.S. over the last two decades. We find that in both countries wealth volatility reaches significantly higher values than income volatility, the effect being mostly driven by changes in the market value of real estate assets. We also show that there is more volatility in both dimensions in the U.S. and that the overall trend in both countries is increasing over time. We conclude by exploring volatility in consumption.

Co-author Statement

The following Chapter is a single-author paper:

- **Chapter 3 - Families at a Loss: The Relationship Between Income Changes and Child Human Capital**

The remaining three Chapters have all been co-authored. Reflecting my relative contribution to the papers, I am listed as first author in all three. Here follows a detailed account of my contribution to each of the papers:

- **Chapter 1 - Maternal Depression and Child Human Capital: A Genetic Instrumental-Variable Approach**, co-authored with Anthony Lepinteur (University of Luxembourg), Andrew E. Clark (Paris School of Economics, CNRS), Simone Ghislandi (Bocconi University), and Conchita D'Ambrosio (University of Luxembourg).

  The order of the authors here follows that of natural sciences, where the most active contributors are listed first and the project's Principal Investigators last. Conchita D'Ambrosio, Anthony Lepinteur, and I are at the origin of the research question. I was fully in charge of the analysis based on the genetic data, the writing of the methodological part, and most of the data cleaning. Anthony and I shared the econometric analysis and the writing of the first draft.

- **Chapter 2 - Boys Don't Cry (or Do the Dishes): Family Size and the Housework Gender Gap**, co-authored with Anthony Lepinteur (University of Luxembourg)

  Anthony and I are both at the origin of the research question (which originated on a late summer afternoon, on the Syracuse seafront). It is hard to separate our contributions to this paper: from the econometric analysis to the writing, each part of the paper is the result of a joint effort.

- **Chapter 4 - Income and Wealth Volatility: Evidence from Italy and the U.S. in the Past Two Decades**, co-authored with Edward N. Wolff (New York University), and Conchita D'Ambrosio (University of Luxembourg).

  Conchita is at the origin of the research question. I was in charge of the whole data analysis and about 80% of the writing of the paper.

General Introduction

General Introduction

Inequality and Parental Background

Prevalent social justice theories in Western societies rely on the concept of equality of opportunity, rather than equality of outcomes. A number of authors have argued that differences in observed individual outcomes are not only based on merit, but reflect personal and environmental differences that are outside of the control of single individuals (Rawls, 1971; Sen, 1980; Dworkin, 1981; Cohen, 1989; Roemer, 1993). Despite differences in the degree of responsibility individuals can be held accountable for and on the exact definition of the object of equalisation, the consensus around egalitarian theories is the equal access to opportunities (typically, opportunities for income), regardless on an individual's background and characteristics. Roemer (1998), among others, formalised this concept by describing outcomes as the result of two components: effort (the set of actions an individual is responsible for) and circumstances (the environment over which the individual has no control). As long as all individuals in a given society have access to the same opportunities and there is no discrimination based on personal characteristics such as gender or age, differences in observed outcomes can be interpreted as the result of preference heterogeneity and effort, and welcomed as a merit-based reward-system.

The aim of this dissertation at large is documenting inequalities deriving from circumstances that are outside of the control of individuals. I mostly focus on the ways parental background can shape children's outcomes, from human capital to the time spent in paid work and housework as adults. Not being able to choose the family they are born into, children cannot be held responsible for the choices (and circumstances) of their parents. Following egalitarian social-choice theories, policy makers should aim at compensating differences in outcomes that emerge from differences in initial circumstances.

Chapter 1 of this thesis addresses differences in child human capital caused by having a mother suffering from depressive symptoms; the identification strategy here relies on the mother's conditionally-exogenous genetic risk for depression. In Chapter 2, I then go on and consider how family size affects children's contribution to housework and the persistence of this effect into adulthood, with consequences for gender inequality. Chapter 3 describes the relationship between family income changes and child human capital accumulation, allowing for income losses and gains to have an asymmetric effect. Last, Chapter 4 further investigates the concept of income changes over time and documents the evolution of income volatility, as compared to wealth volatility, across two longitudinal datasets.

The rest of the introduction is organised as follows: I first review the literature on human capital formation, zooming in onto the role of parental income. I then address how parents can shape children's outcomes beyond human capital, by influencing their set of values, beliefs, and attitudes. Last, I investigate the role of the 'genetic lottery', a circumstance *par excellence*, in explaining differences in outcomes, by summarising the recent advancements in the field of social science genetics.

Human Capital Formation

Investments in human capital, as defined by the pioneering work of Becker (1962), are those "activities that influence future real income through the embedding of resources in people" (Becker, 1962, p. 9). Decisions to invest in human capital can thus be seen as the product of a rational cost-benefit analysis, where individuals decide to spend time and money in the formation of competences that are rewarded in the labour market, in order to achieve better economic outcomes. From the Mincerian wage equation (Mincer, 1970, 1974), assessing the relationship between years of schooling and wages, to more recent developments (Currie and Almond, 2011; Cunha and Heckman, 2008; Gertler *et al.*, 2014; Heckman, Humphries and Veramendi, 2018; Noboa-Hidalgo and Urzua, 2012), a lot of effort has been put in identifying what constitutes 'human capital' and which kind of investments contribute to its production.

Human capital is commonly defined as the ensemble of cognitive and non-cognitive skills.[1] Cunha and Heckman (2007, 2008) developed the theoretical framework for our current understanding of the human capital production function – first applied to cognitive skills only, and then extending the framework to non-cognitive skills. Skills are the product of genes, investments, and the environment, and they develop in accordance with a series of technologies and inputs that may vary according to the child's developmental stage. These technologies are typically characterised by the processes of self-productivity and dynamic

[1] A broader definition of human capital as 'human capabilities' includes also health stocks (Heckman, 2007).

complementarity: the first refers to the self-reinforcement of skills across developmental stages, with skills acquired in a given stage fostering the formation of skills in the next stages; the second instead addresses the synergy of investments across developmental stages, reflecting the fact that the production of skills in one stage increase the productivity of future investments.

Among the determinants of child human capital, parental background plays a fundamental role, by affecting both the parents' potential to invest in their children's skills and the environmental circumstances. In this thesis I first focus on the role of mothers' mental health. While depression is known to negatively affect employment and earnings (Zimmerman and Katon, 2005; Fletcher, 2013; Banerjee, Chatterji and Lahiri, 2017; Hakulinen *et al.*, 2019), productivity (Bubonya, Cobb-Clark and Wooden, 2017), marital status and marital satisfaction (Gotlib, Lewinsohn and Seeley, 1998), and parenting style (Kiernan and Huerta, 2008), little is known about the causal effect of maternal depression on child human capital. Based on the models of skills formation, maternal depression could affect child human capital investments, via a reduction of economic means and/or a worsening of the quality of time investments. Chapter 1 of this thesis addresses the net effect of maternal depression on children's cognitive and non-cognitive skills, by exploiting quasi-exogenous variation in mothers' genetic propensity for depression.

The role of another parental characteristic, namely family income, on child human capital is instead the object of Chapter 3.

The Role of Family Income

What constitutes an investment in human capital? From educational to emotionally nurturing environment, parental investments in child human capital typically involve a cost – be it money or time. Parental income, while arguably not constituting an investment *per-se*, can be seen as the enabler most forms of investments. Even in presence of the provision of public goods such as high-quality public education, parental income can complement the production of skills via ancillary investments and thus contribute to the persistence of intergenerational inequality.

A number of empirical studies document a positive association between family income and child human capital (Duncan and Brooks-Gunn, 1997; Shea, 2000; Yeung, Linver and Brooks-Gunn, 2002; Washbrook, Gregg and Propper, 2014), finding larger effect sizes for measures of cognitive rather than non-cognitive skills. Changes in family income have also been shown to have an effect on child human capital (see Dahl and Lochner, 2012, for a review). Mostly based on variations in earned income tax credit schemes or child-care support, this causal literature shows modest positive effects of an exogenous increase in family income on child cognitive skills, with little evidence on the effect on non-cognitive skills. However, income gains and losses need not necessarily have a symmetric effect on child human capital accumulation: drawing from prospect theory (Kahneman and Tversky, 1979), one may think that individuals are more affected by income losses rather than gains. Additionally, income losses likely interact with market conditions such as the presence of credit constraints; as such, they may well have an asymmetric effect with respect to gains on the parents' ability to foster their children's human capital, either via the provision of material inputs or via the quality and quantity of their time inputs. Using longitudinal data from a UK cohort study of children born at the turn of the millennium, Chapter 3 investigates whether there is any descriptive evidence of an asymmetric effect of parental income gains and losses on child human capital.

Changes in income, however, may not necessarily translate into changes in the parents' ability to invest in their children's human capital. Moved by altruism, parents may have a preference for keeping constant their investment level and may resort to the credit market and/or erode wealth as insurance against adverse economic shocks. Similarly, in standard theoretical models, wealth plays a protective role in that it can help ensuring consumption smoothing. Given its role of insurance, one may then think wealth to be on average more stable than income. Chapter 4 investigates whether this is indeed the case, describing the evolution of individual income and wealth volatility in the U.S. and in Italy.

Cultural Socialisation, Identity, and the Persistence of Gender Inequalities

Parental characteristics can affect their children's later life outcomes in ways beyond human capital formation. Parents can play both a direct and an indirect role in shaping their children's outcomes, by influencing their attitudes, behaviours, and identity formation. For example, Dohmen *et al.* (2012) use German data to confirm some theoretical predictions of models of cultural transmission, providing evidence of the intergenerational transmission of risk and trust attitudes – with consequences on a variety of child outcomes. Doepke and Zilibotti (2008) explain the ascent of the middle-class in post-industrial Britain with the practice of passing values of patience onto the next generation. Using a natural experiment on the US military draft, Fernández, Fogli and Olivetti (2004) argue that men who grew up in families with working mothers develop less stereotypical gender attitudes and are less likely to be the only or main breadwinner in their household. In a similar spirit, Farré and Vella (2013) find evidence of a direct correlation between the gender attitudes of parents and children, showing that a mother's beliefs about the role of women, in the household and in the labour market alike, affect her children's views towards female employment. Additionally, these mothers are more likely to have employed daughters and daughters-in-law.

Much of this literature can be framed in the context of cultural transmission models à la Bisin and Verdier (2001). According to these theoretical frameworks, children develop their preferences via imitation and adaptation mechanisms that can be defined as 'socialisation'. Parents, moved by 'paternalistic' altruism, have a preference for their children acquiring the same set of cultural norms they hold. As such, they exert effort to provide their children with a direct form of socialisation. The cultural mix of the surrounding society provides an 'oblique' form of socialisation instead. The more aligned the preferences of the parents are with the norms embedded in society, the less prevalent is the role of parents (family and society are cultural substitutes). On the contrary, as the societal norms diverge from the parents', families find it optimal to put more effort into direct forms of socialisation for their children. Pairing these insights with those from identity frameworks (Akerlof and Kranton, 2000, 2010) allows to appraise the role of cultural socialisation in the intergenerational transmission of gender norms. If children are socialised into stereotypical cultural prescriptions (about, say, gender or race), they will likely internalise these norms as part of their identity and find it costly to deviate from them. The social and personal costs associated with deviating from norms that are part of one's identity provide a plausible explanation for the cultural persistence of stereotypical norms and behaviours.

A behavioural prescription often associated with stereotypical gender norms is that "women should do more housework than men" (Akerlof and Kranton, 2010). The prevalence of this prescription is consistent with empirical findings showing that women still spend significantly more time in household chores than men (Bittman *et al.*, 2003; Gupta and Ash, 2008;

Lyonette and Crompton, 2015). One may think this evidence to only be the result of rational specialisation patterns illustrated by standard household bargaining models (Chiappori, 1992; Van Klaveren, van Praag and van den Brink, 2008). However, women's intra-household bargaining power has likely never been higher, as reflected by their labour force participation and educational outcomes being at a historical high in most OECD economies (International Labour Office, 2018; World Economic Forum, 2018). So why do we still observe specialisation patterns that are systematically gendered? Norms, identity, and culture could indeed be playing a role, as suggested by Akerlof and Kranton (2010). Women who are more educated or earn more than their partners, and who so deviate from stereotypical gender-role prescriptions, have been shown to compensate via a more traditional division of housework (Bittman *et al.*, 2003; Lyonette and Crompton, 2015; Bertrand, Kamenica and Pan, 2015).

Chapter 2 of this thesis draws from the considerations above and explores the role of parental socialisation on the adult housework gender gap, by exploiting exogenous changes in family size during childhood.

The Role of Genes

Other than through their behaviours and actions, parents have a direct biological effect oh their children, via the transmission of genes. The genetic makeup of a person, fixed at conception, can contribute to shaping a variety of outcomes, from health to educational attainment (Lee *et al.*, 2018) and subjective well-being (Okbay *et al.*, 2016). In the absence of inequality of opportunities, meritocratic societies welcome inequality of outcomes as a way to reward effort (Arrow, Bowles and Durlauf, 2000; Atkinson, 2015): keeping environmental factors constant and assuming better grades are administered to those who study harder, it would be 'fair' for someone with good grades to have access to better-paying jobs with respect to someone with lower grades. However, factoring in genetics considerations makes the line separating effort and circumstances look ever-so-thin. As argued by Harden (2018, par. 10), "success [...] is partially a result of genetic luck. No one earned his or her DNA sequence, yet some of us are benefiting enormously from it".

Heritability: From Twin-Studies to Genome-Wide Association Studies

To what extent can genes explain observed individual variation in socio-economic outcomes? What is the relative weight of a child's DNA as an input in her human capital production function? Decades of evidence from twin-studies summarised in the meta-analysis of Polderman *et al.* (2015) show that the average heritability of a trait (i.e. the share of observed variation in a trait that is only due to genetic differences) is little below 50%. While this figure hides substantial heterogeneity across different traits (from around 90% heritability for height, to around 40% for alcoholism), it provides a first gauge of the importance of genes in shaping observed differences in outcomes.

With the large drop in the costs of genotyping technologies – that is, technologies extracting common variations in the DNA sequence, typically called SNPs (Single-Nucleotide Polymorphisms) – facilitating the direct access to individuals' DNA, the landscape of social-science genetics research has dramatically changed over the past two decades. The ease of access to individual genotype data fostered the diffusion of Genome-Wide Association Studies (GWAS), aimed at detecting systematic associations between SNPs and given traits (e.g. educational attainment) in ethnically-homogeneous populations.[2] One of the research outputs of a GWAS consists in a set of summary statistics reporting the magnitude of the association between each SNP and the target trait, as well as the p-value of such association. With the diffusion of national biobanks and the practice of collecting genetic material from social studies' participants, GWAS sample sizes have substantially increased (reaching over 1 million participants in the GWAS for educational attainment of Lee *et al.*, 2018), allowing for the detection of increasingly precise SNP-trait associations. The larger precision has contributed to a better understanding of the highly polygenic nature of complex traits such as education or depression: there is no 'gene for' a trait, rather a diffused signal over a multitude of genetic variants. Additionally, the genetic variants identified in such studies exert an influence on traits via a series of complex biological processes, which are inextricably linked to the environment.

[2]Results from GWAS have shed a new light on our understanding of heritability: the fraction of a trait's variance that can be explained by the SNPs analysed in a GWAS is typically lower than the estimates of heritability derived from twin studies – a phenomenon known as the 'missing heritability' problem (Young, 2019).

Polygenic Scores in Economics Research

The diffused genetic signal derived from GWAS results can be summarised into a risk score (the so-called 'polygenic score', or PGS) that captures an individual's genetic likelihood of displaying a particular trait (more details on the computation of polygenic scores can be found in Appendix 1.B). The out-of-sample predictive power of polygenic scores, measured with the incremental R^2, is a function of the GWAS sample size, heritability, and number of genetic variants associated with the trait (Dudbridge, 2013). The polygenic score for educational attainment based on the latest GWAS of Lee *et al.* (2018) has been shown to predict around 10% of the variation in years of education in replication cohorts – a predictive power similar to that of father's education and verbal ability (Lee *et al.*, 2018). Additionally, results from within-family studies show that polygenic scores can help explain a large fraction of observed differences between siblings, across a variety of traits (Belsky *et al.*, 2018). Despite their potentially large predictive power, polygenic scores are far from being 'fortune-tellers' (Harden, 2021): just as we cannot foresee an individual's educational attainment solely based on their mother's education level, a high polygenic score for education does not automatically translate into more years of schooling.

What is the role of polygenic scores in economics research? From simple covariates to tools to assess the moderating role of genes in public policy interventions (see Harden and Koellinger, 2020, for a review of potential applications), polygenic scores provide an accessible way of integrating genetic considerations into our understanding of economic phenomena and social change. For example, Belsky *et al.* (2018) show that the polygenic score for educational attainment is a predictor of upward social mobility, independent of socially-transmitted measures of economic advantage. Kweon *et al.* (2020) find that the polygenic score for income has a causal link to socio-economic status and health. The authors emphasise how the mediating role of environmental pathways such as education leave room for policy intervention. Polygenic scores can be also used in the context of Mendelian Randomisation – a term borrowed from epidemiology that describes the use of genetic variants as instrumental variables (Davey Smith and Hemani, 2014; Hemani, Bowden and Davey Smith, 2018; Koellinger and de Vlaming, 2019). This approach, declined in an intergenerational context, is at the basis of the identification strategy of Chapter 1 of this thesis (where the assumptions required in the context of genetic instrumental-variables are described in detail).

Another promising avenue for the use of polygenic scores in social-science research is that of gene-environment interactions ($G \times E$). Differences in genetic endowments can moderate the effect of policy interventions and produce treatment effect heterogeneity. This is the case, for instance, of the relationship between education and health: Barcellos, Carvalho and Turley (2018) exploit the raise in the school-leaving age in the UK (the 1972 RoSLA) to show that the policy was most effective in reducing BMI for individuals with a higher genetic risk for obesity. Other examples explore the moderating role of genes in relation to smoking behaviour and health, using exogenous environmental variations coming from the Vietnam draft (Schmitz and Conley, 2016, 2017) or eligibility for MediCare (Biroli and Zwyssig, 2021); others investigate the interaction between genes, drinking behaviour and alcohol licensing policies in the UK (Biroli and Zünd, 2020).

One could wonder about the policy implications that come with genetic research. First, it is important to mention that genetic heritability does not imply biological determinism. If genes are fixed, the environment is not: even in the presence of a trait (say, myopia) that is 100% heritable, environmental factors will always able to play a role (think of eye surgery or glasses to correct for myopia). Social-science genetics has opened the door to a broader definition of inequality of opportunity, that is not only the result of the social lottery (the circumstances in which we are born), but also of the genetic lottery. Taking into account genetic differences can help designing better targeted policies (Joint Research Centre F7 - Knowledge Health and Consumer Safety, 2019), informed by the integration of genetic research into the evaluation of existing policy interventions.

Dissertation Outline

This thesis draws from the literature described in the section above in order to address the following research questions. First, we ask what is the causal role of parents' mental health in shaping children's human capital trajectories (Chapter 1). Second, we address the role of childhood family structure in explaining the housework gender gap (Chapter 2). I then re-evaluate the role of income changes in the human capital production function, by asking if gains and losses in family income affect child human capital formation in the same way

(Chapter 2). Last, we use two longitudinal datasets to check whether, within each family, income is more volatile than wealth (Chapter 4).

Chapter 1, "Maternal Depression and Child Human Capital: a Genetic Instrumental-Varialble Approach", addresses the causal role of maternal depression in the formation of children's cognitive and non-cognitive skills. Using a unique mix of socio-economic and biological data from a UK-based cohort study, we here exploit the conditionally exogenous variation provided by the mothers' DNA in order to create an instrumental variable that captures their genetic risk of being depressed. We describe in detail the conditions under which mother's genetic variants can be used as valid instruments and address concerns such as pleiotropy (the correlation between genes and multiple traits) and mother-to-child genetic inheritance. The Chapter shows that maternal depression hinders cognitive and non-cognitive outcomes throughout adolescence: one additional episode of maternal depression between the child's birth up to age nine reduces test-scores achievement by 20% of a standard deviation and socio-emotional and behavioural health by 45% of a standard deviation. These results are robust to a battery of sensitivity tests and survive a partial relaxation of the exclusion restriction. Together with evidence showing that the treatment of maternal depression has little benefits in terms of child human capital (Baranov *et al.*, 2020), our results seem to point to the fact that policies aimed at preventing, rather than treating, maternal depression might bring about greater societal returns.

Chapter 2, "Boys Don't Cry (Or Do the Dishes): Family Size and the Housework Gender Gap", deals with the intergenerational transmission of gender inequality, focusing on the persistence of the housework gender gap in developed economies. We here use data from the 1970 British Cohort Study (BCS) to investigate the role of family size in childhood as a determinant of the adult housework gender gap. We do so by linking first family size to the allocation of household chores to children. Using the same-sex instrument introduced by Angrist and Evans (1998) to account for the endogeneity of fertility, we show that an increase in family size has a different effect for boys and girls at age 16: girls in large families are significantly more likely to contribute to housework, while no effect is found for boys. Additionally, girls in larger families spend less time doing homework and in leisure activities. We find the effect to persist into adulthood, translating into a larger housework gender gap at ages 34 and 42. Results are driven by women in our sample who grew up in larger families, who appear to sort into cohabiting relationships where they systematically carry out more housework than their partners. The effect of family size is larger for children from low-SES families and families in which the mother holds conservative gender attitudes. We further show that the persistent effect of family size in adulthood can be partly explained by the adoption of behaviours in line with stereotypical gender roles, such as a lower likelihood of employment and shorter commutes for women, and a higher employment probability for their partners. These results provide evidence that a gendered attribution of household chores to children can contribute to a persistent housework gender gap in the next generation, potentially limiting women's career opportunities and contributing to institutionalise gender inequalities.

Chapter 3, "Families at a Loss: The Relationship Between Income Changes and Child Human Capital" focuses again on the way parental characteristics can affect child human capital. In particular, I test descriptively whether positive and negative shocks in family income have a symmetric effect on child cognitive and non-cognitive development. To do so, I use data from the UK Millennium Cohort Study, looking at the impact of wave-to-wave gains and losses in family income quintiles on the the socio-emotional development and reading test-scores of children aged 3 to 15. Accounting for the dynamic nature of human capital with a value-added model, the chapter shows that income losses are associated with worse cognitive and non-cognitive outcomes of children, while income gains exert a positive impact on cognitive skills only. Consistent with the literature in developmental psychology, the effect of income losses is partly mediated by measures of maternal well-being. Additional results investigating dynamic and distributional effects suggest that past income losses matter only in conjunction with current losses in explaining residualised reading test-scores, and that experiencing an income loss predicts the probability of entering the bottom quintile of the distribution of both cognitive and non-cognitive skills. The evidence further suggests that the bottom quintile of non-cognitive skills is somewhat stickier than the cognitive skills' one, with income gains having no significant effect in predicting the probability of exiting the bottom of the skills distribution.

Chapter 4, "Income and Wealth Volatility: Evidence from Italy and the U.S. in the Past Two Decades", zooms-in on the concept of income changes. The Life-Cycle Theory suggests that wealth can act as form of self-insurance against income shocks: in periods of diminished income, consumption smoothing is ensured through the erosion of wealth and/or access to the credit market. The presence of borrowing constraints might translate into even stronger incentives to accumulate wealth and save 'for a rainy day'. Following a life-cycle approach, one might then believe income volatility in a given economy to be on average larger than wealth volatility: while the former is more subject to life events and transitory conditions, the latter is run down in case of necessity in order to converge towards a stable consumption path. If that was the case, the adverse effects of income losses described in Chapter 3 might be partly compensated with household wealth. However, is wealth really less volatile than income? Chapter 4 provides an empirical test for this assumption, using a variance-decomposition approach on high-detail income and wealth information from individual panel datasets in the US and in Italy. Describing the evolution of individual equivalent income and wealth volatility over time, we find that wealth volatility reaches significantly higher values than income volatility over the period 2002-2014. While the distance between income and wealth volatility is relatively contained in Italy, it appears to be larger in the U.S.. When turning to consumption, we predictably find that consumption volatility is the lowest in both countries. Investigating the process of wealth accumulation, we provide evidence for wealth volatility being mostly driven by changes in the market value of real estate assets.

References

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and identity." *Quarterly Journal of Economics*, 115: 715–753.

Akerlof, George A., and Rachel E. Kranton. 2010. *Identity Economics.* Princeton University Press.

Angrist, Joshua D., and William Evans. 1998. "Children and their parents' labor supply: Evidence from exogenous variation in family size." *American Economic Review*, 88: 450–477.

Arrow, Kenneth, Samuel Bowles, and Steven N. Durlauf. 2000. *Meritocracy and Economic Inequality.* Princeton University Press.

Atkinson, Anthony B. 2015. *Inequality: What Can Be Done?* Harvard University Press.

Banerjee, Souvik, Pinka Chatterji, and Kajal Lahiri. 2017. "Effects of psychiatric disorders on labor market outcomes: a latent variable approach using multiple clinical indicators." *Health Economics*, 26: 184–205.

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko. 2020. "Maternal depression, women's empowerment, and parental investment: evidence from a randomized controlled trial." *American Economic Review*, 110: 824–59.

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: E9765–E9772.

Becker, Gary S. 1962. "Investment in human capital: A theoretical analysis." *Journal of Political Economy*, 70: 9–49.

Belsky, Daniel W., Benjamin W. Domingue, Robbee Wedow, Louise Arseneault, Jason D. Boardman, Avshalom Caspi, Dalton Conley, Jason M. Fletcher, Jeremy Freese, Pamela Herd, *et al.* 2018. "Genetic analysis of social-class mobility in five longitudinal studies." *Proceedings of the National Academy of Sciences*, 115: E7275–E7284.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender identity and relative income within households." *Quarterly Journal of Economics*, 130: 571–614.

Biroli, Pietro, and Christian Lukas Zünd. 2020. *Genes, Pubs, and Drinks: Gene-Environment Interplay and Alcohol Licensing Policy in the UK.* Mimeo.

Biroli, Pietro, and Laura Zwyssig. 2021. *Moral Hazard Heterogeneity: Genes and Health Insurance Influence Smoking after a Health Shock.* IZA Discussion Paper, 14176.

Bisin, Alberto, and Thierry Verdier. 2001. "The economics of cultural transmission and the dynamics of preferences." *Journal of Economic Theory*, 97: 298–319.

Bittman, Michael, Paula England, Liana Sayer, Nancy Folbre, and George Matheson. 2003. "When does gender trump money? Bargaining and time in household work." *American Journal of Sociology*, 109: 186–214.

Bubonya, Melisa, Deborah A. Cobb-Clark, and Mark Wooden. 2017. "Mental health and productivity at work: Does what you do matter?" *Labour Economics*, 46: 150–165.

Chiappori, Pierre-André. 1992. "Collective labor supply and welfare." *Journal of Political Economy*, 100: 437–467.

Cohen, Gerald A. 1989. "On the currency of egalitarian justice." *Ethics*, 99: 906–944.

Cunha, Flavio, and James J. Heckman. 2007. "The technology of skill formation." *American Economic Review*, 97: 31–47.

Cunha, Flavio, and James J. Heckman. 2008. "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation." *Journal of Human Resources*, 43: 738–782.

Currie, Janet, and Douglas Almond. 2011. "Human capital development before age five." In *Handbook of Labor Economics.* Vol. 4, Eds. O. Ashenfelter and D. Card, Chapter 15, 1315–1486. Elsevier.

Dahl, Gordon B., and Lance Lochner. 2012. "The impact of family income on child achievement: Evidence from the earned income tax credit." *American Economic Review*, 102: 1927–56.

Davey Smith, George, and Gibran Hemani. 2014. "Mendelian randomization: genetic anchors for causal inference in epidemiological studies." *Human Molecular Genetics*, 23: R89–R98.

Doepke, Matthias, and Fabrizio Zilibotti. 2008. "Occupational choice and the spirit of capitalism." *Quarterly Journal of Economics*, 123: 747–793.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2012. "The intergenerational transmission of risk and trust attitudes." *Review of Economic Studies*, 79: 645–677.

Dudbridge, Frank. 2013. "Power and predictive accuracy of polygenic risk scores." *PLoS Genetics*, 9: e1003348.

Duncan, Greg J., and Jeanne Brooks-Gunn. 1997. *Consequences of Growing Up Poor.* Russell Sage Foundation.

Dworkin, Ronald. 1981. "What is equality? Part 2: Equality of resources." *Philosophy and Public Affairs*, 283–345.

Farré, Lidia, and Francis Vella. 2013. "The intergenerational transmission of gender role attitudes and its implications for female labour force participation." *Economica*, 80: 219–247.

Fernández, Raquel, Alessandra Fogli, and Claudia Olivetti. 2004. "Mothers and sons: Preference formation and female labor force dynamics." *Quarterly Journal of Economics*, 119: 1249–1299.

Fletcher, Jason. 2013. "Adolescent depression and adult labor market outcomes." *Southern Economic Journal*, 80: 26–49.

Gertler, Paul, James J. Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M. Chang, and Sally Grantham-McGregor. 2014. "Labor market returns to an early childhood stimulation intervention in Jamaica." *Science*, 344: 998–1001.

Gotlib, Ian H., Peter M. Lewinsohn, and John R. Seeley. 1998. "Consequences of depression during adolescence: Marital status and marital functioning in early adulthood." *Journal of Abnormal Psychology*, 107: 686–690.

Gupta, Sanjiv, and Michael Ash. 2008. “Whose money, whose time? A nonparametric approach to modeling time spent on housework in the United States.” *Feminist Economics*, 14: 93–120.

Hakulinen, Christian, Marko Elovainio, Martti Arffman, Sonja Lumme, Sami Pirkola, Ilmo Keskimäki, Kristiina Manderbacka, and Petri Böckerman. 2019. “Mental disorders and long-term labour market outcomes: Nationwide cohort study of 2,055,720 individuals.” *Acta Psychiatrica Scandinavica*, 140: 371–381.

Harden, K. Paige. 2018. *Why Progressives Should Embrace the Genetics of Education.* New York Times. Retrieved from: www.nytimes.com/2018/07/24/opinion/dna-nature-genetics-education.html.

Harden, K. Paige. 2021. ““Reports of my death were greatly exaggerated”: Behavior genetics in the postgenomic era.” *Annual Review of Psychology*, 72: 37–60.

Harden, K. Paige, and Philipp D. Koellinger. 2020. “Using genetics for social science.” *Nature Human Behaviour*, 4: 567–576.

Heckman, James J. 2007. “The economics, technology, and neuroscience of human capability formation.” *Proceedings of the National Academy of Sciences*, 104: 13250–13255.

Heckman, James J., John Eric Humphries, and Gregory Veramendi. 2018. “Returns to education: The causal effects of education on earnings, health, and smoking.” *Journal of Political Economy*, 126: S197–S246.

Hemani, Gibran, Jack Bowden, and George Davey Smith. 2018. “Evaluating the potential role of pleiotropy in Mendelian randomization studies.” *Human Molecular Genetics*, 27: R195–R208.

International Labour Office. 2018. *World Employment and Social Outlook: Trends for Women 2018 – Global snapshot.* ILO, Geneva, Switzerland.

Joint Research Centre F7 - Knowledge Health and Consumer Safety. 2019. *Genome-wide association studies, polygenic scores and social science genetics.* JRC Technical Report, 117414. Luxembourg: European Commission.

Kahneman, Daniel, and Amos Tversky. 1979. “Prospect theory: An analysis of decision under risk.” *Econometrica*, 47: 263–292.

Kiernan, Kathleen E., and M. Carmen Huerta. 2008. “Economic deprivation, maternal depression, parenting and children’s cognitive and emotional development in early childhood.” *British Journal of Sociology*, 59: 783–806.

Koellinger, Philipp D., and Ronald de Vlaming. 2019. “Mendelian randomization: the challenge of unobserved environmental confounds.” *International Journal of Epidemiology*, 48: 665–671.

Kweon, Hyeokmoon, Casper Burik, Richard Karlsson Linnér, Ronald De Vlaming, Aysu Okbay, Daphne Martschenko, Kathryn P. Harden, Thomas A. DiPrete, and Philipp D. Koellinger. 2020. *Genetic Fortune: Winning or Losing Education, Income, and Health.* Tinbergen Institute Discussion Paper, 2020-053/V.

Lee, James J, Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzian, Meghan Zacher, Tuan Anh Nguyen-Viet, Peter Bowers, Julia Sidorenko, Richard Karlsson Linnér, *et al.* 2018. "Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals." *Nature Genetics*, 50: 1112–1121.

Lyonette, Clare, and Rosemary Crompton. 2015. "Sharing the load? Partners' relative earnings and the division of domestic labour." *Work, Employment and Society*, 29: 23–40.

Mincer, Jacob. 1970. "The distribution of labor incomes: a survey with special reference to the human capital approach." *Journal of Economic Literature*, 8: 1–26.

Mincer, Jacob. 1974. *Schooling, Experience, and Earnings.* Columbia University Press.

Noboa-Hidalgo, Grace E., and Sergio S. Urzua. 2012. "The effects of participation in public child care centers: Evidence from Chile." *Journal of Human Capital*, 6: 1–34.

Okbay, Aysu, Bart M.L. Baselmans, Jan-Emmanuel De Neve, Patrick Turley, Michel G. Nivard, Mark Alan Fontana, S. Fleur W. Meddens, Richard Karlsson Linnér, Cornelius A. Rietveld, Jaime Derringer, *et al.* 2016. "Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses." *Nature Genetics*, 48: 624–633.

Polderman, Tinca J.C., Beben Benyamin, Christiaan A. De Leeuw, Patrick F. Sullivan, Arjen Van Bochoven, Peter M. Visscher, and Danielle Posthuma. 2015. "Meta-analysis of the heritability of human traits based on fifty years of twin studies." *Nature Genetics*, 47: 702–709.

Rawls, John. 1971. *A Theory of Justice.* Harvard University Press.

Roemer, John E. 1993. "A pragmatic theory of responsibility for the egalitarian planner." *Philosophy and Public Affairs*, 146–166.

Roemer, John E. 1998. *Theories of Distributive Justice.* Harvard University Press.

Schmitz, Lauren L., and Dalton Conley. 2016. "The long-term consequences of Vietnam-era conscription and genotype on smoking behavior and health." *Behavior Genetics*, 46: 43–58.

Schmitz, Lauren L., and Dalton Conley. 2017. "The effect of Vietnam-era conscription and genetic potential for educational attainment on schooling outcomes." *Economics of Education Review*, 61: 85–97.

Sen, Amartya. 1980. "Equality of what?" In *Tanner Lectures on Human Values.* Vol. 1, Ed. S. McMurrin, Chapter 6, 197–220. Cambridge University Press.

Shea, John. 2000. "Does parents' money matter?" *Journal of Public Economics*, 77: 155–184.

Van Klaveren, Chris, Bernard van Praag, and Henriette Maassen van den Brink. 2008. "A public good version of the collective household model: an empirical approach with an application to British household data." *Review of Economics of the Household*, 6: 169–191.

Washbrook, Elizabeth, Paul Gregg, and Carol Propper. 2014. "A decomposition analysis of the relationship between parental income and multiple child outcomes." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 177: 757–782.

World Economic Forum. 2018. *The Global Gender Gap Report.* World Economic Forum, Geneva.

Yeung, W. Jean, Miriam R. Linver, and Jeanne Brooks-Gunn. 2002. "How money matters for young children's development: Parental investment and family processes." *Child Development*, 73: 1861–1879.

Young, Alexander I. 2019. "Solving the missing heritability problem." *PLoS Genetics*, 15: e1008222.

Zimmerman, Frederick J., and Wayne Katon. 2005. "Socioeconomic status, depression disparities, and financial strain: what lies behind the income-depression relationship?" *Health Economics*, 14: 1197–1215.

Chapter 1

Maternal Depression and Child Human Capital: A Genetic Instrumental-Variable Approach

Maternal Depression and Child Human Capital: A Genetic Instrumental-Variable Approach

with Anthony Lepinteur (University of Luxembourg),
Andrew E. Clark (Paris School of Economics, CNRS),
Simone Ghislandi (Bocconi University),
and Conchita D'Ambrosio (University of Luxembourg).

1.1 Introduction

The prevalence of mental-health disorders has been rising steadily for over two decades (Stansfeld and Hinchliffe, 2016), and these are now estimated to affect over 20% of the population in the UK (www.mind.org.uk) and the US (www.nami.org/mhstats). Depression is one of the most common of these disorders. A vast literature has documented worse outcomes for the depressed in terms of not only health, but also employment and earnings (Zimmerman and Katon, 2005; Fletcher, 2013; Banerjee, Chatterji and Lahiri, 2017; Hakulinen *et al.*, 2019), productivity (Bubonya, Cobb-Clark and Wooden, 2017), marital status and marital satisfaction (Gotlib, Lewinsohn and Seeley, 1998), and parenting style (Kiernan and Huerta, 2008). Major Depressive Disorder has been identified as the largest worldwide contributor to years lost to disability (Prince *et al.*, 2007).

Depression in addition, likely also spills over onto others. There is a great deal of work on the intergenerational correlation between parental and child depression (see Gotlib, Goodman and Humphreys, 2020, for a recent summary). We here consider the consequences of maternal depression on child human-capital in unique British birth-cohort data, beyond the intergenerational inheritance of the genes associated with depression.

While a broad range of descriptive evidence has underlined the negative association between maternal depression and child outcomes (see Goodman *et al.*, 2011; O'hara and McCabe, 2013, for meta-analyses and reviews of the psychological literature), there has been little causal analysis of this intergenerational link. One exception is Dahlen (2016), who uses non-parametric bounds to estimate ranges of the negative causal impact of maternal depression on the test scores and socioemotional outcomes of US kindergarten children. More notably, Von Hinke *et al.* (2019) rely on unexpected life experiences (the illness or death of friends and family members) to isolate the effect of perinatal maternal depression on children's cognitive and non-cognitive skills in a UK birth cohort (ALSPAC; the same dataset that we use here). They find that mother's worse mental health around birth negatively affects their children's non-cognitive skills, with the effects fading away as the child approaches adolescence. No effect is found on cognitive outcomes.

A small number of contributions have focused on the beneficial causal effects of the successful treatment of depressed mothers. Perry (2008), exploiting the arguably-exogenous variation in US primary-care physicians' propensity to diagnose depression, shows that treating maternal depression improved children's asthma outcomes. Using data from a randomised controlled trial, Baranov *et al.* (2020) find that prenatally-depressed mothers in rural Pakistan who were offered psychotherapy had better mental-health outcomes, and invested more time and money in their children (although there is only limited evidence that this investment improved child-development outcomes).

The causal link between parental mental health and child outcomes is of primary policy importance, but is in general not particularly easy to establish. The interplay between maternal mental health and child human-capital development is complex and subject to potential endogeneity concerns. For instance, poor child school performance or behavioural problems might themselves produce maternal depression; alternatively, environmental variables (shared by parents and children who live in the same household), such as local public goods or criminality, could feed through to both parental mental health and child outcomes. In both cases it is difficult to establish causality.

We here address endogeneity via recent advances in Epidemiology and Molecular Genetics. In particular, we adopt a genetic instrumental-variable approach (similar to DiPrete, Burik and Koellinger, 2018), and instrument maternal depression using a synthetic measure (the polygenic score) based on the mother's genetic variants that are robustly associated with the trait of depression.

Our empirical analysis is based on genetic and socio-economic information on mother-child pairs from the Avon Longitudinal Study of Parents and Children (ALSPAC), a UK-based cohort study that recruited about 14,000 pregnant mothers in the early 1990s. The key explanatory variable is reported maternal depression: this is a summary measure from the answers mothers give to questions about recent depression in seven different data waves from childbirth up to child age nine. We instrument this cumulative depression score by the polygenic score (PGS) for maternal depression, using Genome-Wide Association Studies (GWAS) summary statistics from the depression meta-analysis in Turley *et al.* (2018). Our methodological approach is similar to that in Von Hinke *et al.* (2016), who illustrate the assumptions under which an individual's genetic variants can be used as instrumental variables for that individual's traits (in their empirical application, child fat mass). Our question differs from theirs, as the trait we instrument (depression) and the outcome (human capital) refer to different individuals (respectively, the mother and her child). In this intergenerational analysis, additional concerns need to be addressed, such as those deriving from genetic inheritance that we will discuss below.

Following the human-capital development and skill-formation literature (see, for example, Cunha and Heckman, 2008), we consider child cognitive and non-cognitive skills as human capital components. The cognitive element is given by the measurement of child skills and knowledge at different stages of compulsory education in the UK. We in particular analyse the child's average Key Stage test-scores at ages 11 and 14, and their total GCSE score at age 16 (at the end of compulsory education); all three of these test scores come from administrative data. Non-cognitive skills come from the child's score from the questions in the Strengths and Difficulties Questionnaire (as reported by their principal carer) at child ages 11, 13 and 16.

The genetic instrument allows us to isolate an exogenous change in maternal depression up to child age nine and establish its causal impact on the child's later human capital. We find that one additional episode of maternal depression (out of the seven recorded) has a persistent negative impact on both cognitive and non-cognitive skills, with an effect size of around 20% of a standard deviation for the former and 40% for the latter.

Our identification strategy relies on a number of assumptions: while the relevance of the mother's PGS in predicting her depression can be formally tested, the genetic nature of this instrument calls for a more thorough investigation of the exclusion restriction. We illustrate the potential pathways that may compromise identification here, and discuss some ways in which these concerns can be addressed. Pleiotropy (when one genetic variant can explain a number of different traits) is arguably the main issue with genetic instrumentation in general. The intergenerational nature of our research introduces a second potential problem, that of genetic inheritance: as the child inherits about 50% of each parent's genetic variants, the direct effect of the child's inherited genetic variants may confound the relationship between mother's instrumented depression and child human capital. Child outcomes will be affected by the child's own depression, and this is partly due to the genetic propensity for depression that was inherited from the mother. But this is not what we understand by asking if depressed mothers affect their children's outcomes: we here rather wish to establish the effect of maternal depression net of genetic inheritance.

We tackle some of the pleiotropic concerns by controlling for a set of maternal and child traits that might be affected by the genetic variants used in the construction of the PGS, and that are in turn likely to affect human-capital development (e.g. educational attainment and fertility decisions). Following Lawlor *et al.* (n.d.) and DiPrete, Burik and Koellinger (2018), we control for genetic inheritance by holding constant the child's own PGS for depression, as well as their PGSs for cognitive and non-cognitive skills. Our results are robust to these and other sensitivity tests.

The remainder of this paper is organised as follows. Section 1.2 describes the birth-cohort data that we use. Section 1.3 then provides an overview of the conditions under which genetic variants can be used as instrumental variables in observational data, and considers the specific issues when the treatment and the outcome refer to different individuals who are genetically-related. The main results, of a sizeable causal effect of maternal depression in childhood on adolescent children's cognitive and non-cognitive skills, and a variety of robustness checks, appear in Section 1.4. Last, Section 1.5 concludes.

1.2 Data: The Avon Longitudinal Study of Parents and Children

We will use mother's genetic information as an instrument to establish the causal effect of her depression on her children's cognitive and non-cognitive outcomes. The data requirements to carry out this analysis are stringent. We need information on mother's reported depression during her child's young years, the adolescent outcomes of her child, and both the mother's and the child's genotype. Few datasets contain all of this information. One that does is the Avon Longitudinal Study of Parents and Children (ALSPAC) survey, also known as 'The Children of the 90s'.

ALSPAC is an English birth-cohort study designed to investigate the influence of environmental, genetic, and socio-economic variables on health and development over the life course. Over 14,000 pregnant women who were due to give birth between April 1991 and December 1992 in the county of Avon (Bristol and its surrounding areas) were recruited. These women and their families have been followed ever since, even if they move out of the original recruitment area (see www.bristol.ac.uk/alspac/). The pregnancy outcomes of the participants resulted in a total of 14,062 live births, with 13,988 children surviving their first year. The sample is broadly representative of the early 1990s UK population of mothers with children under age one, although higher socio-economic status groups as well as Whites are over-represented (see Fraser *et al.*, 2013; Boyd *et al.*, 2013, for a full description of the cohort profile). The study includes detailed information about the family environment, as well as indicators of child development, wellbeing and skills over time, and rich information on the parents' characteristics and background.[1] Biological samples from the children and their parents were collected at different points in time, allowing for DNA genotyping. We here use imputed genotype data from around 9,000 children and their mothers (Taylor *et al.*, 2018,

[1]The study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool: http://www.bristol.ac.uk/alspac/researchers/our-data/.

provide technical details on the genotyping technology, imputation, and quality control in ALSPAC).

When the child was aged 8 months and 2, 3, 4, 5, 6 and 9 years, their mothers were asked whether they had experienced depression since the last interview in which they were asked about their health (or since the birth of the child the first time this question was asked). Although the wording of the question changed slightly across waves, the potential responses were the same: "Yes and consulted a doctor", "Yes but did not consult a doctor" and "No". We consider a mother to have had an episode of depression between two periods if she replied "Yes and consulted a doctor" or "Yes but did not consult a doctor". We combine these seven reported depression scores to produce an index of reported maternal depression from the child's birth to the child's ninth birthday, with index values running from zero to seven.

Our child non-cognitive skill measures come from the Strengths and Difficulties Questionnaire (SDQ) (as used in Flèche, 2017; Briole, Le Forner and Lepinteur, 2020; Clark, D'Ambrosio and Barazzetta, 2021). The SDQ is a 25-question behavioural-screening tool for children, including questions on whether the child is considerate of others, and her concentration span, worries and fears, degree of obedience, and social isolation (Goodman, 1997). The full list of the SDQ items appears in Appendix Table 1.A1. The main carer (this is the mother in the vast majority of cases) was asked to rate the child's SDQ seven times between child ages 4 and 16. We will relate maternal depression during the child's first 9 years to the child's subsequent SDQ scores at ages 11, 13 and 16.

The 25 SDQ items are split up into five sub-scales covering emotional problems, peer problems, conduct problems, hyperactivity/inattention and pro-social behaviour. Consistent with Goodman, Lamping and Ploubidis (2010) and the SDQ scores produced by ALSPAC, our main analysis will use the total SDQ score, which is the sum of the first four sub-scales. We code total SDQ so that higher values represent better outcomes (i.e. strengths rather than difficulties). In the robustness checks (Section 1.4.4.2), we will consider additional non-cognitive skill measures to test for convergent validity (teacher-reported SDQ scores, and an alternative measure of non-cognitive skills from the Short Moods and Feelings Questionnaire, SMFQ, reported by the main carer).

Child cognitive development is measured by their national exam results in linked administrative data from the UK National Pupil Database. We use the average Key Stage fine-grading test-scores at ages 11 and 14 and the total GCSE score in all of the exams that the child took at the end of compulsory education at age 16.[2]

[2] At the end of Key Stages 2 (age 11) and 3 (age 14), children's progress in Mathematics, Science and English is assessed using National Curriculum tests. While these tests do not produce exit certificates, the national exams taken at the end of Key Stage 4 (age 16), the General Certificate of Secondary Education (GCSE), do produce qualification certificates. Students in the UK typically take at least 5 GCSEs (one per

The next section first sets out the principle of using genotype data as an instrument, and then describes the way in which we will apply this method using ALSPAC data.

1.3 A Genetic Instrumental-Variable Approach

1.3.1 Mendelian Randomisation

Establishing causality in non-experimental data is very often challenging, and particularly so for variables that are unlikely to be targeted by policies or be subject to quasi-experimental variation. One recent approach in Social Sciences and Epidemiology is Mendelian Randomisation (MR). This term refers to Mendel's Laws of Segregation and Independent Assortment, which are involved in the formation of reproductive cells (i.e. gametes) through meiosis and which ensure genetic variability across individuals. Traits that are regulated by one gene are defined by a sequence of two alleles (one inherited from each parent); the Law of Segregation states that each individual has a 50% chance of inheriting one of the two maternal (paternal) alleles for a given gene. The Law of Independent Assortment, on the other hand, ensures that alleles for different traits are passed on independently of each other.[3] As a result, conditional on the parental genotypes, the child's genotype can be seen as the outcome of a lottery.[4]

MR in practice refers to a variety of different approaches, the common denominator being the use of genetic variants as instrumental variables for a given endogenous trait (see Koellinger and de Vlaming, 2019; Hemani, Bowden and Davey Smith, 2018, for reviews of some recent developments). While some traits can be linked to a clear small set of genetic variants through well-characterised biological pathways (this is the case for severe health problems, such as Huntington's disease), most traits that interest economists and other social scientists (e.g. socio-economic status, education, and subjective well-being) are highly polygenic and, as such, involve a greater degree of genetic complexity. The burgeoning literature on large-scale GWAS, which aims to estimate the relationship between a given trait and known genetic variants (typically Single-Nucleotide Polymorphisms, or SNPs) in large samples, has brought about significant advances in the understanding of the genetic

subject), with Mathematics, Science and English being compulsory.

[3]The Law of Independent Assortment does however come with a caveat: genes that are close to each other on a chromosome strand have a higher chance of being transmitted together. This leads to what is known as *linkage disequilibrium*: in a given population, alleles for different genes have higher association rates than those that would be expected from random matching.

[4]Note that if parents were to match to each other independently of the trait that a given genotype regulates, we would not even need to condition on parental genotypes for the genotype of the child to be a random draw from the population genetic pool.

architecture of genetically-complex traits such as education (Lee *et al.*, 2018; Demange *et al.*, 2021), depression (Okbay *et al.*, 2016; Turley *et al.*, 2018) and risk behaviour (Karlsson Linnér *et al.*, 2019).

One issue with the use of genetic variants of complex traits as instrumental variables is weak instruments, as each single SNP identified in a GWAS likely has only relatively little predictive power on its own. Polygenic scores have then come into widespread use as linear combinations of all of the relevant genetic markers into one synthetic measure (Appendix 1.B provides more details on the PGS and its functional form), capturing a greater portion of trait variance as compared to single SNPs (DiPrete, Burik and Koellinger, 2018; Davies *et al.*, 2015).

We now consider the various relationships between maternal genes and her child's outcomes, and how these can be addressed to establish a plausible causal relationship.

1.3.2 Instrumental Variables Assumptions in the Context of Genetic Instruments

While others have laid down the assumptions for drawing inference from genetic instruments within the same individual (notably Von Hinke *et al.*, 2016), we here consider instrumentation between parent and child, as illustrated by the solid black lines in Figure 1.1. We aim to measure the causal effect of a mother's trait D_M on her child's outcome Y_C (i.e. the value of the parameter β), where G_M^D is a vector of independent genetic variants of the mother that are robustly associated with this trait D_M. In the ALSPAC analysis that we undertake here, D_M is maternal depression between child birth and child age nine, Y_C the adolescent-child's human capital, and G_M^D the mother's PGS for depression, based on the 88 most-relevant SNPs (p-value threshold of $10^{\breve{}6}$) derived from the single-trait meta-analysis in Turley *et al.* (2018). The results throughout the paper are robust to the use of a more-stringent threshold, identifying what are called genome-wide significant SNPs, with a p-value threshold of 5×10^{-8}.

Just as in a standard instrumental variables (IV) analysis, the validity of the identification strategy relies on the following assumptions:

- *Relevance*: the genetic variants G_M^D are correlated with the trait D_M ($\alpha \neq 0$).
- *Independence*: the G_M^D are not correlated with any confounders (U) of the association between the mother's trait and the child outcome ($\eta = 0$).
- *Exclusion restriction*: the G_M^D are causally related to the outcome Y_C only through the trait D_M (so that β is not confounded by any of the dashed grey lines in Figure 1.1).

The relevance assumption, while being the easiest to prove in most IV contexts, is particularly straightforward in the case of genetic instrumental variables. The task of identifying which genetic variants are robustly associated with a given trait is typically left to summary data from published GWAS (see Appendix 1.B for further details). As noted above, single genetic variants per se might not be sufficiently strong predictors of a trait, especially when the latter is genetically-complex. In these cases, it is more appropriate to use synthetic measures such as the PGS to avoid weak-instrument problems.

The independence assumption is typically assumed to hold in the context of MR due to the randomness of genetic variants, with very few exceptions suggesting otherwise(e.g. Koellinger and de Vlaming, 2019). It is worth underlining, however, that the mother's genotype can be considered as truly random only when conditioning on the maternal grandparents' genotype. In practice, for data-availability reasons, it is seldom possible to partial out the genes of the mother's parents when analysing G_M^D. Some common established good practices in MR analyses are controlling for population stratification[5] and documenting the absence of systematic correlations between the instrument and observable confounders (Smith *et al.*, 2007; Boef, Dekkers and Le Cessie, 2015).

In the context of multivariate regressions, controlling for a selected set of grandparental traits, as well as environmental characteristics, should also attenuate the concerns regarding the independence assumption. Consider, as an illustration of U in Figure 1.1, the potential influence of grandparental depression. Depressed grandparents are first more likely to have genetic variants associated with depression: via genetic inheritance, their daughters will then also likely display a higher PGS for depression (in Figure 1.1 this would translate into $\eta \neq 0$). In addition, grandparental depression may increase the chances of their daughter's depression through non-genetic pathways, e.g. by increasing familial stress and anxiety (this is represented by the line from U to D_M in Figure 1.1). Last, grandparental depression can affect child outcomes directly, as depicted in the line from U to Y_C : this could reflect, for example, the crowding-out effect of the time that mothers with depressed parents can dedicate to their children. As they may simultaneously affect all of the variables of interest (via the three unbroken grey lines in Figure 1.1), not controlling for grandparental genes and/or their associated traits can violate the independence assumption.[6] Introducing controls for the

[5]This stratification reflects drifts in allelic frequencies within the population of interest. A popular solution, which is particularly well-suited in contexts of considerable geographical and ethnic diversity, is controlling for principal components derived from genotyped data. These account for systematic associations between the alleles in subsets of a given population that are produced, among other things, by within-group assortative-matching patterns.

[6]While the grandparents' non-transmitted alleles can be unobserved confounders of the D_M-Y_C association (through genetic nurture, as they likely affect the way in which the grandparents bring up the mother), they

depression of both of the grandparents, as well as for other grandparental traits and environmental characteristics, can attenuate the bias in this case.

The exclusion restriction is well-known to be the most problematic assumption in all IV setups, and this is particularly true in the context of MR (Koellinger and de Vlaming, 2019). In our mother-child framework, phenomena such as horizontal pleiotropy and genetic inheritance can link G_M^D and Y_C through pathways other than D_M.

Under horizontal pleiotropy, an individual's genetic variant directly affects two or more of her traits through separate biological pathways (e.g. the genetic variants causing maternal depression might also affect other maternal traits, such as educational attainment). This will pose identification problems if these additional traits affected by G_M^D (the X_M in Figure 1.1) are correlated with the outcome of interest ($\gamma_1 \neq 0$).

One simple way to account for the confounding effects of the X_M in Figure 1.1 is to control for them. While this may sound rather trivial, most MR applications are actually bivariate associations (accompanied, in most cases, by statistical tools to account for pleiotropy), typically due to data limitations. We do of course need to be careful when controlling for maternal traits: while some might indeed capture part of the observable pleiotropic effects, they can also partly mediate the relationship between D_M and Y_C and, as such, be 'bad controls' (Angrist and Pischke, 2008). In relation to Figure 1.1, holding 'bad controls' constant would lead to the attenuation of the estimated value of β.

Genetic inheritance also poses a problem for identification. Each child inherits 50% of each parent's genetic variants. This produces the path from G_M^D to G_C^D, the child's genetic variants that are associated with child trait D, in Figure 1.1. There are then two pathways from G_C^D to the child's outcome. The first is the direct biological pathway from G_C^D to Y_C ($\delta_1 \neq 0$); the second is due to vertical pleiotropy ($\gamma_2 \neq 0$), i.e. the effect of G_C^D on Y_C that is mediated by one or more child traits (X_C).

The issues around genetic inheritance might not only concern the transmission of the genetic variants for depression. The child's genetic variants explaining Y_C, G_C^Y may also partly be inherited from the mother's G_M^D and/or result from linkage disequilibrium (LD from here onwards; see footnote 3 for the definition) with it ($\delta_2 \neq 0$). This is important here, as mental health and cognitive achievement partly share the same genetic aetiology (see Rajagopal *et al.*, 2020). Similarly to G_C^D, the vector G_C^Y can affect the child outcomes either directly or via vertical pleiotropy ($\gamma_2 \neq 0$).

do not play a role in the independence assumption of the effect of G_M^D on Y_C, as they cannot appear in the mother's PGS for depression (Mendel's Law of Segregation).

Were the genetic data detailed enough, we could deal with all the concerns arising from genetic inheritance by using only the mother's non-transmitted alleles as instruments: mechanically, there would then be no correlation between the mother's and the child's genotypes (unless there is assortative matching between the parents over trait D and/or Y). Another possibility, which is what we do here, is to control for the child's genotypes, so as to hold constant all of the pathways between G_C^D, G_C^Y and Y_C. We will in addition control for the child's traits, X_C, in case these partly result from genetic variants other than G_C^D and G_C^Y. We are to the best of our knowledge the first to be able to control for both the child's PGS for depression, G_C^D (Lawlor *et al.*, n.d.), and cognitive (non-cognitive) skills, G_C^Y (DiPrete, Burik and Koellinger, 2018),[7] in the empirical analysis. Note that once we have controlled for the relevant child PGSs, the residual part of the bivariate pathway between X_M and Y_C in Figure 1.1, γ_1, reflects genetic nurture (Kong *et al.*, 2018), i.e. the effect of the maternal traits caused by the part of G_M^D that is not inherited by the child.

We next describe the equations that to be estimated using ALSPAC data.

1.3.3 Empirical Strategy

We address endogeneity by estimating the following Two-Stage Least Squares (2SLS) regressions using a genetic instrument:

$$D_M = \alpha_1 PGS_M^D + \alpha_2 X_M + \alpha_3 X_C + \alpha_4 PC_M + \epsilon_M \quad (1.1)$$

$$HK_{Ct} = \beta_1 D_M + \beta_2 X_M + \beta_3 X_C + \beta_4 PC_M + v_{Ct} \quad (1.2)$$

In Equation 1.1, D_M is the number of self-reported episodes of maternal depression, from the child's birth up to age nine, taking on values from 0 to 7. In Equation 1.2, the outcome HK_{Ct} is successively different measures of child C's human capital at age t: the fine-grading average Key Stage test-scores at ages 11 and 14, the total GCSE score at age 16, and total (carer-reported) SDQ at child ages 11, 13 and 16. We standardise the different HK_{Ct} variables for comparison purposes, as they are not measured on the same scale.

We address pleiotropy by controlling for a set of both mother and child traits. X_M is a vector of

[7] Note that controlling for G_C^Y is what DiPrete, Burik and Koellinger (2018) refer to as Unconditional Genetic Instrumental Variables (GIV-U), that is simply controlling for all the genetic variants associated with Y_C (i.e. G_C^Y). In particular, they show that GIV-U regression provides a reasonable lower bound for the true effect of D_M on Y_C under several violations of the IV assumptions (either through moderate pleiotropy or other genetic confounds).

mother's traits when the child is aged nine: age at the birth of the child and dummies for being employed, having at least an A-level,[8] having a partner, having a partner with at least an A-level, having an employed partner, the number of additional children, and banded household income (the latter two variables are measured at child age 8). It can be argued that some of these are potentially bad controls, as they may themselves be influenced by maternal depression (for example, mother's labour-force status and the household's income). We will address this issue in the robustness checks. Last, the X_C are time-invariant child traits: gender, birth year and birth order.

In Equation 1.1, PGS_M^D, the maternal polygenic score for depression (our measure of mother's genetic variants, G_M^D, in Figure 1.1) is used as an instrument for maternal depression D_M. We calculate PGS_M^D using the command-line program PLINK 1.9, with summary statistics from the single-trait depression meta-analysis GWAS in Turley *et al.* (2018). 68 of the 88 SNPs identified in the GWAS are genotyped in ALSPAC participants and were used in the PGS: see Appendix 1.B for the details of the calculation. We standardise PGS_M^D, as polygenic scores have no natural scale. With polygenic scores being based on genetic variants that are determined at conception, the exogenous variation in maternal depression provided by PGS_M^D is fixed prior to the child's birth: this rules out reverse-causality concerns (e.g. mothers' mental health being affected by their children's poor cognitive and/or non-cognitive performance).

We address population stratification by excluding mothers of non-European descent: Hansell *et al.* (2015) find no evidence of any remaining population stratification in ALSPAC after this selection and other standard quality-control (QC) procedures (see Taylor *et al.*, 2018, for a complete overview of the QC procedures that were applied to ALSPAC data prior to its release). While the documented lack of stratification provides evidence in favour of the independence assumption in our context, we always control for 10 ancestry-informative principal components PC_M (as in Von Hinke *et al.*, 2016) and carry out additional tests for the influence of grandparental characteristics and partners' depression on the effect of maternal depression (see Section 1.4.3).

Our estimation sample consists of observations with non-missing values for mothers' genetic information, depression history, and the controls measured at child age nine.[9] As there are only

[8]An A-level (Advanced Level) qualification is a subject-based school-leaving certificate that is typically obtained at the end of Upper-Secondary School at around age 18.

[9]The mother's traits that are measured at child age eight (household income and the number of children in the household) are missing in roughly 9% of the cases in our different estimation samples. Where the respondents have missing information, we create a variable-specific dummy to flag this missing information (the Missing Indicator method) and replace the missing value by the sample mean. We in addition drop the 32 cases with multiple births for the mother over the 18-month initial survey period (although our results are robust to including these observations).

1,065 families with non-missing information on all six human-capital measures and we cannot reject concerns about weak instruments in this balanced sample, we here use a different estimation sample for each dependent variable to maximise statistical power. Our final samples consist of between 2,036 and 2,993 observations per equation estimated. Due to attrition, the size of the non-cognitive skills estimation samples falls naturally with child age (from 2,993 to 2,076 observations). For cognitive skills, the estimation samples consist of 2,828 observations at age 16 (GCSE), 2,601 observations at age 11 (KS2) and 2,036 at age 14 (KS3). The discrepancy between the sample sizes at age 16 and earlier child ages reflects that the average KS2 and KS3 grades are retrospectively matched when the child takes her GCSE exams at age 16. 10% of the 227-observation difference between the GCSE and KS2 samples is due to either missing values in the school and academic year identifiers or in the grades, while the remaining 90% is due to the NPD data-cleaning process. For the gap between the GCSE and KS3 samples, 258 observations are missing for these two reasons, while the remaining 534 are due to the KS3 grades of ALSPAC children taking their GCSE in academic year 2008-09 no longer being collected.[10] The influence of maternal depression and PGS_M^D on attrition and the different sample sizes is discussed in the next section.

The distribution of self-reported maternal depression in the different estimation samples appears in Table 1.1, where depression takes on values between zero and seven. Around half of the women in our samples reported at least one episode of depression between the birth and ninth birthday of their child. This figure is consistent with data from a nationally-representative survey, the British Household Panel Study (BHPS), over the same time period: around 45% of the mothers observed for at least two consecutive years from 1991 to 2000 reported a least one episode of depression. The distribution of the measures of children's human capital is shown in Appendix Figure 1.A1, and the complete descriptive statistics are listed in Tables 1.A2 (cognitive skills) and 1.A3 (non-cognitive skills).

While Equation 1.2 partly addresses pleiotropy by controlling for both the mother's and the child's traits (X_M and X_C), the maternal PGS may still be directly linked to child human capital via the child's genome (G_C^D and G_C^Y in Figure 1.1). We here follow Lawlor *et al.* (n.d.) and DiPrete, Burik and Koellinger (2018), and address these concerns by controlling for the child's PGS for depression and, respectively, cognitive and non-cognitive outcomes (see Section 1.4.2.3).

[10]Technical details about the NPD cleaning process and the collection of the KS3 average grades we use here can respectively be found at http://www.bristol.ac.uk/media-library/sites/cmpo/migrated/documents/ks5userguide2011.pdf and https://find-npd-data.education.gov.uk/en/data_elements/11e50a8a-78d6-425c-871d-9d9fd3330dd9.

1.4 Results

1.4.1 Main Results

Table 1.2 presents the OLS and 2SLS estimates of Equation 1.2 for the effect of maternal depression on the different measures of child human capital. All of the estimated coefficients are negative and significantly different from zero at the 10% level at least. In the 2SLS results in columns (2), (4) and (6), one additional episode of maternal depression before child age nine reduces child test-scores by on average 23% of a standard-deviation (SD) and total SDQ by roughly 45% of a SD.[11] Although the 2SLS estimates become a little larger as the child grows older, none of them are significantly different from each other. This pattern does not reflect the different sample compositions: restricting our analysis to families with valid information on either all of the cognitive-skill measures or all of the non-cognitive skill measures yields similar conclusions (these results are available upon request). Our specification exploits the longitudinal dimension of the dataset by looking at the impact of the observed history of a mother's depression on the subsequent cognitive and non-cognitive outcomes of her children. Our estimates may thus reflect the predictive effect of the PGS on unobserved later episodes of maternal depression occurring between child age nine and the time the child's outcome of interest is observed. Maternal depression during a child's puberty could have a greater impact on their schoolwork and behaviour, producing larger coefficients at ages 14 and 16. In either case, maternal depression produces worse child outcomes.

Instrument relevance is evaluated in the first-stage estimates below the 2SLS results in Table 1.2. As expected, a higher PGS for depression significantly predicts more maternal-depression episodes in all specifications (at the 0.1% level at least). We also list the Cragg-Donald Wald F-statistics for the first-stages, which are sufficiently large to alleviate weak-instrument concerns in most cases.[12] This F-statistic is under 10 only for the effect of maternal depression on total SDQ at age 16, which may show selective attrition. The probability of dropping out of the total SDQ estimation sample between two periods rises with maternal depression, but does not depend on the value of the instrument. It is thus unsurprising to see a lower first-stage F-statistic in the last column of the bottom panel of

[11] The reduced-form estimates for the PGS for depression range from -0.036 to -0.051 SD for cognitive skills and from -0.071 to -0.063 SD for non-cognitive skills, with significance levels identical to those in our baseline 2SLS estimates. While reduced-form estimates rely on weaker assumptions, they come at a cost in terms of interpretation, as they do not identify a mediating trait in the maternal genes - child outcome relationship. Under the assumptions described in Section 1.3.2, our 2SLS estimates reveal that the causal effect of the PGS for depression of the mother on the human child capital of children is only mediated by maternal depression.

[12] We cannot make strong statements about whether the effect of maternal depression differs by gender, birth-order, maternal education and household-income band, as the smaller samples produce F-statistics that are mainly too low for robust inference.

Table 1.2.[13]

Columns (1), (3) and (5) show the OLS results. Although these are qualitatively similar to the 2SLS estimates, they are four to ten times smaller in size. This gap may reflect that the GWAS summary statistics from Turley *et al.* (2018) are based on discovery samples where the trait is mostly measured as clinically diagnosed depression or self-diagnosed major depressive disorder (in around 80% of cases). As such, it is normal that the 2SLS estimates be larger than those in OLS, as the instrument captures more extreme forms of depression, that in turn play a larger role in human-capital accumulation. When analysing a non-binary trait (like our measure of maternal depression) genetic compliers can be seen as the whole population (see Dixon et al., 2020). Our 2SLS estimates then capture the average treatment effect of the trait that is most prevalent in the GWAS discovery cohorts – that is, 'severe' forms of depression (clinically-diagnosed depression, or major depressive disorder). On the contrary, the OLS estimates reveal the average effect of all forms of depression, both mild and severe.

The difference between the OLS and 2SLS estimates is larger for cognitive than non-cognitive skills: this may reflect the relative importance of diagnosed and undiagnosed symptoms of maternal depression in these two dimensions of human capital. While we do not observe formal diagnoses of depression, we know whether the mother consulted a doctor due to her depressive symptoms. When we separately consider episodes of maternal depression that were followed up by a medical visit and those that were not, the descriptive evidence from the OLS estimates suggests that, while both measures matter equally for non-cognitive skills, only the former is significantly associated with child cognitive skills (results available upon request).[14]

[13]Note that neither maternal depression nor the instrument predict retrospective attrition for cognitive skills in the top panel of Table 1.2. As information on Key Stages 2 and 3 (child ages 11 and 14, respectively) test-scores are obtained retrospectively, attrition here is the probability of being in the age-16 sample for cognitive skills and being absent from, respectively, the analogous age-11 and age-14 samples. The results on attrition in the cognitive and non-cognitive samples are available upon request.

[14]It might be thought then that we would be better-off restricting our analysis only to episodes of maternal depression that are followed by a medical consultation. When doing so, we find coefficients that are on average twice as large as the baseline 2SLS estimates from Table 1.2 (all significant at least at the 10% level). However, the F-statistics for episodes of maternal depression followed by a medical visit take values that are systematically lower than those in Table 1.2. Using only depressive episodes followed by a medical visit comes at a greater risk of weak-instrument issues.

1.4.2 Addressing the Exclusion Restriction

1.4.2.1 Horizontal Pleiotropy

The credibility of the exclusion restriction relies on there being no relationship between the PGS for maternal depression and the child outcomes, other than via maternal depression. However, as set out in Section 1.3.2, a genetic variant may predict more than one trait: this is horizontal pleiotropy. While we already control for a set of maternal traits in our main specification, we here provide additional evidence against pleiotropy playing a significant role in our analysis. Table 1.A4 in Appendix 1.A shows the bivariate associations between the PGS for depression and a variety of maternal traits. Unsurprisingly, the association between the PGS and maternal depression is positive and very significant. Just as importantly, none of the other traits is significantly associated with this instrument. While we cannot entirely rule out an effect of the genetic variants in the mother's PGS on other unobserved traits involved in child human-capital development, the lack of any correlation with the observed traits is reassuring.

We also address the risk of pleiotropy more directly, investigating the known biological functions that are linked to the 68 SNPs used in the mother's PGS for depression. We do so using the NHGRI-EBI online GWAS Catalog to review all of the biological functions associated with our SNPs. In line with Von Hinke *et al.* (2016), we then calculate a new PGS discarding the six lead SNPs linked to either the cognitive or non-cognitive outcomes,[15] as these are likely to violate the exclusion restriction via their effect on the mother's human capital. Columns (2), (4) and (6) of Table 1.A5 list the 2SLS estimates with this restricted PGS: these are very similar to those in the baseline (reproduced in columns (1), (3) and (5)). We also calculate the mother's PGS for depression excluding the lead SNPs that predict any trait other than depression, even those that may appear unrelated to human capital (e.g. bone density). The last two sets of mother's PGS exclude the SNPs in LD with genetic variants explaining other traits (first, only the cognitive and/or non-cognitive outcomes, and second an expanded set of traits made up of these two outcomes, along with BMI, and smoking), using a window of 500k base-pairs and a squared pairwise correlation of at least 0.6. Although both approaches reduce the variability in our instrument on which identification is based, the 2SLS estimates remain qualitatively the same. These results are available upon request.

We last address unobserved associations between the SNPs for depression and mother's human capital (e.g. unknown biological pathways) by computing her PGS for both cognitive and non-cognitive skills based on the GWAS summary statistics in Demange *et al.* (2021), and introducing

[15] These are the following: rs10514301, rs10789340, rs10045971, rs11876620, rs12958048, and rs174548

these as controls in our main specification. Table 1.A6 shows that partialling out maternal genetic variation in cognitive and non-cognitive skills does not qualitatively change the results (and these latter are mostly not significant predictors of maternal depression in the first-stage regressions).

1.4.2.2 Bad Controls

As discussed in Section 1.2, controlling for mother's and child's traits attenuates pleiotropy concerns. It can nonetheless be argued that some of these traits (for example, mother's labour-force status, the presence of a partner in the household and household income) are bad controls as they could themselves result from depression. We thus re-estimate our 2SLS regressions first with no controls, then controlling for the mother's traits, and finally for the child's traits. The results, as compared to the baseline estimates (which control for both sets of traits), are depicted in Figure 1.2. The inclusion of potentially 'bad' controls makes relatively little difference, and the estimated coefficients on maternal depression remain negative and significant in every specification for every outcome.

1.4.2.3 Genetic Inheritance and Trait Overlap

We can expect about half of the genetic variants included in the PGS for maternal depression to be passed on to the child (and an even higher figure if the parents match assortatively on the basis of depression). As noted in Section 1.2, if the inherited variants are correlated with the child's cognitive/non-cognitive outcomes, then the exclusion restriction will be violated. Controlling for the child's polygenic scores for depression and cognitive/non-cognitive skills will effectively shut off any confounding effect from genetic inheritance that affects these traits.

We here again use the summary statistics from the depression meta-analysis GWAS in Turley *et al.* (2018) to calculate a PGS for depression in children. We use the GWAS-by-subtraction summary data from Demange *et al.* (2021) and the summary statistics from the genome-wide association meta-analysis of Middeldorp *et al.* (2016) to calculate the PGSs for cognitive and non-cognitive skills. We do not use Demange *et al.* (2021) to calculate the PGS for non-cognitive skills, as the weights from a GWAS on an adult population might not be relevant for children (see Zhang *et al.*, n.d.). Furthermore, as Demange *et al.* (2021) define as 'non-cognitive' those SNPs associated with educational attainment independent of cognitive ability, a PGS based on their summary statistics may not be appropriate for our measure of non-cognitive skills (SDQ). In contrast, Middeldorp *et al.* (2016) use a discovery sample of children under 13 to identify the SNPs associated with attention-deficit/hyperactivity disorder (ADHD) symptoms (a condition arguably captured by the 'inattention/hyperactivity' subscale of the SDQ).[16]

Columns (1), (5) and (9) in Table 1.3 show the baseline 2SLS estimated coefficients for maternal depression from Table 1.2; the other columns introduce various child PGS measures. As the child genotype is missing in roughly 10% of the cases, we replace the missing values with the sample average and use a missing-indicator flag (dropping missing-genotype children from the estimation produces similar results). Columns (2), (6) and (10) in the top panel of Table 1.3 control for the child's cognitive-skill PGS: this is positively correlated with the child's average test-scores (as expected), but not with the PGS for maternal depression (there is little change in the F-statistics). Analogous results pertain for the child's non-cognitive PGS in the bottom panel of Table 1.3.

We then turn to the child's depression PGS, part of which is inherited from the mother. As expected, we find a 50% unconditional correlation between the mother's and the child's PGSs for depression (which explains the lower F-statistics in columns (3), (7) and (11) when controlling for the latter). However, in regressions including the mother's PGS, the child's depression PGS does not significantly influence the dependent variables (as shown in the table) or maternal depression (not reported – results available upon request). As the PGS uses weights derived from an adult population, the genetic variants identified there may not work in the same way for children.

Columns (4), (8) and (12) introduce the two scores simultaneously, which does not change our conclusions: the children of more-depressed mothers have significantly worse cognitive and non-cognitive skills. Although the estimated maternal-depression coefficients change a little in size as we introduce different PGS controls, they are never significantly different from each other.[17]

1.4.2.4 Plausible Exogeneity

While the analyses above have put considerable effort into tackling potential violations of the exclusion restriction, there may still be unobserved pathways for which we do not control. For instance, while we do account for horizontal pleiotropy from the mother's genetic variants by

[16] Note that the discovery sample of Middeldorp *et al.* (2016) includes the ALSPAC cohort. We also used alternative summary statistics from other GWAS (Benke *et al.*, 2014; Pappa *et al.*, 2016; Demange *et al.*, 2021) to calculate alternative polygenic scores for non-cognitive skills, but none of these significantly correlates with total SDQ other than that from Middeldorp *et al.* (2016). These results are available upon request.

[17] Another way of ruling out confounding genetic-inheritance effects is to recalculate the PGS for maternal depression excluding the genetic variants that are also associated with children's cognitive and non-cognitive skills (either directly or through LD patterns). Out of the 68 top variants for maternal depression genotyped in ALSPAC, we find that none coincides with top variants for cognitive skills, while fourteen others are in LD with at least one cognitive top variant. In contrast, we find no overlap with the eight main genetic variants for non-cognitive outcomes (as measured by ADHD). The results, available upon request, remain qualitatively unchanged.

controlling for a set of maternal covariates, there are still channels we do not observe or, if observed, are subject to measurement error and reporting bias. Additionally, although their impact is likely to be marginal, there might be yet some other sources of pleiotropy confounding our main estimates (see, for instance, network pleiotropy in Boyle, Li and Pritchard, 2017).

We thus follow the analysis in Conley, Hansen and Rossi (2012), and consider the implications of our instrument being only 'plausibly exogenous'. Here the instrumental variable is allowed to have a direct effect, λ, on the outcomes. As in Nybom (2017), λ is the share of the reduced-form effect of the instrument on child human capital that is independent of the variable we instrument, maternal depression. Considering different values of λ allows us to identify the threshold at which our 2SLS estimated coefficients are no longer significant at the 10% level.

Figure 1.A2 depicts the 2SLS estimates from Equation 1.2 for λ in the interval $[0, 1]$. We follow Nybom (2017) and assume that λ is known with certainty. For cognitive skills at ages 11 and 14, once λ reaches 0.1 the 2SLS estimates are no longer significant at the 10% level (as revealed by the grey shaded areas). For all other outcomes, the threshold is larger (λ from 0.3 up to 0.5). In other words, as long as the direct effect of the PGS for maternal depression on the child outcomes is under 30% of the total reduced-form effect, most of our 2SLS estimates remain significantly different from zero at the 10% level.[18]

1.4.3 The Influence of Maternal Grandparents and the Partner

Based on the ethnic composition of our subsample of ALSPAC participants and the fact that we always control for 10 ancestry-informative principal components, we have little reason to believe that residual population stratification is a threat to the independence assumption (see Section 1.3.3). However, other concerns regarding the independence assumption remain. Mendel's laws of Segregation and Independent Assortment imply that, conditional on the parental genotype, the child's genotype is the result of a lottery. The genotypes of the maternal grandparents are not available in ALSPAC, so that the mother's genotype, and consequently her PGS for depression, might partly capture the effect of her parents' genotypes, with the latter also potentially being correlated with the U variables in Figure 1.1 (see Section 1.3.2).

[18] For the sake of transparency, the dashed grey lines in Figure 1.A2 show the 90% confidence intervals when following the 'local-to-zero' approach described in Van Kippersluis and Rietveld (2018), where λ is assumed to follow a Normal distribution and where there is no subsample for which the first-stage is zero. When we do so half of our baseline estimates, i.e. $\lambda = 0$, are no longer significantly different from zero at the 10% level. Note that Van Kippersluis and Rietveld (2018) apply this method to an estimation sample with over 100,000 observations. With roughly 3,000 observations at best, our estimation samples may well be too small to provide sufficient precision here.

While we cannot control for the genetic variants of the maternal grandparents, we do have data on a set of grandparental traits: their education, social status, and a dummy for at least one of the maternal grandparents having had a severe mental illness prior to the birth of the child. The results controlling for these variables appear in Table 1.A7. The 2SLS estimates are virtually unchanged from those in the baseline. The F-statistics are slightly lower. This is unsurprising: even though, after conditioning on the mother's traits, none of the grandparental characteristics is correlated with child human-capital, the mother having at least one parent with a history of mental illness is positively and significantly associated with both our measure of maternal depression and her PGS for depression.

We finally consider assortative matching between the child's parents: depressed mothers might choose their partners according to certain traits (depression itself, and/or other traits), which may in turn affect child human capital. Our main specification, which includes a number of the mother's partner's controls, partly addresses this. We can further show that these traits (having a partner, partner's working status and education) are not systematically explained by the mother's PGS for depression (see Table 1.A4). While this alleviates concerns about cross-trait assortative matching, mothers with a higher genetic risk of being depressed might be more likely to have a depressed partner. We have information on the mother's partner's number of depression episodes, measured at child ages 2, 4 and 6. While the unconditional correlation between the partner's depression and maternal depression is relatively high (0.44) and significant, its correlation with the PGS for maternal depression is not statistically different from zero (in both bivariate and multivariate analyses). Introducing partner's depression makes little difference to our main results (Table 1.A8).

1.4.4 Robustness Checks

1.4.4.1 The Measurement of Maternal Depression

We carry out a battery of robustness checks. We first show that our results hold with different maternal-depression measures (the descriptive statistics of which appear in Table 1.A9). Our baseline count of reported depressive episodes between child ages 0 and 9 weights recent and more-distant episodes equally, but those at younger child ages may matter more (as children then have greater developmental plasticity and spend more time with their mothers). Panels B and C of Table 1.A10 however reveal larger estimates for more-recent depressive episodes (although the estimated coefficients between these panels are not significantly different from each other). The results continue to hold using only the number of episodes net of post-partum depression (i.e. between child ages 2 and 9) in Panel D, and with a dummy for any episode of depression in Panel E. Panel F considers a dummy for recent depression and Panel G the average of the six maternal scores on the Edinburgh Postnatal Depression Scale (EPDS) between child ages 0 and 8 (at child age 8 months and 2, 3, 5, 6 and 8 years). Although the results continue to be of the same nature, the F-statistics are notably worse. The instrument weakness here reveals that our PGS has greater predictive power when maternal depression is measured over longer time periods and in a similar way to that in the GWAS meta-analysis (the EPDS does not appear in Turley *et al.*, 2018).

1.4.4.2 The Measurement of Non-Cognitive Skills

The SDQ measure of non-cognitive skills we use is reported by the mother. As depressed mothers may over- or under-estimate their children's non-cognitive skills (Del Bono, Kinsler and Pavan, 2020) we turn to teacher-reported SDQ (which is only available when the child was aged 11). In the first column of Table 1.A11, an additional episode of maternal depression continues to reduce total SDQ with an effect size identical to that in Table 1.2.[19] We also test for convergent validity using the SMFQ (reported by the main carer) in columns (2) to (4) of Table 1.A11: the resulting estimates are not significantly different from those in the baseline (although that at age 16 is statistically insignificant). See Table 1.A12 for a complete list of the items in the SMFQ questionnaire.

1.5 Conclusion

Social scientists are interested in causal phenomena, and research agendas are partly limited to the analysis of variables that can be influenced, either directly or via policy intervention. However, there are many variables and pathways that are either costly or impossible to manipulate. We believe that it is possible to make causal statements about some of these latter via the increasing availability of genetic data and recent developments in the fields of Epidemiology and Molecular Genetics. This is the approach that we have taken here. However, the use of genetic data as instruments is not a quick fix, as it comes with a number of quite-stringent assumptions. We have here discussed a number of tests and tools that can be applied in this empirical setting.

We illustrate how genetic data can be used to identify the effect of maternal depression on children's human capital, using data from a British birth-cohort study. We first show that genetic

[19] Total SDQ can be split into two finer subscales: internalising SDQ (emotional health: the sum of 'peer problems' and 'emotional problems') and externalising SDQ (behavioural issues: the sum of 'hyperactivity/inattention' and 'conduct problems'). Maternal depression produces worse outcomes for both internalising and externalising SDQ. These results are available upon request.

variants, combined into a synthetic polygenic score, are a strong instrumental variable for maternal depression. In 2SLS estimation, we then exploit the exogenous differences in maternal depression resulting from the mother's genes to identify its negative consequences on the cognitive and non-cognitive outcomes of their adolescent children.

Our results suggest that fewer episodes of maternal depression will not only benefit mothers, but also improve their children's human capital. In turn, better cognitive and non-cognitive skills in childhood are known to have positive returns on a variety of outcomes during adulthood, such as income and labour-market experience (Heckman, Stixrud and Urzua, 2006; Heckman, Humphries and Veramendi, 2018; Clark, 2018; Clark and Lepinteur, 2019). As revealed by the evaluation of the Improving Access to Psychological Therapies programme in the UK in Clark (2018), the costs of effective treatments for depression are extremely low compared to their expected benefits. If treatment also produces positive spillovers on children, the benefit-cost ratio will be even higher, making treatment more attractive.

However, as we compare depressed to not-depressed or less-depressed mothers using cross-section data on adolescents, our results do not tell us how changes in depression (in particular, due to its treatment) would affect children. Baranov *et al.* (2020) find only small long-term effects on child development following the treatment of prenatally-depressed mothers in rural Pakistan. The socio-economic, geographical and temporal contexts of our work and those in Baranov *et al.* (2020) are of course dissimilar. More importantly, they look at mothers who were already depressed pre-birth, whereas we consider a general sample of mothers, some of whom experience episodes of depression after birth and some of whom do not. While we show that the experience of maternal depression has large scarring effects on adolescent children, we do not know how easy it is to erase these scars. Policies that aim to prevent depression, rather than treat it once it occurs, may have a greater return from a societal perspective.

The use of polygenic scores as instrumental variables is a promising avenue for causal inference in observational data. It is however important to keep in mind that the genetic component of complex traits, such as mental health, is far from deterministic. The same polygenic score can be found in individuals with a very wide range of values of the trait of interest. This may reflect that the individual genetic architecture predicts outcomes partly via individuals' reactions to their environment. This opens the door to policy intervention: while genes are fixed, the environment is not. Future research on which stressors are the most important in this context will help advance our understanding of the sign and size of causal relationships that can serve as inputs to public-policy debate.

References

Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Banerjee, Souvik, Pinka Chatterji, and Kajal Lahiri. 2017. "Effects of psychiatric disorders on labor market outcomes: a latent variable approach using multiple clinical indicators." *Health Economics*, 26: 184–205.

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko. 2020. "Maternal depression, women's empowerment, and parental investment: evidence from a randomized controlled trial." *American Economic Review*, 110: 824–59.

Benke, Kelly S., Michel G. Nivard, Fleur P. Velders, Raymond K. Walters, Irene Pappa, Paul A. Scheet, Xiangjun Xiao, Erik A. Ehli, Lyle J. Palmer, Andrew J.O. Whitehouse, *et al.* 2014. "A genome-wide association meta-analysis of preschool internalizing problems." *Journal of the American Academy of Child and Adolescent Psychiatry*, 53: 667–676.

Boef, Anna G.C., Olaf M. Dekkers, and Saskia Le Cessie. 2015. "Mendelian randomization studies: a review of the approaches used and the quality of reporting." *International Journal of Epidemiology*, 44: 496–511.

Boyd, Andy, Jean Golding, John Macleod, Debbie A. Lawlor, Abigail Fraser, John Henderson, Lynn Molloy, Andy Ness, Susan Ring, and George Davey Smith. 2013. "Cohort profile: the 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children." *International Journal of Epidemiology*, 42: 111–127.

Boyle, Evan A., Yang I. Li, and Jonathan K. Pritchard. 2017. "An expanded view of complex traits: from polygenic to omnigenic." *Cell*, 169: 1177–1186.

Briole, Simon, Hélène Le Forner, and Anthony Lepinteur. 2020. "Children's socio-emotional skills: Is there a quantity–quality trade-off?" *Labour Economics*, 64: 101811.

Bubonya, Melisa, Deborah A. Cobb-Clark, and Mark Wooden. 2017. "Mental health and productivity at work: Does what you do matter?" *Labour Economics*, 46: 150–165.

Clark, Andrew E., and Anthony Lepinteur. 2019. "The causes and consequences of early-adult unemployment: Evidence from cohort data." *Journal of Economic Behavior and Organization*, 166: 107–124.

Clark, Andrew E., Conchita D'Ambrosio, and Marta Barazzetta. 2021. "Childhood circumstances and young adult outcomes: The role of mothers' financial problems." *Health Economics*, 30: 342–357.

Clark, David M. 2018. "Realizing the mass public benefit of evidence-based psychological therapies: The IAPT program." *Annual Review of Clinical Psychology*, 14: 159–183.

Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi. 2012. "Plausibly exogenous." *Review of Economics and Statistics*, 94: 260–272.

Cunha, Flavio, and James J. Heckman. 2008. "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation." *Journal of Human Resources*, 43: 738–782.

Dahlen, Heather M. 2016. "The impact of maternal depression on child academic and socioemotional outcomes." *Economics of Education Review*, 52: 77–90.

Davies, Neil M., Stephanie von Hinke, Helmut Farbmacher, Stephen Burgess, Frank Windmeijer, and George Davey Smith. 2015. "The many weak instruments problem and Mendelian randomization." *Statistics in Medicine*, 34: 454–468.

Del Bono, Emilia, Josh Kinsler, and Ronni Pavan. 2020. *Skill Formation and the Trouble with Child Non-Cognitive Skill Measures.* IZA Discussion Papers, 13713.

Demange, Perline A., Margherita Malanchini, Travis T. Mallard, Pietro Biroli, Simon R. Cox, Andrew D. Grotzinger, Elliot M. Tucker-Drob, Abdel Abdellaoui, Louise Arseneault, Elsje Van Bergen, *et al.* 2021. "Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction." *Nature Genetics*, 53: 35–44.

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: E4970–E4979.

Flèche, Sarah. 2017. *Teacher Quality, Test Scores and Non-Cognitive Skills: Evidence from Primary School Teachers in the UK.* CEP Discussion Paper, 1472.

Fletcher, Jason. 2013. "Adolescent depression and adult labor market outcomes." *Southern Economic Journal*, 80: 26–49.

Fraser, Abigail, Corrie Macdonald-Wallis, Kate Tilling, Andy Boyd, Jean Golding, George Davey Smith, John Henderson, John Macleod, Lynn Molloy, Andy Ness, *et al.* 2013. "Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort." *International Journal of Epidemiology*, 42: 97–110.

Goodman, Anna, Donna L. Lamping, and George B. Ploubidis. 2010. "When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers and children." *Journal of Abnormal Child Psychology*, 38: 1179–1191.

Goodman, Robert. 1997. "The Strengths and Difficulties Questionnaire: a research note." *Journal of Child Psychology and Psychiatry*, 38: 581–586.

Goodman, Sherryl H., Matthew H. Rouse, Arin M. Connell, Michelle Robbins Broth, Christine M. Hall, and Devin Heyward. 2011. "Maternal depression and child psychopathology: A meta-analytic review." *Clinical Child and Family Psychology Review*, 14: 1–27.

Gotlib, Ian H., Peter M. Lewinsohn, and John R. Seeley. 1998. "Consequences of depression during adolescence: Marital status and marital functioning in early adulthood." *Journal of Abnormal Psychology*, 107: 686–690.

Gotlib, Ian H., Sherryl H. Goodman, and Kathryn L. Humphreys. 2020. "Studying the intergenerational transmission of risk for depression: Current status and future directions." *Current Directions in Psychological Science*, 29: 174–179.

Hakulinen, Christian, Marko Elovainio, Martti Arffman, Sonja Lumme, Sami Pirkola, Ilmo Keskimäki, Kristiina Manderbacka, and Petri Böckerman. 2019. "Mental disorders and long-term labour market outcomes: Nationwide cohort study of 2,055,720 individuals." *Acta Psychiatrica Scandinavica*, 140: 371–381.

Hansell, Narelle K., Graeme S. Halford, Glenda Andrews, David H.K. Shum, Sarah E. Harris, Gail Davies, Sanja Franic, Andrea Christoforou, Brendan Zietsch, Jodie Painter, *et al.* 2015. "Genetic basis of a cognitive complexity metric." *PloS One*, 10: e0123886.

Heckman, James J., John Eric Humphries, and Gregory Veramendi. 2018. "Returns to education: The causal effects of education on earnings, health, and smoking." *Journal of Political Economy*, 126: S197–S246.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor Economics*, 24: 411–482.

Hemani, Gibran, Jack Bowden, and George Davey Smith. 2018. "Evaluating the potential role of pleiotropy in Mendelian randomization studies." *Human Molecular Genetics*, 27: R195–R208.

Karlsson Linnér, Richard, Pietro Biroli, Edward Kong, S. Fleur W. Meddens, Robbee Wedow, Mark Alan Fontana, Maël Lebreton, Stephen P. Tino, Abdel Abdellaoui, Anke R. Hammerschlag, *et al.* 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: 245–257.

Kiernan, Kathleen E., and M. Carmen Huerta. 2008. "Economic deprivation, maternal depression, parenting and children's cognitive and emotional development in early childhood." *British Journal of Sociology*, 59: 783–806.

Koellinger, Philipp D., and Ronald de Vlaming. 2019. "Mendelian randomization: the challenge of unobserved environmental confounds." *International Journal of Epidemiology*, 48: 665–671.

Kong, Augustine, Gudmar Thorleifsson, Michael L. Frigge, Bjarni J. Vilhjalmsson, Alexander I. Young, Thorgeir E. Thorgeirsson, Stefania Benonisdottir, Asmundur Oddsson, Bjarni V. Halldorsson, Gisli Masson, *et al.* 2018. "The nature of nurture: Effects of parental genotypes." *Science*, 359: 424–428.

Lawlor, Deborah, Rebecca Richmond, Nicole Warrington, George McMahon, George Davey Smith, Jack Bowden, and David M. Evans. n.d.. "Using Mendelian randomization to determine causal effects of maternal pregnancy (intrauterine) exposures on offspring outcomes: Sources of bias and methods for assessing them." *Wellcome Open Research.*

Lee, James J, Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzian, Meghan Zacher, Tuan Anh Nguyen-Viet, Peter Bowers, Julia Sidorenko, Richard Karlsson Linnér, *et al.* 2018. "Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals." *Nature Genetics*, 50: 1112–1121.

Middeldorp, Christel M., Anke R. Hammerschlag, Klaasjan G. Ouwens, Maria M. Groen-Blokhuis, Beate St. Pourcain, Corina U. Greven, Irene Pappa, Carla MT Tiesler, Wei Ang, Ilja M. Nolte, *et al.* 2016. "A genome-wide association meta-analysis of attention-deficit/hyperactivity disorder symptoms in population-based pediatric cohorts." *Journal of the American Academy of Child and Adolescent Psychiatry*, 55: 896–905.

Nybom, Martin. 2017. "The distribution of lifetime earnings returns to college." *Journal of Labor Economics*, 35: 903–952.

O'hara, Michael W., and Jennifer E. McCabe. 2013. "Postpartum depression: Current status and future directions." *Annual Review of Clinical Psychology*, 9: 379–407.

Okbay, Aysu, Bart M.L. Baselmans, Jan-Emmanuel De Neve, Patrick Turley, Michel G. Nivard, Mark Alan Fontana, S. Fleur W. Meddens, Richard Karlsson Linnér, Cornelius A. Rietveld, Jaime Derringer, *et al.* 2016. "Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses." *Nature Genetics*, 48: 624–633.

Pappa, Irene, Beate St. Pourcain, Kelly Benke, Alana Cavadino, Christian Hakulinen, Michel G. Nivard, Ilja M. Nolte, Carla M.T. Tiesler, Marian J. Bakermans-Kranenburg, Gareth E. Davies, *et al.* 2016. "A genome-wide approach to children's aggressive behavior: The EAGLE consortium." *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 171: 562–572.

Perry, Cynthia D. 2008. "Does treating maternal depression improve child health management? The case of pediatric asthma." *Journal of Health Economics*, 27: 157–173.

Prince, Martin, Vikram Patel, Shekhar Saxena, Mario Maj, Joanna Maselko, Michael R. Phillips, and Atif Rahman. 2007. "No health without mental health." *Lancet*, 370: 859–877.

Rajagopal, Veera M., Andrea Ganna, Jonathan R.I. Coleman, Andrea Giuseppe Allegrini, Georgios Voloudakis, Jakob Grove, Thomas D. Als, Henriette Thisted Horsdal, Liselotte Vogdrup Petersen, Vivek Appadurai, *et al.* 2020. *Genome-Wide Association Study of School Grades Identifies a Genetic Overlap Between Language Ability, Psychopathology and Creativity.* BioRxiv, 075226.

Smith, George Davey, Debbie A. Lawlor, Roger Harbord, Nic Timpson, Ian Day, and Shah Ebrahim. 2007. "Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology." *PLoS Medicine*, 4: e352.

Stansfeld, Stephen, Clark Charlotte Bebbington Paul King Micheal Jenkins Rachel, and Stephen Hinchliffe. 2016. "Common mental disorders." In *Mental Health and Wellbeing in England: Adult Psychiatric Morbidity Survey 2014*. Eds. S. McManus, P. Bebbington, R. Jenkins and T. Brugha, Chapter 2, 37–68. NHS Digital.

Taylor, Amy E., Hannah J Jones, Hannah Sallis, Jack Euesden, Evie Stergiakouli, Neil M Davies, Stanley Zammit, Debbie A. Lawlor, Marcus R. Munafò, George Davey Smith, *et al.* 2018. "Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children." *International Journal of Epidemiology*, 47: 1207–1216.

Turley, Patrick, Raymond K Walters, Omeed Maghzian, Aysu Okbay, James J. Lee, Mark Alan Fontana, Tuan Anh Nguyen-Viet, Robbee Wedow, Meghan Zacher, Nicholas A. Furlotte, *et al.* 2018. "Multi-trait analysis of genome-wide association summary statistics using MTAG." *Nature Genetics*, 50: 229–237.

Van Kerm, Philippe. 2003. "Adaptive kernel density estimation." *Stata Journal*, 3: 148–156.

Van Kippersluis, Hans, and Cornelius A. Rietveld. 2018. "Pleiotropy-robust Mendelian randomization." *International Journal of Epidemiology*, 47: 1279–1288.

Von Hinke, Stephanie, George Davey Smith, Debbie A. Lawlor, Carol Propper, and Frank Windmeijer. 2016. "Genetic markers as instrumental variables." *Journal of Health Economics*, 45: 131–148.

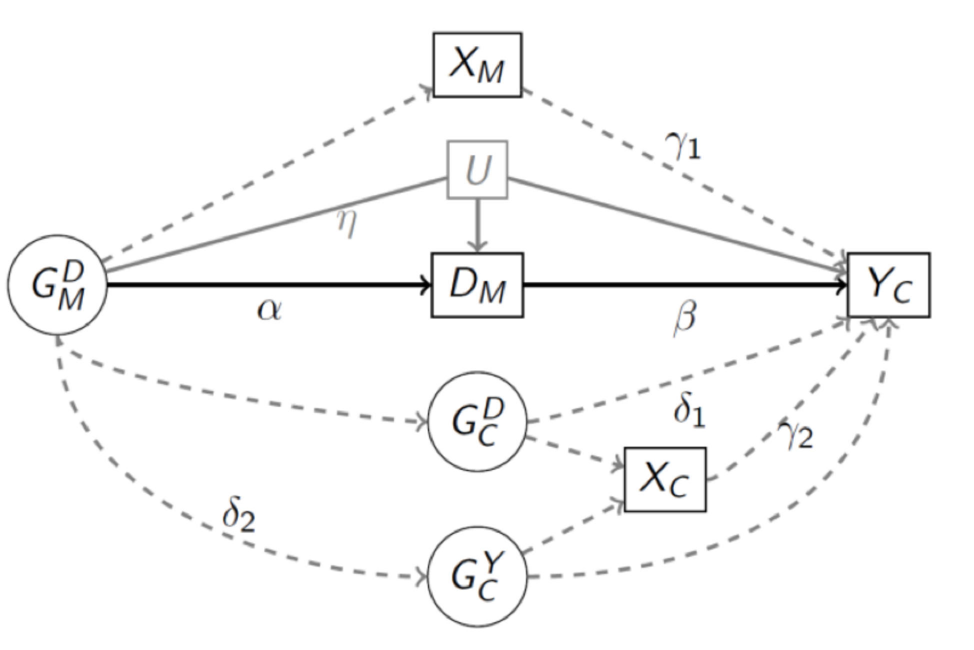
Von Hinke, Stephanie, Nigel Rice, Emma Tominey, ***et al.*** 2019. *Mental Health around Pregnancy and Child Development from Early Childhood to Adolescence.* IZA Discussion Paper, 12544.

Zhang, Ge, Jonas Bacelis, Candice Lengyel, Kari Teramo, Mikko Hallman, Øyvind Helgeland, Stefan Johansson, Ronny Myhre, Verena Sengpiel, Pål Rasmus Njølstad, ***et al.*** n.d.. "Assessing the causal relationship of maternal height on birth size and gestational age at birth: a mendelian randomization analysis." *PLoS Medicine*, 12: 1001865.

Zimmerman, Frederick J., and Wayne Katon. 2005. "Socioeconomic status, depression disparities, and financial strain: what lies behind the income-depression relationship?" *Health Economics*, 14: 1197–1215.

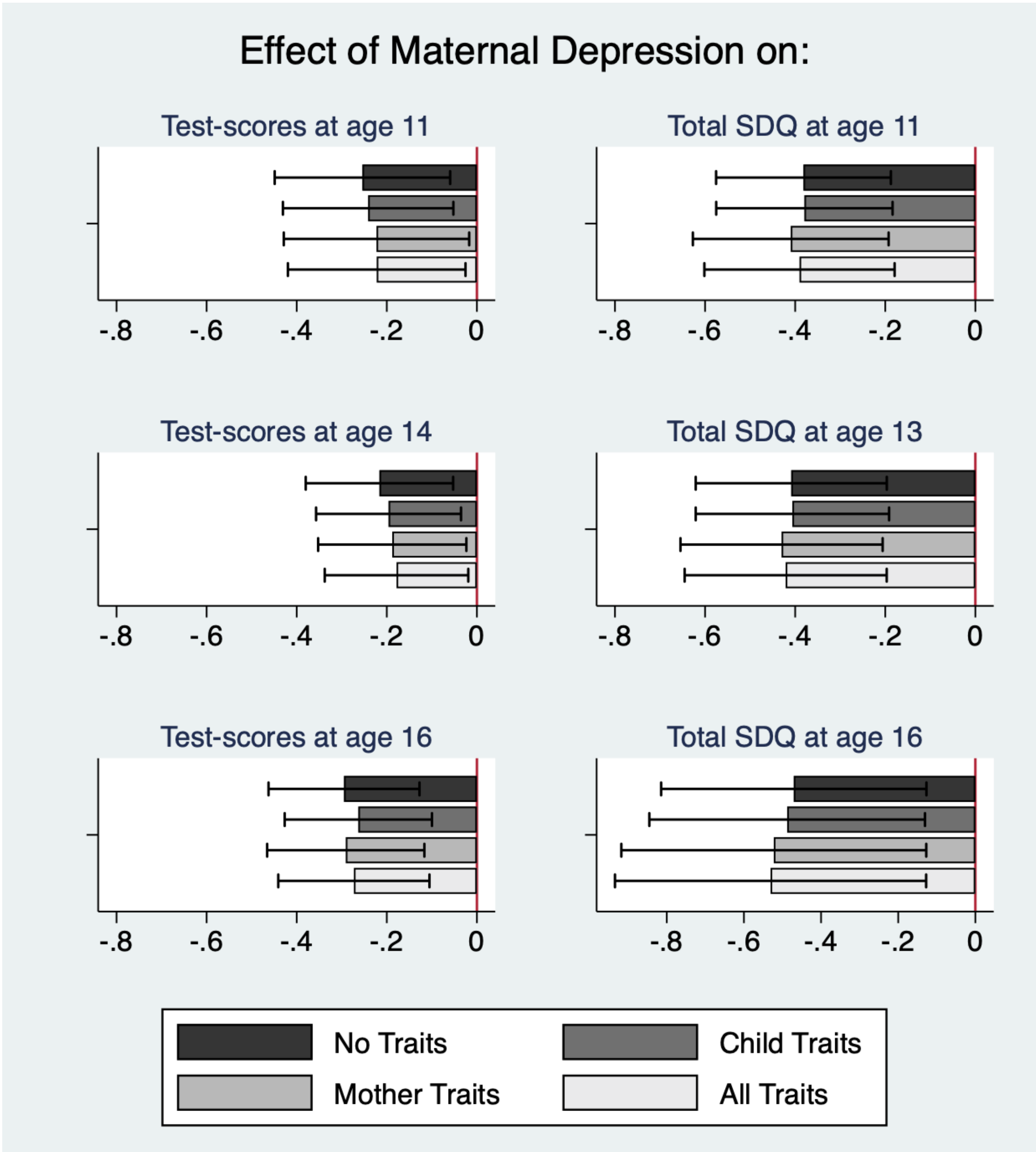
Figures and Tables

Figure 1.1: A Directed Acyclic Graph Illustrating the IV Setup and its Assumptions



Notes: The solid black lines depict the standard IV setup, where G_M^D is a (vector of) instrument(s) for a maternal trait D_M and Y_C is the child-level outcome of interest. U is a set of unobservable confounders of the trait-outcome association that should not be correlated with G_M^D (the independence assumption, i.e. $\eta = 0$). X_M is a set of maternal traits that are influenced by G_M^D through horizontal pleiotropy or other confounding pathways (e.g. genetic nurture) and have an impact on Y_C, thus violating the exclusion restriction. The identification issues in the bottom half of the figure reflect genetic inheritance (G_C^D and G_C^Y are, respectively, the child's genetic variants for traits D and Y). Lines with arrows at the end represent causal relationships; the line between G_M^D and U does not have an arrow and therefore reflects a correlational relationship.

Figure 1.2: Maternal Depression and Child Human Capital: 2SLS Results with Different Sets of Covariates



Notes: The horizontal lines in each bar show the 90% confidence intervals. All of the dependent variables are standardised. The child traits are the child's gender, birth year and birth-order dummies. The mother's traits are the child's number of siblings in the household, the age of the mother at birth of the cohort member, dummies for the mother having at least an A-level, working, having a partner, having a partner with at least an A-level, and having a working partner, and dummies for banded household income. All regressions using test-scores as the dependent variable include school and school-year fixed effects. The Cragg-Donald Wald F-statistics for weak identification, going from the "No traits" specification to the "All traits" specification, are the following: for KS2, 15.5, 16.0, 13.0, 14.0; for KS3, 20.9, 20.7, 18.8, 19.5; for KS4, 22.8, 22.2, 20.6, 21.2; for SDQ11, 23.8, 23.0, 20.6, 21.0; for SDQ13, 22.1, 21.4, 21.1, 20.8; for SDQ16, 9.7, 9.2, 8.3, 8.1.

Table 1.1: The Distribution of Maternal Depression

	Estimation Sample:					
	Test-scores			Total SDQ		
	Age 11 (1)	Age 14 (2)	Age 16 (3)	Age 11 (4)	Age 13 (5)	Age 16 (6)
Maternal Depression by Age 9:						
No episodes	47.1%	47.4%	47.0%	48.6%	50.3%	50.5%
1 episode	17.0%	16.6%	17.2%	16.7%	16.8%	17.1%
2 episodes	11.1%	11.1%	11.0%	10.9%	10.6%	11.1%
3 episodes	8.0%	7.5%	7.7%	7.2%	7.1%	6.7%
4 episodes	6.2%	6.4%	6.2%	6.0%	5.4%	6.1%
5 episodes	4.5%	4.6%	4.5%	4.4%	4.2%	3.7%
6 episodes	4.0%	3.9%	3.9%	3.9%	3.5%	2.5%
7 episodes	2.4%	2.6%	2.4%	2.4%	2.1%	2.3%
Observations	2601	2036	2828	2993	2585	2076

Note: These figures refer to the estimation samples used in the main analysis.

Table 1.2: Maternal Depression and Child Human Capital: OLS and 2SLS Results

	Test-scores					
	Age 11		Age 14		Age 16	
	OLS (1)	2SLS (2)	OLS (3)	2SLS (4)	OLS (5)	2SLS (6)
Maternal Depression	-0.018*	-0.222*	-0.031***	-0.178*	-0.016*	-0.273***
	(0.010)	(0.120)	(0.010)	(0.097)	(0.009)	(0.102)
First Stage:						
Mother's PGS		0.158***		0.205***		0.179***
		(0.042)		(0.046)		(0.039)
F-statistics		14.1		19.5		21.2
Observations	2601	2601	2036	2036	2828	2828
	Total SDQ					
	Age 11		Age 13		Age 16	
	OLS (1)	2SLS (2)	OLS (3)	2SLS (4)	OLS (5)	2SLS (6)
Maternal Depression	-0.100***	-0.390***	-0.087***	-0.421***	-0.083***	-0.531**
	(0.009)	(0.128)	(0.010)	(0.136)	(0.012)	(0.245)
First Stage:						
Mother's PGS		0.160***		0.167***		0.114***
		(0.035)		(0.037)		(0.040)
F-statistics		20.9		20.8		8.1
Observations	2993	2993	2585	2585	2076	2076

Notes: Standard errors appear in parentheses. All dependent variables are standardised. Maternal depression is the number of episodes of depression reported by the mother from the birth of the child to the child's ninth birthday. The controls are: the child's gender, birth year, birth order, number of siblings, the age of the mother at the birth of the child; dummies for the mother having at least an A-level, being employed, having a partner, having a partner with at least an A-level, having an employed partner, dummies for banded household income and ancestry-informative principal components. All regressions using test-scores as the dependent variable include school and school-year fixed effects. The reported F-statistics are those for the Cragg-Donald Wald weak-identification test. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

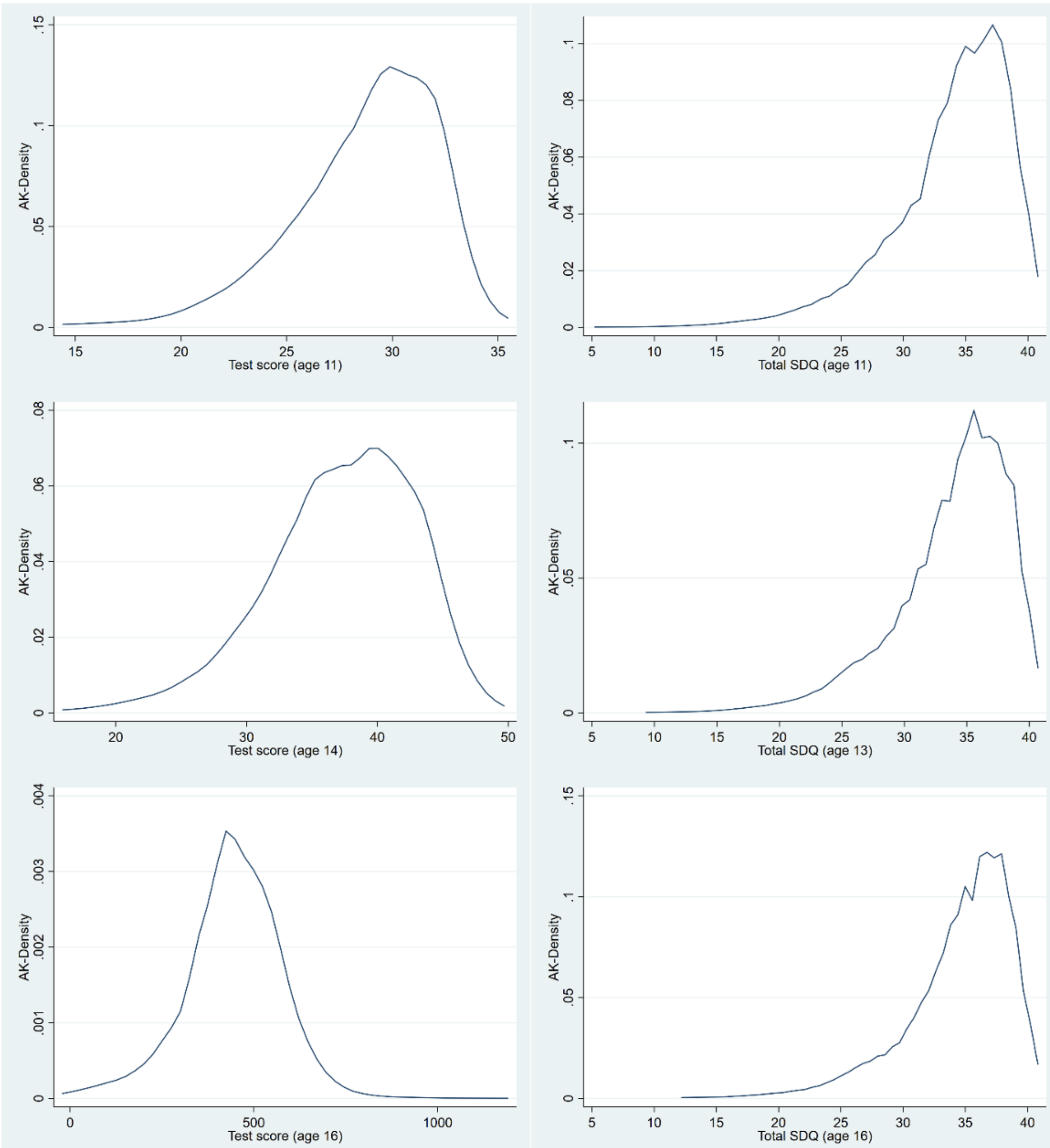
Table 1.3: Addressing Genetic Inheritance for Maternal Depression and Child Human Capital: 2SLS results

	Test-scores											
	Age 11				Age 14				Age 16			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Maternal Depression	-0.222*	-0.192	-0.238*	-0.225*	-0.178*	-0.164*	-0.182**	-0.177**	-0.273***	-0.255**	-0.294***	-0.288***
	(0.120)	(0.117)	(0.123)	(0.121)	(0.097)	(0.096)	(0.089)	(0.089)	(0.102)	(0.100)	(0.106)	(0.106)
Child PGS:												
Cognitive Skills		0.071***		0.072***		0.047**		0.047**		0.061***		0.062***
		(0.020)		(0.021)		(0.022)		(0.022)		(0.020)		(0.020)
Depression			0.005	0.012			0.002	0.006			0.009	0.014
			(0.022)	(0.021)			(0.021)	(0.021)			(0.021)	(0.021)
F-statistics	14.1	13.9	13.7	13.7	19.5	19.2	23.0	22.9	21.2	21.3	20.3	20.4
Observations	2601	2601	2601	2601	2036	2036	2036	2036	2828	2828	2828	2828
	Total SDQ											
	Age 11				Age 14				Age 16			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Maternal Depression	-0.390***	-0.391***	-0.295**	-0.296**	-0.421***	-0.421***	-0.384***	-0.383***	-0.531**	-0.529**	-0.478*	-0.482*
	(0.128)	(0.128)	(0.124)	(0.124)	(0.136)	(0.136)	(0.127)	(0.127)	(0.245)	(0.244)	(0.249)	(0.250)
Child PGS:												
Non-Cognitive Skills		0.044**		0.039*		0.058**		0.056**		0.077**		0.074**
		(0.022)		(0.021)		(0.025)		(0.024)		(0.031)		(0.031)
Depression			-0.036*	-0.036*			-0.015	-0.015			-0.014	-0.013
			(0.022)	(0.022)			(0.024)	(0.024)			(0.031)	(0.031)
F-statistics	20.9	20.9	19.5	19.4	20.8	20.8	22.4	22.4	8.1	8.1	7.1	7.0
Observations	22993	2993	2993	2993	2585	2585	2585	2585	2076	2076	2076	2076

Notes: Standard errors are in parentheses. All dependent variables are standardised. Maternal depression is the number of episodes of depression reported by the mother from the birth of the child to the child's ninth birthday. The controls are: the child's gender, birth year, birth order, number of siblings, the age of the mother at the birth of the child; dummies for the mother having at least an A-level, being employed, having a partner, having a partner with at least an A-level, having an employed partner, dummies for banded household income and ancestry-informative principal components. All regressions using test-scores as the dependent variable include school and school-year fixed effects. The reported F-statistics are those for the Cragg-Donald Wald weak-identification test. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

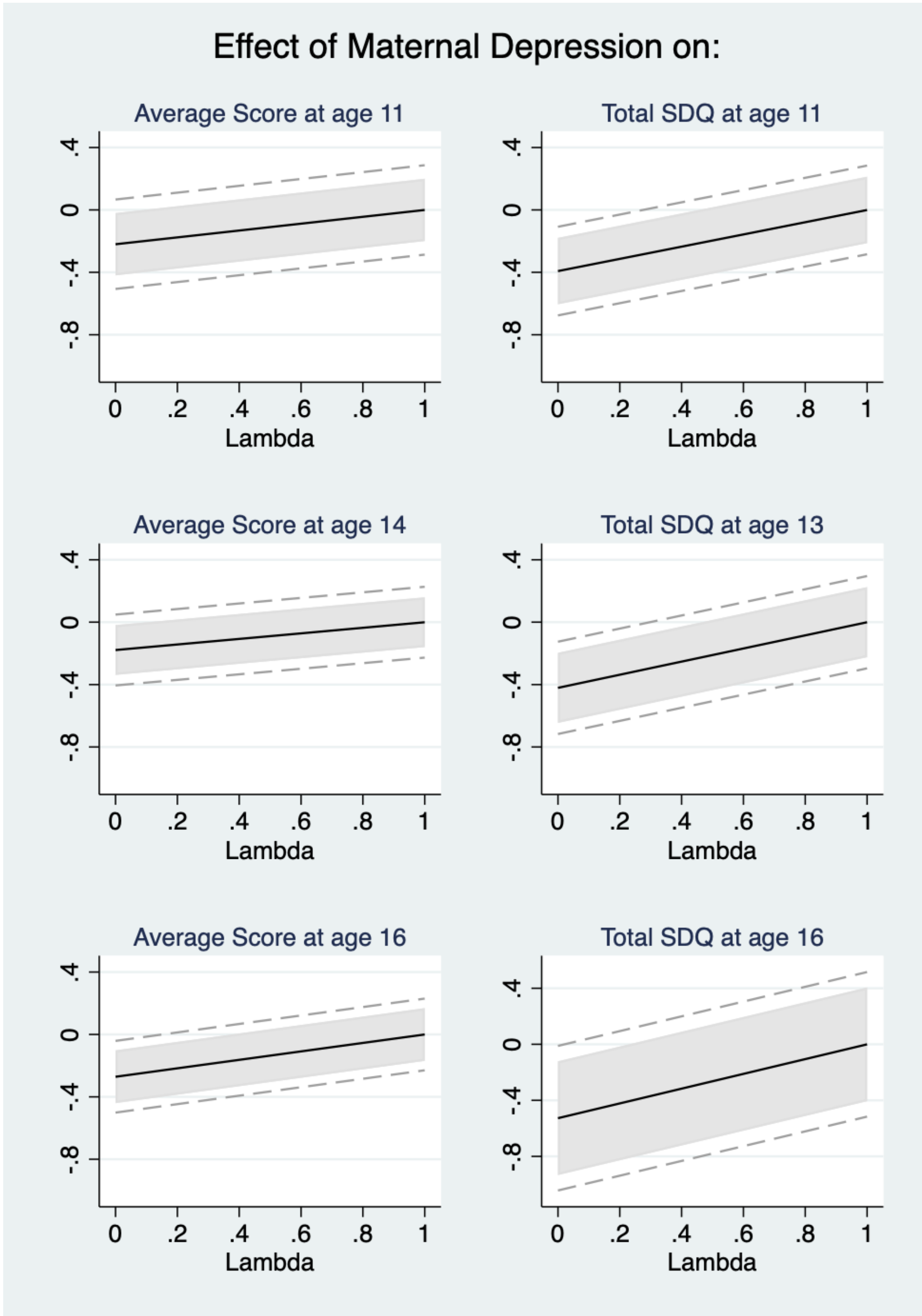
Appendix 1.A: Additional Figures and Tables

Figure 1.A1: The Distribution of Test-scores and Total SDQ



Notes: Each figure refers to one of our estimation samples. The test-scores at age 11, 14 and 16 respectively refer to the Key-Stage 2 average score, Key-Stage 3 average score and GCSE total score. The densities are plotted using an adaptive-kernel (see Van Kerm, 2003, for the technical details).

Figure 1.A2: Plausible Exogeneity and Pleiotropy-robust MR



Notes: Lambda represents the share of the reduced-form effect of the instrument on the outcome that is independent of maternal depression. The black line is the 2SLS point estimate of maternal depression for different values of lambda; the grey solid area represents 90% confidence intervals using the Nybom (2017) approach, while the grey dashed lines are the 90% confidence intervals following Van Kippersluis and Rietveld (2018).

Table 1.A1: The Strengths and Difficulties Questionnaire (SDQ)

Please think about this child's behaviour over the last 6 months if you can:

This child:	NOT TRUE	SOMEWHAT TRUE	CERTAINLY TRUE
Emotional health:			
Often complains of headaches, stomachaches or sickness	0	1	2
Has many worries, often seems worried	0	1	2
Is often unhappy, down-hearted or tearful	0	1	2
Is nervous or clingy in new situations, easily loses confidence	0	1	2
Has many fears, is easily scared	0	1	2
Total emotional health score: 0-10			
Conduct problems:			
Has temper tantrums or hot tempers	0	1	2
Is generally obedient, usually does what adults request	2	1	0
Often fights with other children or bullies them	0	1	2
[a]Often lies or cheats	0	1	2
[b]Steals from home/school/elsewhere	0	1	2
Total conduct problems score: 0-10			
Hyperactity/Inattention:			
Is restless, overactive, cannot stay still for long	0	1	2
Constantly fidgets or squirms	0	1	2
Is easily distracted, concentration wandered	0	1	2
[c]Thinks things out before acting	2	1	0
Sees tasks through to the end, good attention span	2	1	0
Total hyperactivity score: 0-10			
Peer relationship problems:			
Is rather solitary, tends to play alone	0	1	2
Has at least one good friend	2	1	0
Is generally liked by other children	2	1	0
Is picked on or bullied by other children	0	1	2
Gets on better with adults than with other children	0	1	2
Total peer relationship problems score: 0-10			
Total internalising behaviour = emotional + peer relationship (0-20)			
Total externalising behaviour = behaviour + hyperactivity (0-20)			

[a] Changed to "Often argumentative with adults" in the questionnaire for 3-4 years old.

[b] Changed to "Can be spiteful to others" in the questionnaire for 3-4 years old.

[c] Changed to "Can stop and think things out before acting" in the questionnaire for 3-4 years old.

Table 1.A2: Descriptive Statistics: Cognitive Skills Estimation Samples

	Estimation samples:											
	Test-Scores (age 11) 2601 observations				Test-Scores (age 14) 2036 observations				Test-Scores (age 16) 2828 observations			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Cohort member characteristics:												
KS2 average score (fine-grading)	28.76	3.48	15	35								
KS3 average score (fine-grading)					37.33	5.65	17	49				
GCSE total score									445.48	132.63	0	1171
Female	0.51		0	1	0.52		0	1	0.51		0	1
Birth year:												
1991	0.36		0	1	0.43		0	1	0.36		0	1
1992	0.63		0	1	0.57		0	1	0.64		0	1
1993	0.01		0	1	0.00		0	0	0.01		0	1
Birth order:												
1st-born	0.46		0	1	0.45		0	1	0.46		0	1
2nd-born	0.37		0	1	0.38		0	1	0.37		0	1
3rd-born	0.14		0	1	0.13		0	1	0.14		0	1
4th-born or higher	0.03		0	1	0.03		0	1	0.03		0	1
Mother and family characteristics:												
No. of episodes of maternal depression	1.50	1.94	0	7	1.51	1.96	0	7	1.50	1.94	0	7
Employed mother	0.74		0	1	0.74		0	1	0.73		0	1
Mother has an A-level or above	0.33		0	1	0.33		0	1	0.35		0	1
Age of the mother at child's birth	29.29	4.23	18	44	29.29	4.23	18	44	29.43	4.28	18	44
Presence of partner	0.94		0	1	0.94		0	1	0.94		0	1
Employed partner	0.89		0	1	0.89		0	1	0.88		0	1
Partner has an A-level or above	0.47		0	1	0.47		0	1	0.49		0	1
Number of siblings	2.29	0.77	0	6	2.29	0.77	0	6	2.28	0.76	0	6
Average family income per week:												
<£100	0.01		0	1	0.01		0	1	0.01		0	1
£100-£199	0.07		0	1	0.07		0	1	0.07		0	1
£200-£299	0.16		0	1	0.16		0	1	0.15		0	1
£300-£399	0.20		0	1	0.21		0	1	0.20		0	1
£400+	0.47		0	1	0.46		0	1	0.48		0	1
Do not know	0.05		0	1	0.05		0	1	0.05		0	1
Missing	0.04		0	1	0.04		0	1	0.04		0	1

Note: These figures refer to our estimation samples.

Table 1.A3: Descriptive Statistics: Non-Cognitive Skills Estimation Samples

	Estimation samples:											
	Total SDQ (age 11) 2601 observations				Total SDQ (age 13) 2036 observations				Total SDQ (age 16) 2828 observations			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Cohort member characteristics:												
Total SDQ	33.79	4.85	6	40	33.86	4.55	10	40	34.55	4.26	13	40
Female	0.50		0	1	0.50		0	1	0.53		0	1
Birth year:												
1991	0.36		0	1	0.36		0	1	0.35		0	1
1992	0.63		0	1	0.63		0	1	0.64		0	1
1993	0.01		0	1	0.01		0	1	0.01		0	1
Birth order:												
1st-born	0.47		0	1	0.47		0	1	0.48		0	1
2nd-born	0.37		0	1	0.37		0	1	0.37		0	1
3rd-born	0.13		0	1	0.13		0	1	0.12		0	1
4th-born or more	0.03		0	1	0.03		0	1	0.02		0	1
Mother and family characteristics:												
No. of episode of maternal depression	1.46	1.94	0	7	1.37	1.88	0	7	1.33	1.84	0	7
Employed mother	0.73		0	1	0.74		0	1	0.74		0	1
Mother has an A-level or more	0.39		0	1	0.41		0	1	0.44		0	1
Age of the mother at child's birth	29.59	4.27	18	44	29.63	4.24	18	43	29.77	4.26	18	43
Presence of partner	0.94		0	1	0.94		0	1	0.95		0	1
Employed partner	0.89		0	1	0.89		0	1	0.90		0	1
Partner has an A-level or above	0.53		0	1	0.54		0	1	0.56		0	1
Number of siblings	2.27	0.76	0	6	2.26	0.75	0	6	2.25	0.74	0	6
Average family income per week:												
<£100	0.01		0	1	0.01		0	1	0.01		0	1
£100-£199	0.06		0	1	0.06		0	1	0.05		0	1
£200-£299	0.14		0	1	0.13		0	1	0.13		0	1
£300-£399	0.19		0	1	0.19		0	1	0.17		0	1
£400+	0.51		0	1	0.53		0	1	0.55		0	1
Do not know	0.05		0	1	0.04		0	1	0.04		0	1
Missing	0.04		0	1	0.04		0	1	0.04		0	1

Note: These figures refer to our estimation samples.

Table 1.A4: The PGS for Maternal Depression and Maternal Traits: Bivariate Associations

	PGS for Maternal Depression
Mother's Traits:	
Maternal Depression [0-7]	0.177***
	(0.035)
Mother is Employed	-0.007
	(0.008)
Mother has at least an A-level	0.003
	(0.009)
Age at Birth of CM	0.066
	(0.076)
Mother has a Partner	-0.005
	(0.004)
Working Partner†	0.002
	(0.004)
Partner has at least an A-level	0.012
	(0.009)
Number of Siblings	0.013
	(0.014)
Family Income above the Median†	-0.013
	(0.009)

Notes: Each cell reports the estimate and standard errors (in parentheses) of separate bivariate associations between the PGS for maternal depression and various maternal traits. The bivariate regressions are based on the observations coming from our largest estimation sample (the one for which the SDQ at age 11 is not missing: 2993 observations). Using our five other estimation samples produces similar estimates.
† indicates that we excluded the missing values of the maternal traits from the estimation sample (including the missing values and introducing a missing-indicator flag produces similar estimates). Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Table 1.A5: Maternal Depression and Child Human Capital: 2SLS Results using the Mother's PGS for Depression excluding Genetic Variants Linked to Known Biological Pathways

	Test-scores					
	Age 11		Age 14		Age 16	
	(1)	(2)	(3)	(4)	(5)	(6)
Maternal Depression	-0.222*	-0.214*	-0.178*	-0.142	-0.273***	-0.290***
	(0.120)	(0.124)	(0.097)	(0.103)	(0.102)	(0.110)
First Stage:						
Mother's PGS	0.158***	0.152***	0.205***	0.189***	0.179***	0.167***
	(0.042)	(0.042)	(0.046)	(0.047)	(0.039)	(0.039)
F-statistics	14.1	13.0	19.5	16.5	21.2	18.7
Observations	2601	2601	2036	2036	2828	2828
	Total SDQ					
	Age 11		Age 13		Age 16	
	(1)	(2)	(3)	(4)	(5)	(6)
Maternal Depression	-0.390***	-0.396***	-0.421***	-0.418***	-0.531**	-0.534*
	(0.128)	(0.145)	(0.136)	(0.157)	(0.245)	(0.314)
First Stage:						
Mother's PGS	0.160***	0.142***	0.167***	0.145***	0.114***	0.090***
	(0.035)	(0.035)	(0.036)	(0.037)	(0.040)	(0.040)
F-statistics	20.9	16.5	20.8	15.5	8.1	5.0
Observations	2993	2993	2585	2585	2076	2076

Notes: Standard errors appear in parentheses. All dependent variables are standardised. Maternal depression is the number of episodes of depression reported by the mother from the birth of the child to the child's ninth birthday. The controls are: the child's gender, birth year, birth order, number of siblings, the age of the mother at the birth of the child; dummies for the mother having at least an A-level, being employed, having a partner, having a partner with at least an A-level, having an employed partner, dummies for banded household income and ancestry-informative principal components. All regressions using test-scores as the dependent variable include school and school-year fixed effects. The reported F-statistics are those for the Cragg-Donald Wald weak-identification test. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Table 1.A6: Maternal Depression and Child Human Capital: 2SLS Results Controlling for Mother's PGS for Cognitive and Non-Cognitive Skills

	Test-scores					
	Age 11		Age 14		Age 16	
	(1)	(2)	(3)	(4)	(5)	(6)
Maternal Depression	-0.222*	-0.214*	-0.178*	-0.177*	-0.273***	-0.270***
	(0.120)	(0.121)	(0.097)	(0.098)	(0.102)	(0.103)
First Stage:						
Mother's PGS for Depression	0.158***	0.156***	0.205***	0.202***	0.179***	0.177***
	(0.042)	(0.042)	(0.046)	(0.046)	(0.039)	(0.039)
Mother's PGS for Cognitive Skills		-0.049		-0.093*		-0.030
		(0.046)		(0.050)		(0.042)
F-statistics	14.1	13.5	19.5	18.9	21.2	20.8
Observations	2601	2601	2036	2036	2828	2828
	Total SDQ					
	Age 11		Age 13		Age 16	
	(1)	(2)	(3)	(4)	(5)	(6)
Maternal Depression	-0.390***	-0.392***	-0.421***	-0.422***	-0.531**	-0.533**
	(0.128)	(0.128)	(0.136)	(0.136)	(0.245)	(0.244)
First Stage:						
Mother's PGS for Depression	0.163***	0.164***	0.171***	0.172***	0.174***	0.129***
	(0.035)	(0.035)	(0.036)	(0.036)	(0.055)	(0.043)
Mother's PGS for Non-Cognitive Skills		-0.025		-0.044		-0.027
		(0.035)		(0.037)		(0.040)
F-statistics	20.9	21.0	20.8	20.9	8.1	8.2
Observations	2993	2993	2585	2585	2076	2076

Notes: Notes: Standard errors appear in parentheses. All dependent variables are standardised. Maternal depression is the number of episodes of depression reported by the mother from the birth of the child to the child's ninth birthday. The controls are: the child's gender, birth year, birth order, number of siblings, the age of the mother at the birth of the child; dummies for the mother having at least an A-level, being employed, having a partner, having a partner with at least an A-level, having an employed partner, dummies for banded household income and ancestry-informative principal components. All regressions using test-scores as the dependent variable include school and school-year fixed effects. The reported F-statistics are those for the Cragg-Donald Wald weak-identification test. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Table 1.A7: Maternal Depression and Child Human Capital: 2SLS Results Controlling for Grandparental Characteristics

	Test-scores			Total SDQ		
	Age 11 (1)	Age 14 (2)	Age 16 (3)	Age 11 (4)	Age 13 (5)	Age 16 (6)
Maternal Depression	-0.255*	-0.186*	-0.296***	-0.385***	-0.437***	-0.525*
	(0.132)	(0.100)	(0.112)	(0.138)	(0.147)	(0.273)
First Stage:						
Mother's PGS	0.146***	0.196***	0.164***	0.147***	0.158***	0.102**
	(0.042)	(0.047)	(0.039)	(0.035)	(0.037)	(0.040)
F-statistics	11.8	17.5	17.7	17.6	18.2	6.4
Observations	2601	2036	2828	2993	2585	2076

Notes: Standard errors are in parentheses. All dependent variables are standardised. Maternal depression is the number of episodes of depression reported by the mother from the birth of the child to the child's ninth birthday. The controls are: the child's gender, birth year, birth order, number of siblings, the age of the mother at the birth of the child; dummies for the mother having at least an A-level, being employed, having a partner, having a partner with at least an A-level, having an employed partner, dummies for banded household income and ancestry-informative principal components. The regressions also include the following grandparental characteristics: dummies for the mother having at least one parent who had a mental illness before the birth of the child and the highest diploma obtained by the maternal grandmother and grandfather, and a social-status index for the maternal grandmother and grandfather. All of the regressions using test-scores as the dependent variable include school and school-year fixed effects. The reported F-statistics are those for the Cragg-Donald Wald weak-identification test. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Table 1.A8: Maternal Depression and Child Human Capital: 2SLS Results Controlling for Partner's Depression

	Test-scores			Total SDQ		
	Age 11	Age 14	Age 16	Age 11	Age 13	Age 16
	(1)	(2)	(3)	(4)	(5)	(6)
Maternal Depression	-0.240*	-0.187*	-0.297***	-0.406***	-0.436***	-0.566**
	(0.129)	(0.102)	(0.113)	(0.136)	(0.143)	(0.270)
First Stage:						
Mother's PGS	0.148***	0.195***	0.163***	0.151***	0.160***	0.105***
	(0.040)	(0.044)	(0.037)	(0.033)	(0.035)	(0.039)
Partner's Depression	0.965***	0.963***	0.942***	0.912***	0.859***	0.873***
	(0.068)	(0.073)	(0.062)	(0.056)	(0.060)	(0.066)
F-statistics	13.4	19.3	19.3	20.3	20.4	7.4
Observations	2601	2036	2828	2993	2585	2076

Notes: Standard errors appear in parentheses. All dependent variables are standardised. Maternal depression is the number of episodes of depression reported by the mother from the birth of the child to the child's ninth birthday. The controls are: the child's gender, birth year, birth order, number of siblings, the age of the mother at the birth of the child; dummies for the mother having at least an A-level, being employed, having a partner, having a partner with at least an A-level, having an employed partner, dummies for banded household income and ancestry-informative principal components. All regressions using test-scores as the dependent variable include school and school-year fixed effects. The reported F-statistics are those for the Cragg-Donald Wald weak-identification test. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Table 1.A9: Descriptive Statistics: Alternative Measures of Maternal Depression

	Mean	SD	Min	Max
Estimation Sample: Test-scores (age 11)				
No. of depressive episodes between child age 0 and 5	1.06	1.44	0	5
No. of depressive episodes between child age 5 and 9	0.44	0.70	0	2
No. of depressive episodes between child age 2 and 9	1.25	1.70	0	6
Had at least one episode of depression	0.48		0	1
Recent depression	0.24		0	1
Mean EPDS	5.60	3.80	0	24.33
Estimation Sample: Test-scores (age 14)				
No. of depressive episodes between child age 0 and 5	1.08	1.46	0	5
No. of depressive episodes between child age 5 and 9	0.43	0.70	0	2
No. of depressive episodes between child age 2 and 9	1.25	1.71	0	6
Had at least one episode of depression	0.48		0	1
Recent depression	0.23		0	1
Mean EPDS	5.62	3.84	0	24
Estimation Sample: Test-scores (age 16)				
No. of depressive episodes between child age 0 and 5	1.06	1.44	0	5
No. of depressive episodes between child age 5 and 9	0.44	0.70	0	2
No. of depressive episodes between child age 2 and 9	1.25	1.70	0	6
Had at least one episode of depression	0.48		0	1
Recent depression	0.24		0	1
Mean EPDS	5.56	3.76	0	24
Estimation Sample: Total SDQ (age 11)				
No. of depressive episodes between child age 0 and 5	1.03	1.44	0	5
No. of depressive episodes between child age 5 and 9	0.43	0.70	0	2
No. of depressive episodes between child age 2 and 9	1.21	1.70	0	6
Had at least one episode of depression	0.47		0	1
Recent depression	0.23		0	1
Mean EPDS	5.47	3.72	0	24
Estimation Sample: Total SDQ (age 13)				
No. of depressive episodes between child age 0 and 5	0.97	1.40	0	5
No. of depressive episodes between child age 5 and 9	0.40	0.68	0	2
No. of depressive episodes between child age 2 and 9	1.14	1.66	0	6
Had at least one episode of depression	0.45		0	1
Recent depression	0.22		0	1
Mean EPDS	5.30	3.60	0	24
Estimation Sample: Total SDQ (age 16)				
No. of depressive episodes between child age 0 and 5	0.95	1.38	0	5
No. of depressive episodes between child age 5 and 9	0.39	0.66	0	2
No. of depressive episodes between child age 2 and 9	1.10	1.62	0	6
Had at least one episode of depression	0.21		0	1
Recent depression	0.44		0	1
Mean EPDS	5.27	3.54	0	24

Note: These figures refer to the estimation samples used in Table 1.A10.

Table 1.A10: Alternative Measures of Maternal Depression and Child Human Capital: 2SLS results

	Test-scores			Total SDQ		
	Age 11 (1)	Age 14 (2)	Age 16 (3)	Age 11 (4)	Age 13 (5)	Age 16 (6)
Panel A (Baseline from Table 1.2)						
No. of depressive episodes	-0.227*	-0.179*	-0.284***	-0.394***	-0.416***	-0.498**
between child ages 0 and 9	(0.118)	(0.093)	(0.102)	(0.125)	(0.132)	(0.211)
F-statistics	14.8	21.3	21.8	22.1	22.2	10.3
Observations	2601	2036	2828	2993	2585	2076
Panel B						
No. of depressive episodes	-0.273**	-0.223*	-0.364***	-0.512***	-0.579***	-0.597**
up to child's 5th birthday	(0.139)	(0.115)	(0.130)	(0.162)	(0.187)	(0.244)
F-statistics	18.2	24.7	23.7	23.5	20.5	12.5
Observations	2601	2036	2828	2993	2585	2076
Panel C						
No. of depressive episodes	-0.681*	-0.545*	-0.777**	-1.019***	-0.969***	-1.750
between child ages 5 and 9	(0.409)	(0.320)	(0.328)	(0.375)	(0.335)	-1.112
F-statistics	6.0	8.4	10.4	12.4	15.6	3.0
Observations	2601	2036	2828	2993	2585	2076
Panel D						
No. of depressive episodes	-0.282*	-0.218*	-0.345***	-0.468***	-0.475***	-0.572**
between child ages 2 and 9	(0.148)	(0.114)	(0.126)	(0.151)	(0.152)	(0.245)
F-statistics	12.6	18.8	19.1	20.4	21.9	10.0
Observations	2601	2036	2828	2993	2585	2076
Panel E						
Had at least one episode	-0.948*	-0.843*	-1.234***	-1.555***	-1.631***	-1.520**
of depression (dummy)	(0.502)	(0.457)	(0.463)	(0.502)	(0.532)	(0.610)
F-statistics	12.6	14.6	17.0	20.9	20.2	14.8
Observations	2601	2036	2828	2993	2585	2076
Panel F						
Recent depression (dummy)	-3.218	-1.972	-2.565**	-3.162**	-2.826**	-7.374
	-2.759	-1.339	-1.283	-1.358	-1.107	-7.789
F-statistics	1.6	3.8	5.5	7.1	9.9	0.9
Observations	2601	2036	2828	2993	2585	2076
Panel G						
Average EPDS	-0.889	-0.583*	-1.038**	-1.220***	-0.961***	-2.048
	(0.568)	(0.336)	(0.485)	(0.443)	(0.315)	-1.394
F-statistics	3.7	7.1	6.4	9.1	15.2	2.2
Observations	2559	1998	2783	2947	2548	2053

Notes: Standard errors appear in parentheses. All dependent variables are standardised. All non-binary measures of maternal depression are standardised. The controls are: the child's gender, birth year, birth order, number of siblings, the age of the mother at the birth of the child; dummies for the mother having at least an A-level, being employed, having a partner, having a partner with at least an A-level, having an employed partner, dummies for banded household income and ancestry-informative principal components. All of the regressions with test-scores as the dependent variables include school and school-year fixed effects. The difference in observations in the sixth panel is due to missing values for the EPDS (Edinburgh Postnatal Depression Score) when the cohort member was eight years old. The reported F-statistics are those for the Cragg-Donald Wald weak-identification test. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Table 1.A11: Maternal Depression and Children's Non-Cognitive Skills: Validity Tests - 2SLS Results

	Total SDQ (Teacher reported)	SMFQ (Main-carer reported)		
	Age 11 (1)	Age 11 (2)	Age 13 (3)	Age 16 (4)
Maternal Depression	-0.332**	-0.399***	-0.362***	-0.130
	(0.165)	(0.128)	(0.133)	(0.188)
First Stage:				
Mother's PGS	0.172***	0.160***	0.163***	0.116***
	(0.048)	(0.035)	(0.037)	(0.041)
F-statistics	12.6	20.9	19.4	8.1
Observations	1559	2993	2558	2015

Notes: Standard errors are in parentheses. All dependent variables are standardised. SMFQ takes on integer values from 0 to 26. The mean values (standard deviations) at ages 11, 13 and 16 are 23.8 (3.1), 23.9 (3.1) and 24.2 (3.0). Teacher-reported total SDQ at age 11 is on the same scale as carer-reported total SDQ, with mean (standard deviation) of 35.5 (5.1). Maternal depression is the number of episodes of depression reported by the mother from the birth of the child to the child's ninth birthday. The controls are: the child's gender, birth year, birth order, number of siblings, the age of the mother at the birth of the child; dummies for the mother having at least an A-level, being employed, having a partner, having a partner with at least an A-level, having an employed partner, dummies for banded household income and ancestry-informative principal components. The reported F-statistics are those for the Cragg-Donald Wald weak-identification test. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Table 1.A12: The Short Moods and Feelings Questionnaire (SMFQ)

These questions are about how your teenager may have been feeling or acting recently. For each question, please say how much you think he/she has felt or acted this way in the past two weeks.

Your teenager :	NOT TRUE	SOMEWHAT TRUE	TRUE
1. Felt miserable or unhappy	2	1	0
2. Didn't enjoy anything at all	2	1	0
3. Felt so tired that he/she just sat around and did nothing	2	1	0
4. Was very restless	2	1	0
5. Felt he/she was no good anymore	2	1	0
6. Cried a lot	2	1	0
7. Found it hard to think properly or concentrate	2	1	0
8. Hated himself/herself	2	1	0
9. Was a bad person	2	1	0
10. Felt lonely	2	1	0
11. Thought nobody really loved him/her	2	1	0
12. Thought he/she could never be as good as other kids	2	1	0
13. Felt he/she did everything wrong	2	1	0

Appendix 1.B: Polygenic Scores

Genetic variants, such as single-nucleotide polymorphisms (SNPs), are locations in the human DNA at which a certain degree of variation is observed across individuals in a population. There are only two possible nucleotide variations for each SNP, which are called alleles. Some specific alleles, the ‘effect alleles’, are associated with particular diseases or traits (due to evolutionary mechanisms of natural selection, these are typically the alleles appearing less frequently in the population, i.e. the ‘minor alleles’). The number of effect alleles that an individual possesses for a given SNP (the so-called ‘allelic dosage’) can either be 0 (no effect allele), 1 (only one effect allele), or 2 (both alleles are the effect allele). Polygenic scores (PGS) are weighted sums or averages of individual allelic dosages for a given set of SNPs. Both the weights and the set of relevant SNPs are obtained from the publicly-available summary statistics of an existing GWAS. These are typically tables or text files providing a list of SNPs that are robustly associated with a trait, accompanied by a range of characteristics (e.g. the effect size of the SNP-trait association, the p-value of such association, and the effect allele). For a given individual j in a prediction sample (independent of the training sample used in the GWAS), the default formula for the calculation of her PGS in the command line program PLINK 1.9 (www.cog-genomics.org/plink/) is:

$$PGS_j = \frac{\sum_{i=1}^{N} \hat{\beta}_i X_{ij}}{PM_j}$$

where $\hat{\beta}_i$ is the estimated effect size of SNP i on the trait of interest (obtained from a GWAS), X_{ij} the number of effect alleles observed in individual j for SNP i, P the ploidy of the sample (i.e. the number of sets of chromosomes in a cell, which is generally two for humans), N the total number of SNPs included in the PGS, and M_j the number of non-missing SNPs observed in individual j. If individual j has a missing genotype for SNP i, then the population minor-allele frequency multiplied by the ploidy ($MAF_i \times P$) is used instead of X_{ij}.

The allelic dosages of SNPs that are close to each other on a DNA strand tend to be correlated due to linkage disequilibrium (LD), i.e. $Cov(X_{ij}, X_{sj}) \neq 0$ for i and s that are close enough. As each $\hat{\beta}_i$ coming from the GWAS results is separately estimated from a linear regression of the trait of interest on SNP i in the training sample, some of the $\hat{\beta}_i$'s will be biased. Due to the overweighting of SNPs in long LD blocks, the resulting PGS will also be biased, and will have worse predictive accuracy. There are several ways to account for LD when interpreting the results of a GWAS: some, such as pruning or clumping,[20] are based on recursive algorithms where only approximately-independent SNPs are retained for PGS construction; others, such as LDpred (Vilhjálmsson *et al.*, 2015), use more complex machine-learning algorithms and Bayesian inference to obtain corrected effect-size estimates $\hat{\beta}_i$ that take LD into account. It is increasingly common for GWAS authors to identify a 'clean' set of approximately-independent SNPs (either via pruning or clumping). This is the case for the depression meta-analysis in Turley *et al.* (2018) that we use to calculate our instrumental variable. The single-trait meta-analysis expands on the SNPs for depressive symptoms already identified in Okbay et al. (2016), using a larger sample of 465,337 individuals from UK Biobank, 23andMe, and the Resource for Genetic Epidemiology Research on Adult Health and Aging (GERA). In particular, using clumping (for more details see the Online Methods of Turley *et al.*, 2018), the single-trait depression GWAS of Turley and co-authors identified 88 SNPs for depression significant at least at a 10^{-6} p-value threshold, of which 68 were available in our ALSPAC genotyped data and used for the construction of the PGS for maternal depression.[21] Out of those 88 SNPs, 32 have a p-value lower than the genome-wide significant threshold 5×10^{-8} and were used in robustness checks to test the sensitivity of our PGS to the number of SNPs included in its computation (available upon request).

Other PGSs used for the production of Table 3 were derived from the GWASs of Demange *et al.* (2021) and Middeldorp *et al.* (2016). The former, which captures the cognitive aspects of educational attainment, was used to calculate the PGS for cognitive skills. We clumped the GWAS summary statistics using p-value thresholds of, respectively, 5×10^{-8} for the lead SNPs and 10^{-6} for the SNPs in the clumps. Clumps were defined based on windows of 1000 kb from the lead SNPs and squared pairwise correlations of at least 0.1 (LD patterns were inferred from the sequenced genotypes of 379 individuals of European descent from Phase 1 of the 1000 Genomes Project). Only 179 out of the

[20]Pruning takes the available SNPs in the prediction sample as the starting point. For each SNP in a defined window, a pruning algorithm generally calculates the Variance Inflation Factor (VIF) or the squared pairwise correlation between each pair of SNPs, and removes the pair if the LD is greater than a certain threshold (e.g. 0.5). The procedure is then repeated shifting the window a certain number of SNPs forward. Clumping, on the other hand, uses the GWAS summary results as the starting point. This procedure starts by taking the SNP whose association with the trait of interest has the smallest p-value (the 'lead' SNP) and constructing a symmetric window around it; SNPs in the window that have a squared pairwise correlation with the lead SNP above a certain cut-off are assigned to the lead SNP's clump. The algorithm continues by taking the next-most significant SNP that is not yet assigned to a clump and repeating the above procedure until there are no more significant SNPs (based on user-defined significance thresholds; for large GWASs, the genome-wide significance p-value threshold of 5×10^{-8} is often used for lead SNPs). The clumped set of SNPs is the list of all lead SNPs.

[21]The full list of SNPs is available in the supplementary material of Turley *et al.* (2018). Those we could not use, as they were not available in the genotyped data of ALSPAC participants, are the following: rs1806153, rs3806843, rs4799936, rs9291059, rs9813064, rs10172121, rs10965565, rs113092725, rs11643097, rs11663393, rs12501627, rs12515229, rs1520081, rs189383553, rs192796028, rs28383313, rs28567442, rs413130, rs7126679, rs9663959.

225 lead SNPs were found in ALSPAC children and used for the calculation of the child's PGS for cognitive skills.

For non-cognitive skills, we first applied the same clumping procedure to the SNPs that were significantly associated with the non-cognitive aspects of educational attainment from the GWAS-by-subtraction in Demange *et al.* (2021). However, the resulting PGS is not predictive of any of our measures of non-cognitive development in ALSPAC. This could reflect that the measure used in Demange *et al.* (2021) to determine non-cognitive skills (i.e. the portion of educational attainment that is not explained by cognitive skills) is very different from our measures of non-cognitive skills (the SDQ and SMFQ). Furthermore, with respect to cognitive ability, non-cognitive skills encompass a broader and harder-to-define variety of traits, which are also more likely to change over the life course. Consequently, the weights derived from GWASs in adult populations may not provide an appropriate summary of the importance of these genetic variants in the prediction of child non-cognitive outcomes. As such, we considered the GWASs of non-cognitive skills in populations of children and/or adolescents. The GWASs from the Early Genetics and Lifecourse Epidemiology (EAGLE) consortium are to our knowledge the only appropriate ones, as they use cohorts of children and adolescents to analyse internalising problems (Benke *et al.*, 2014), ADHD, (Middeldorp *et al.*, 2016) and aggressive behaviour (Pappa *et al.*, 2016).[22] As the sample sizes here are much smaller than those in adult populations GWASs, the SNP-trait associations are less-precisely estimated. We thus use less-stringent p-value thresholds to select a subset of approximately-independent SNPs via clumping: leading SNPs should have p-values of 5×10^{-5} at most, and only SNPs with a p-value lower than 0.001 can form the clumps. Other than that, the clumping procedure is the same as that described above for the depression PGS. After clumping, the PGS from the eight lead SNPs from the ADHD GWAS of Middeldorp *et al.* (2016) proved to be the most predictive of total SDQ across ages 11, 13 and 16. The PGSs from the clumped SNPs of the other two GWASs (Benke *et al.*, 2014; Pappa *et al.*, 2016), although less-robustly associated with our measures of non-cognitive skills, were used for sensitivity analysis in Table 3 (results available upon request).

Last, our results throughout the paper do not depend on either clumping or the clumping p-value thresholds. The results remain qualitatively similar when using PGSs based on all SNPs above certain p-value thresholds, regardless of LD concerns.[23]

[22]Note that the training samples here include children from the ALSPAC cohort, thus violating the standard non-overlapping condition between the training and prediction sample (as we simply use the derived PGSs as controls in Table 3, however, we do not believe that this constitutes a major problem in our context).

[23]In detail, we used p-value thresholds of either 5×10^{-8} or 10^{-6} for the PGSs for depression and cognitive skills, and thresholds of 5×10^{-5} and 0.001 for all the PGSs for non-cognitive skills.

References

Benke, Kelly S., Michel G. Nivard, Fleur P. Velders, Raymond K. Walters, Irene Pappa, Paul A. Scheet, Xiangjun Xiao, Erik A. Ehli, Lyle J. Palmer, Andrew J.O. Whitehouse, *et al.* 2014. "A genome-wide association meta-analysis of preschool internalizing problems." *Journal of the American Academy of Child & Adolescent Psychiatry*, 53: 667–676.

Demange, Perline A., Margherita Malanchini, Travis T. Mallard, Pietro Biroli, Simon R. Cox, Andrew D. Grotzinger, Elliot M. Tucker-Drob, Abdel Abdellaoui, Louise Arseneault, Elsje Van Bergen, *et al.* 2021. "Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction." *Nature Genetics*, 53: 35–44.

Middeldorp, Christel M., Anke R. Hammerschlag, Klaasjan G. Ouwens, Maria M. Groen-Blokhuis, Beate St. Pourcain, Corina U. Greven, Irene Pappa, Carla MT Tiesler, Wei Ang, Ilja M. Nolte, *et al.* 2016. "A genome-wide association meta-analysis of attention-deficit/hyperactivity disorder symptoms in population-based pediatric cohorts." *Journal of the American Academy of Child & Adolescent Psychiatry*, 55: 896–905.

Pappa, Irene, Beate St. Pourcain, Kelly Benke, Alana Cavadino, Christian Hakulinen, Michel G. Nivard, Ilja M. Nolte, Carla M.T. Tiesler, Marian J. Bakermans-Kranenburg, Gareth E. Davies, *et al.* 2016. "A genome-wide approach to children's aggressive behavior: The EAGLE consortium." *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 171: 562–572.

Turley, Patrick, Raymond K Walters, Omeed Maghzian, Aysu Okbay, James J. Lee, Mark Alan Fontana, Tuan Anh Nguyen-Viet, Robbee Wedow, Meghan Zacher, Nicholas A. Furlotte, *et al.* 2018. "Multi-trait analysis of genome-wide association summary statistics using MTAG." *Nature Genetics*, 50: 229–237.

Vilhjálmsson, Bjarni J., Jian Yang, Hilary K. Finucane, Alexander Gusev, Sara Lindström, Stephan Ripke, Giulio Genovese, Po-Ru Loh, Gaurav Bhatia, Ron Do, *et al.* 2015. "Modeling linkage disequilibrium increases accuracy of polygenic risk scores." *American Journal of Human Genetics*, 97: 576–592.

Chapter 2

Boys Don't Cry (or Do the Dishes): Family Size and the Housework Gender Gap

Boys Don't Cry (or Do the Dishes): Family Size and the Housework Gender Gap

with Anthony Lepinteur (University of Luxembourg)

Published in the *Journal of Economic Behavior and Organization*

2.1 Introduction

Recent decades have seen a shift in the distribution of housework within couples. The time women devote to household chores has fallen, while men's participation in housework has risen. However, the housework gender gap is yet to be closed in most countries: the cross-national trends in Altintas and Sullivan (2016) show that convergence has stalled since the 1980s, especially in those countries where the gap was initially smaller. The burden of housework and childcare continues to disproportionately weigh on women, with consequences in terms of labour-market outcomes and well-being. Using data from the Multinational Time Use survey, Sayer (2010) shows that women in the early 2000s carried out 1.5 to 2 times as much housework as men in developed countries. Extending the analysis to the more recent waves of the American Time Use Survey, Bianchi *et al.* (2012) confirm that a decade later American women were still responsible for about 1.6 times more housework than men. Along the same lines, McMunn *et al.* (2020) show that, in 2010, women in 93% of British couples spent more time on housework than their partners.

Standard models of household decision-making suggest that differences in bargaining power, from labour-market earnings and non-market work productivity, help determine intra-household time allocation (Chiappori, 1992; Van Klaveren, van Praag and van den Brink, 2008). Female labour-force participation and educational outcomes are at a historical high in most OECD economies (International Labour Office, 2018; World Economic Forum, 2018), with considerable consequences for women's bargaining power and the quality of their outside options (Antman, 2014; Bittman *et al.*, 2003). A large body of empirical work shows that the time spent in home production falls as either absolute or relative earnings rise (Bittman *et al.*, 2003; Gupta, 2007; Gupta and Ash, 2008; Bertrand, Kamenica and Pan, 2015) and that, conditional on being employed, educated women participate less in housework (Baxter, Hewitt and Haynes, 2008). Despite this progress, why do women then still devote disproportionately more time to housework than men?

Many researchers have turned their attention towards explanations based on gender identity formation (Akerlof and Kranton, 2000, 2010). In societies where the definition of masculinity is based on the principle that men should be the family breadwinners and should not engage in 'feminine' forms of housework, individuals will find it costly to adopt behaviours that deviate from this prescription, as it would be at odds with their identity and would translate into a utility loss. This is consistent with empirical work showing that women who are more educated or earn more than their partners, and who so deviate from gender-role prescriptions, compensate via a more traditional division of housework (Bittman *et al.*, 2003; Lyonette and Crompton, 2015; Bertrand, Kamenica and Pan, 2015). Considering the housework gender gap as a by-product of utility-maximising behaviour is however at odds with the literature on subjective well-being: Flèche, Lepinteur and Powdthavee (2018, 2020) show that the housework gender gap can be perceived as unfair and, as such, it produces lower levels of happiness and marital stability.

Comparatively little is known about the role of childhood characteristics in shaping adult differences in housework participation. Based on the intergenerational cultural socialisation framework of Bisin and Verdier (2001), some authors have shown that children's perception of gender roles are directly linked to their parents' attitudes, contributing to the persistence of unequal gender norms (Farré and Vella, 2013). Children's socialisation into traditional gender roles can also happen indirectly, as a result of the household's demographic structure. While we know that decisions involving marital status and fertility affect adults' labour-force participation (Angrist and Evans, 1998; Cruces and Galiani, 2007; Bloom *et al.*, 2009; Baxter, Hewitt and Haynes, 2008), the concomitant effects on intra-household time allocation may well involve not only parents but also children. We here consider the role of family size: assuming that the amount of housework rises with family size (Blundell, Chiappori and Meghir, 2005; Cherchye, De Rock and Vermeulen, 2012), then the time that parents move out of the labour market may not suffice to satisfy the greater demand for home production, so that children may be asked to step in and contribute more to housework (Brody and Steelman, 1985). If the effect of family size on children's housework contribution depends on their gender, a larger family size might then feed through to the adult housework gender gap, through factors such as educational achievement, future labour-market outcomes, fertility and gender attitudes.

To the best of our knowledge, the causal impact of family size on the allocation of childhood household tasks, and the persistence of this effect in adulthood, has not been explored. We address the endogeneity of family size via an instrumental-variables approach, as in Angrist and Evans (1998). In the latter the impact of fertility on women's labour supply in the US is considered using an instrument reflecting parental preferences for child sex composition: parents whose first two children are of the same sex are more likely to have a third. Similarly, we restrict the analysis to families with two or more children and exploit parents' preferences for variety in the sex mix of the offspring to predict the number of children in the household. When presenting our results, we extensively discuss the validity of the instrument in our context, following Conley, Hansen and Rossi (2012) and carrying out a number of additional tests.

In our sample of the 1970 British Cohort Study (BCS), a larger family size during childhood increases the share of housework performed by girls at age 16, but not that of boys. This conclusion is robust to different measures of housework. Girls also consistently spend less time on other activities, namely homework and leisure. The effect of family size is mostly found in low-SES and conservative households. We then show that family size at age 16 also affects the division of household tasks in adulthood: at age 34, women in the BCS who grew up in large families are more likely to perform a larger share of housework as compared to women from smaller families, and they additionally sort into households where the housework gender gap is significantly larger. We again find that the effect of childhood family size is significantly higher for cohort members who grew up in low-SES and conservative households. The results at age 42 are similar. We then argue that this persistence is in large part due to the adoption of behaviours conforming to traditional gender roles: women who grew up in large families are more likely to be not-employed and to have an employed husband.

Our paper contributes to the existing literature in a number of ways. To the best of our knowledge, we are the first to use both cohort data and an instrumental-variable strategy to estimate the causal effect of family size on the contribution of children to household tasks. The richness of our data allows us to explore the effect of family size on the time spent in other activities, such as leisure or homework. Second, we use the same instrumental variable strategy to estimate whether the effect of family size at age 16 is persistent and affects the housework gender gap of the cohort members once partnered at age 34. Last, we consider some of the channels through which childhood family size affects the adult division of housework, namely education, labour-market outcomes and fertility.

The remainder of the paper is organized as follows. Section 2 reviews two strands of the literature: the first on the link between family size and children's contribution to housework, and the second on the influence of family size on a set of determinants of the adult housework gender gap. Section 3 then describes the data and identification strategy, and the empirical results at age 16 appear in Section 4. The results at age 34 are then discussed in Section 5. Last, Section 6 concludes.

2.2 Literature review

2.2.1 The determinants of the division of housework among children

Theoretical models of household time-allocation usually consider that only adults carry out household tasks, while children, if anything, create the need for more housework (Blundell, Chiappori and Meghir, 2005; Cherchye, De Rock and Vermeulen, 2012). However, time-use surveys reveal that children actually spend a significant amount of time performing household tasks (Peters and Haldeman, 1987; Bianchi and Robinson, 1997). The contribution of children to housework can first be explained by parental time constraints: employed parents may not have sufficient time to handle the housework load and may ask their children to help them. We expect children to be imperfect substitutes for their parents, as they are likely to be less productive than adults and can only contribute to a limited set of tasks. It can also be argued that parents ask their children to help with household tasks as they wish to transmit a set of skills to them and foster their human capital (Blair, 1992*a*).

The empirical literature on children's contribution to household tasks is small and mostly non-causal. Using US data, Gager, Cooney and Call (1999) show that girls aged between 3 and 11 spend more time on housework than boys do. Girls also carry out more household tasks when their mother is employed full-time (Peters and Haldeman, 1987; Blair, 1992*a*), while the evidence is inconclusive for boys (Blair, 1992*b*). Antill *et al.* (1996) find that parental involvement in household tasks positively predicts children's housework participation.

We here focus on the effect of family size on the allocation of housework among children: according to Brody and Steelman (1985), this is ambiguous. An additional household member increases the housework load and may lead to parents asking their children to participate to a greater extent. At the same time, an additional child also increases the number of potentially helping hands in the household. The net effect of family size on the housework load per child will then be positive (negative) if the new household member's contribution is higher (lower) than the marginal increase in housework her presence entails.

Using US samples of children aged from 3 to 11 and 12 to 16 respectively, both Bianchi and Robinson (1997) and Gager, Cooney and Call (1999) find a positive relationship between family size and children's time spent on housework. These papers are the most-closely related to our first empirical question here, but neither addresses endogenenity. Family size is considered as a simple control variable. However, fertility decisions are not random and depend on confounding factors that may also be directly related to the allocation of housework among children.

2.2.2 The effects of family size in childhood on intra-household time allocation in adulthood

The housework gender gap can be defined as the difference between women and men in the time spent on housework. A body of theoretical and empirical work has aimed to understand why women still devote more time to household tasks than men do. On the theoretical side, both unitary and collective models of household decision-making suggest that the partner with the lowest earnings should spend relatively more time on housework (Stratton, 2015). Bittman *et al.* (2003); Gupta (2007) and Gupta and Ash (2008) confirm this prediction empirically: women contribute less to housework as their earnings rise. As earnings are positively correlated with human capital, we expect the housework gender gap to be smaller in households where the wife is highly-educated. Baxter, Hewitt and Haynes (2008) use Australian data to show that women with a Bachelor's degree spend less time on average on household tasks than do women without a Bachelor's degree, conditional on being employed.

While the education gap between men and women has almost closed (World Economic Forum, 2018), the housework gender gap remains. The stream of literature burgeoning from the seminal work on gender identity by Akerlof and Kranton (2000, 2010) attributes part of this persistence to gender norms. In Bittman *et al.* (2003), couples that deviate from the norm that *'a husband should make more money than his wife'* compensate by a more traditional division of housework in the US and Australia. This finding is corroborated in Bertrand, Kamenica and Pan (2015), who also show that, controlling for the absolute level of income, women with a higher probability of out-earning their husbands are less likely to participate in the labour force. An extensive Sociological literature has confirmed that, holding earnings constant, egalitarian attitudes about the gender division of labour are associated with a smaller housework gender gap (see Carlson and Lynch, 2013, for a detailed review).

Baxter (2005) and Baxter, Hewitt and Haynes (2008) emphasize the role of life-course transitions in the housework gender gap: while men's contribution to household tasks is relatively insensitive to marital status and the number of children, marriage and motherhood significantly increase that of women. Using respectively British and German data, Schober (2011) and Grunow, Schulz and Blossfeld (2012) confirm the asymmetric effect of parenthood on parental contributions to housework. Here again, Schober (2011) shows that parents with more egalitarian gender attitudes share housework more equally.

The three groups of determinants of the housework gender gap described above (i.e. education and earnings, gender norms and demographics) have one thing in common: they are all likely to be influenced by childhood family structure and, as such, are good candidates for mediating an effect of childhood family size on the adulthood division of household tasks. The paragraphs below review some of the literature describing the relationship between childhood family size and adult outcomes.

Björklund and Salvanes (2011) note that a number of contributions have found large and robust negative associations between family size and different measures of child quality, such as educational achievement and adult labour-market outcomes. This is in line with the theoretical literature on the trade-off between child quality and quantity (Becker, 1960; Becker and Lewis, 1973). However, the use of instrumental variables to address the endogeneity of fertility decisions produces more nuanced results. In Angrist, Lavy and Schlosser (2010) and Åslund and Grönqvist (2010) there is no causal effect of family size on adult educational achievement or labour-market outcomes, while other authors find negative and significant effects on private-school attendance (Conley and Glauber, 2006; Cáceres-Delpiano, 2006) and IQ (Black, Devereux and Salvanes, 2010).

Anderton *et al.* (1987), Booth and Kee (2009), Kolk (2014), and Fasang and Raab (2014) find evidence supporting of the intergenerational transmission of fertility decisions. Instrumenting for family size in Norwegian data, Cools and Hart (2017) find a differential effect of childhood family size on adult fertility by gender: an additional sibling increases male fertility but reduces female fertility. The authors argue that this difference comes from mothers reducing their labour supply relatively less when they have a daughter than when they have a son. Cools and Hart (2017) also provides descriptive evidence of a substitution effect, as girls are more likely than boys to help with housework as family size rises. Girls then become more aware of the associated strain of large families and limit their own number of children in adulthood.

One may also expect adulthood gender attitudes to be influenced by childhood family size. We know that family structure and parental background play a role in the intergenerational transmission of gender attitudes. Vella (1994) uncovers a relationship between young women's attitudes towards female employment and her parents' educational backgrounds and labor-market behaviour. Using differences in the male draft across US states as an exogenous source of variation in mothers' labour-force participation, Fernández, Fogli and Olivetti (2004) argue that men who grew up in families with working mothers develop less stereotypical gender attitudes and are less likely to be the household breadwinner. The intergenerational transmission of gender attitudes can be tested directly by correlating parental gender attitudes with those of their children. Using the NLSY1979, Farré and Vella (2013) find that the mother's views of the role of women, both in the family and in the labour market, affect the views of her children. They also show that mothers with less-traditional views about the role of women are more likely to have working daughters and working daughters-in-law (consistent with Fernández, Fogli and Olivetti, 2004). Using the British Cohort Study, Johnston, Schurer and Shields (2013) test for the external validity of Farré and Vella (2013) and find similar results. Last, Giménez-Nadal, Mangiavacchi and Piccoli (2019) use Russian panel data to infer the gender norms of the parents from the share of housework carried out by the mother, and find that conservative parents have sons and sons-in-law who perform less housework in adulthood. As it affects the allocation of household tasks among boys and girls, family size is then likely to affect children's gender norms.

2.3 Data and empirical strategy

2.3.1 The British Cohort Study (BCS)

Our empirical analysis is based on the British Cohort Study (BCS). The 1970 BCS follows the lives of more than 17,000 people born in England, Scotland and Wales in a single week of 1970. Over the course of the lives of cohort members, the 1970 BCS has collected information on, amongst others things, physical, educational and social development, health, economic circumstances and gender attitudes. Since the birth wave of the survey in 1970, there have been nine other waves ('sweeps") at ages 5, 10, 16, 26, 30, 34, 38, 42, and 46. At each sweep, different sources and methods were used to gather information on the cohort members. In the birth survey, the main questionnaire was completed by the midwife present at birth and supplementary information was obtained from clinical records. As the cohort members aged, questionnaires were administered to parents, teachers and, eventually, cohort members themselves. Medical examinations were also carried out and cohort members participated in thorough assessments.

The first outcome variable of interest during childhood is cohort member's contribution to household tasks. This is derived from the question 'What kind of things do you help with at home?', asked when the cohort member is about 16 years old. The question is followed by a set of twelve items, each depicting a particular area of contribution to housework. The items are listed in the questionnaire as follows: 'Shopping', 'Washing up', 'Cleaning the house', 'Making the bed', 'Cooking', 'Looking after elderly relatives', 'Looking after pets', 'Washing and/or ironing clothes', 'Gardening', 'Cleaning car if any', 'Painting or decorating' and 'Looking after younger children if any'. The possible answers are 'Regularly', 'Sometimes' or 'Rarely or never'.

A second housework outcome variable refers to the cohort member and their partner at age 34. Cohort members married to or cohabiting with a partner were asked to report who does most of the following household tasks: 'Shopping', 'Washing up', 'Cleaning the house', 'Cooking', 'Paying the bills', 'Looking after children when they are ill', 'Washing and/or ironing clothes' and 'Looking after the children in general'. The possible answers were 'I do most of it', 'My partner does most of it', 'We share more or less equally' or 'Someone else does it'.[1]

We further combine information on household composition at birth and age 16 to create a measure of family size. We do so by adding the number of younger siblings of the cohort member at age 16 to the number of her older siblings as reported in the birth sweep.[2] We also know the gender and birth date of all of the cohort member's siblings, which we will use to create our instrumental variable.

2.3.2 The endogeneity of family size

Our first goal is to estimate the impact of family size on teenagers' contributions to household tasks. To do so, we first show the Kernel density of the contribution to household chores of BCS cohort members at age 16 (calculated as the share of household tasks the cohort member helps with 'Regularly', as opposed to 'Sometimes' or 'Rarely or never'), by gender and family size (Figure 2.1). Consistent with the extant literature, we find that, for any family size, girls contribute more to household tasks than boys do. The descriptive results in Figure 2.1 further suggest that while family size does not much affect boys' contribution to housework, girls in larger families spend more time on household tasks than girls in smaller families do.

[1] Survey-derived housework measures are usually biased as compared to the more accurate time-use diaries (Geist, 2010). Kan (2008) documents a systematic inflation in housework measures derived from stylised questionnaires rather than time-use data, the gap being larger for men than for women. In a similar spirit, Achen and Stafford (2005) show that the amount of time married men report spending in housework is larger than the time their wives report for them. However, if reporting biases are orthogonal to the instrumental variable we use in our empirical analysis, there is no reason to believe that our main estimate will be affected. Second, in contrast with the literature, we here do not measure hours spent in housework, but the relative contribution of individuals as compared to their partners – a measure which we believe to be less susceptible to measurement error.

[2] As our identification strategy relies on the gender composition of the two first-born children in the household, we measure family size as the total number of siblings. We are aware that this measure may include siblings who had already left the household when the cohort member is age 16. We address this potential concern by using the number of children living in the household at the fourth survey sweep as an alternative measure of family size. The use of this alternative measure produces even stronger results (available upon request), although, as expected, the instrument appears to be slightly weaker.

The evidence in Figure 2.1 is suggestive of a role of family size for girls, but does not address endogeneity. The distribution of fertility across households cannot be assumed to be random, as it depends on a set of both observable and unobservable household characteristics that may well be correlated with household tasks both during childhood and adulthood. For example, BCS family size at age 16 is larger when the mother is not employed and has conservative opinions about maternal employment. Being on average less educated and less likely to be employed, mothers in large families will mechanically have more time to spend on housework, which in turn has a crowding-out effect on childrens' own contribution. Naive specifications that do not account for negative selection into parenthood and the time-use of mothers may thus underestimate the true effect of family size on childrens' contribution to housework.

The endogeneity of family size is commonly addressed via instrumental variables. A first popular strategy is to instrument the size of families with at least two children by the sex composition of the two first-born children. To the best of our knowledge, Angrist and Evans (1998) is the first influential work to use this strategy, and estimates the causal impact of family size on women's labour supply. The rationale here is that parents have a preference for variety: a couple with the first two children of the same sex is more likely to try and have a third, relative to a couple whose first two children are a boy and a girl. As the sex mix of children can be seen as random, the instrument provides the exogenous variation necessary for plausible identification. This approach has been widely-used in the literature to assess the impact of family size on a variety of child outcomes, such as education, fertility and labour market outcomes (Angrist, Lavy and Schlosser, 2010; Black, Devereux and Salvanes, 2010; Cools and Hart, 2017). Section 4.2 provides a discussion of the validity of the same-sex instrument applied to our context.

Multiple births can also be seen as a source of exogenous change in family size. A number of articles have used twin births as an instrument to estimate the causal impact of family size on outcomes such as women's labour supply (Rosenzweig and Wolpin, 2000; Angrist, Lavy and Schlosser, 2010) and children's education (Black, Devereux and Salvanes, 2005; Cáceres-Delpiano, 2006; Åslund and Grönqvist, 2010). However, using individual data on 17 million births over 72 countries, Bhalotra and Clarke (2019) underline that twin births are systematically positively correlated with maternal health. This finding is robust to a battery of tests and casts doubt on the validity of multiple births as an instrument for family size.

2.3.3 Empirical strategy

We account for the endogeneity of family size by following the instrumental-variable approach in Angrist and Evans (1998). Our instrument is a dummy for the first two children of a couple being of the same sex. We do not make use of multiple births in our main analysis for a number of reasons. First, as noted above, Bhalotra and Clarke (2019) suggests that multiple births may not be random and can reflect positive selection into motherhood that could bias our estimates. Second, our estimation sample is of limited size and the lack of statistical power may be prejudicial to our analysis; a similar concern is raised by Black, Devereux and Salvanes (2005) when considering the results from estimation samples of sizes comparable to ours.[3]

We first estimate the following model by Two-Stages Least Squares (2SLS):

$$FamSize_i^{16} = \alpha_1 SameSex_i + \delta_1 X_i + \epsilon_i$$

$$HhTasks_i^{16} = \alpha_2 \widehat{FamSize}_i^{16} + \delta_2 X_i + \mu_i \tag{2.1}$$

where $FamSize_i^{16}$ is family size at age 16, calculated as the total number of siblings of the cohort member plus one, and $HhTasks_i^{16}$ is the contribution to household tasks of individual i at age 16, calculated as the share of household tasks the cohort member helps with 'Regularly' (as opposed to 'Sometimes' or 'Rarely or never'). We use $SameSex_i$, a dummy for the first- and second-born children being of the same sex, as an instrument for $FamSize_i^{16}$. X_i is a vector of standard controls, including dummies for sex and the child being of European descent. We measure parental education by the age at which the parents left school, and include a dummy for the child's parents still living together in 1986. We additionally exploit information on the study child's father's socio-economic status (SES), observed at child age 0, as proxied by his occupational status (following the Registrar General's 1966 Classification of Occupation). In particular, we control for a dummy equal one for being in either a professional, managerial, or non-manual skilled occupation ('high SES'), and zero for being either a skilled manual, semi-skilled, or unskilled worker, or unemployed ('low SES'). We also include an index constructed by the data providers to measure the mother's attitudes towards maternal employment, when the study child was 5 years old. Controlling for this

[3] Despite its limitations, we do ultimately use twin births as an instrument for family size as a supplementary robustness check, as suggested in Angrist, Lavy and Schlosser (2010). While the estimates are in line with our main results, none are significantly different from zero at conventional levels. As suggested by Black, Devereux and Salvanes (2005) and Cools and Hart (2017), this lack of precision may come from the limited sample size. The results from this additional identification approach are available upon request.

variable attenuates the possibility of the $SameSex_i$ instrument to be spuriously capturing differences in fertility responses correlated with heterogeneous adherence to conservative gender norms. We control for a potential independent effect of the gender mix of all the siblings by adding a dummy named *Balanced* for there being at least two siblings of different sex in the family. We finally control for birth order dummies.

Our first estimation sample covers individuals from families with at least two children and with valid information on both the household tasks performed at age 16 and the controls. This produces 3,389 observations. One reported task out of four is performed 'regularly' and the average family size in our estimation sample is 2.8. The full descriptive statistics on this estimation sample can be found in Table 2.A1 in Appendix 2.A. Only 6,349 out of roughly 13,000 solicited families completed and returned the questionnaire measuring children's contribution to housework (the 'Document G: Home and All That'). We ask in Table 2.A2 whether children from our estimation sample differ significantly from those with similar characteristics but who did not complete 'Document G: Home and All That' (i.e. children with at least one sibling and with valid information on the controls, but no information on the household tasks performed at age 16). Children in our estimation sample have on average a better family background (higher-SES households, more educated parents and a more stable parental relationship) and are mostly girls. This is not surprising as male survey respondents, as well as respondents whose parents have low levels of education, typically have higher attrition rates and non-response rates with respect to females (Mostafa and Wiggins, 2015).[4] We find a similar pattern of selection when comparing our estimation sample to the overall BCS population with non-missing information on the controls.

We then ask whether family size at age 16 continues to influence the time devoted to housework at age 34, via a second 2SLS model:

$$FamSize_i^{16} = \beta_1 SameSex_i + \delta_3 X_i + \epsilon_i$$

$$Y_i^{34} = \beta_2 \widehat{FamSize}_i^{16} + \delta_4 X_i + \mu_i \tag{2.2}$$

Here Y_i^{34} corresponds to one of the three following dependent variables measured at age 34: the share of household tasks carried out by the wife (or female partner), the share of household

[4]Additionally, based on observable characteristics, we do not find any evidence that girls and boys select into non-response in systematically different ways (except for father's age: having an older father significantly increases the probability of returning the questionnaire for boys, while it plays no role for girls). These results are available upon request.

tasks carried out by the husband (or male partner) and the housework gender gap (the difference between these two shares). The vector X_i includes the same control variables as in model 2.1. We do not control for the socio-demographic characteristics of the cohort members at age 34 (e.g. labour-force status, number of children) as we suspect that these may mediate the effect of family size in childhood and, as such, are 'bad controls' (Angrist and Pischke, 2008). We explore this potential mediation in Section 5.2 using the decomposition approach in Gelbach (2016).

Our second estimation sample covers individuals who are in a partnership at age 34, with at least one sibling at age 16, and with valid information on the household tasks performed at age 34 and on the controls. This produces a sample of 3,200 observations.[5] The cohort members in our estimation sample sort on average into couples where about half of housework is only carried out by women. Only 15 percent of the tasks are only carried out by men. Additional descriptive statistics for this sample are shown in Table 2.A3.

2.4 Family size and the contribution to household tasks at age 16

2.4.1 Main results

Table 2.1 shows both the OLS and 2SLS estimates of model 2.1. The first two columns refer to the whole sample of households with at least two children, while the sub-samples by child sex appear in columns (3) through (6). The main variable of interest is family size. In the OLS estimates in column (1), an additional household member has a positive and significant impact on the contribution of the study child to household tasks. When we instrument *Family size* by *Same sex*, the 2SLS results in column (2) also reveal a positive and significant coefficient on *Family size*.[6] This result is in line with Bianchi and Robinson (1997) and Gager, Cooney and Call (1999): larger family size increases the contribution of children to household tasks. Looking at the estimates in column (2), one additional sibling increases the share of tasks performed 'regularly' by 5.4 percentage points.

[5]The estimation samples in childhood and adulthood are not the same size. This is because only a sub-sample of the cohort members reported their contribution to housework tasks at age 16, while the age-34 housework questionnaire was administered to all partnered cohort members. Keeping only the cohort members who appear in both estimation samples reduces the number of observations drastically (under 1,500 observations). While such selection does not affect the size of our estimates, the smaller sample size does reduce their precision. The results are available upon request.

[6]We have also re-estimated all our regressions using a dummy 'Having at least three children' rather than all discrete values of family size, with the results remaining qualitatively unchanged.

This effect is equal to 25 percent of a standard deviation of the dependent variable. In terms of magnitude, the effect of family size on the share of housework is equivalent to about 60% of the effect of being a girl.[7]

Figure 2.1 suggested that the effect of family size was mainly found for girls. We formally check whether there is a difference between boys and girls in columns (3) to (6) of Table 2.1. Both OLS and 2SLS estimates confirm that an increase in family size translates into a significantly higher contribution of girls to household tasks at age 16, while it does not affect the contribution of boys. The positive family size coefficient in columns (1) and (2) is thus mostly driven by girls.[8] Note that when we do control for variables that are arguably endogenous (e.g. children's cognitive skills and mothers' labour force participation), estimates for family size are not statistically different from results in Table 2.1.

While both the OLS and 2SLS estimates are positive and significantly different from zero at 5% level, both in the overall sample and in the subsample of girls, the OLS estimate is smaller than that from IV. This is in line with our hypothesis that the negative selection into parenthood reduces the true effect of family size in OLS. The difference between the OLS and 2SLS estimates might also be due to the fact that the former is expected to capture the influence of family size over the whole estimation sample while the latter is a Local Average Treatment Effect (LATE) for the compliers, that is children whose parents are more likely to have a third child if the first two are of the same sex. One could argue that the preferences for variety in the offspring sex-mix are not randomly distributed across the population: individuals with more conservative views may care more about a mixed-sex offspring pool, as compared to progressive individuals. If that was the case, parents whose first two children are of the same sex and who decide not to have a third one may hold systematically more progressive views than parents in the same position who instead decide to have a third child.

What does this mean for the interpretation of our LATE? The effect sizes found in columns 2 and 4 might be partly driven by the fact that larger weights are attached to the second-stage treatment effect of those families in which fertility decisions are more sensitive to the same-sex instrument – that is, arguably more conservative families. As conservativism is likely correlated with a more gender-stereotypical assignment of housework to children, our 2SLS estimates might be thus inflated. We address this potential source of concern in two ways: first, all our regression control

[7] One may argue that we should not compare the effect of a discrete variable to that of dummies. When we dichotomise family size as a dummy for having at least three children, we find that the estimated coefficient has the same magnitude of that of the female dummy.

[8] We also follow Wooldridge (2010) and use same-sex and its interaction with the cohort member's gender to instrument family size and its interaction with a female dummy. The difference between boys and girls continues to be significantly different from zero at the 5% level.

for the mother's attitudes towards maternal employment, a proxy of the family's conservativism. Second, we check whether we find any evidence of conservative parents holding stronger preferences for a mixed-sex offspring. As families with the same-sex instrument equal to zero already achieved a composite gender mix, we here focus only on families in which the first two children are of the same sex. Using a range of indices capturing the mother's gender attitudes in 1975, Table 2.A4 shows that, conditional on being assigned to the treatment (i.e. same-sex equal one), women who have a third child do not hold systematically more conservative views than those who stop at two children. This suggests that preferences for mixed-sex offspring are not systematically driven by conservative attitudes in our estimation samples.

One may suspect a smaller effect of family size in families that outsource their home-production in the market. The outsourcing of housework is not accurately measured in the BCS, so we use the father's SES as a proxy for the probability of hiring help. We expect the effect of family size to be smaller in high-SES families, as they are more likely to hire in help for home production, thus decreasing the need for children to contribute to housework. Net of concerns on families' reactivity to the same-sex instrument, we may also expect gender attitudes to be a source of heterogeneity for the effect of family size – the effect being arguably larger for children from conservative households. Traditional parents might believe, for instance, that their daughters (but not their sons) will face a marriage-market premium when endowed with a set of domestic skills (this is consistent with the matching model developed by Chiappori, Iyigun and Weiss, 2009, under the assumption of traditional household roles). We empirically test the presence of these two sets of channels in Table 2.2, where we split the estimation sample by gender, father's SES, and mother's adherence to conservative gender norms (above or below the third quartile of the distribution of the measure of mothers' attitudes towards maternal employment discussed above).[9] Panel A of the table shows that family size significantly increases the contribution of girls to housework in low-SES families but not in high-SES families. In Panel B, the effect of family size on boys' contribution to household tasks remains insignificant in both cases. We also see that girls with conservative mothers are more likely to contribute to household tasks as family size rises; there is no significant effect for girls with non-conservative mothers or for boys.[10]

In Appendix 2.B we show that our results are robust to alternative measures of children's contribution to housework, as well as to the aggregation of subsets of tasks according to their

[9]The results using the median of the mothers' gender attitudes distribution are qualitatively similar.

[10]We have also looked at other sources of heterogeneity, and find that the effect of family size is stronger for girls who are first- or second-born and whose mother is young.

characteristics (e.g. ‘feminine’ vs ‘masculine’ tasks). Additionally, we explore the time children spend in other activities (homework and leisure) and find consistent results suggesting that girls from larger families dedicate less time to homework and leisure activities.

2.4.2 Assessment of the identification assumptions

For a valid causal interpretation of our 2SLS estimates, the following five assumptions need to hold (Angrist, Imbens and Rubin, 1996):

- Assumption 1, SUTVA: in our context this is equivalent to saying that the time spent in household tasks is not affected by the sex composition of the first two children born in other families. Additionally, SUTVA requires the treatment (an increase in family size after the first two children) to be the same across all individuals.

- Assumption 2, Independence: the sex composition of the two first-born children is uncorrelated with any confounder of the association between family size and household tasks.

- Assumption 3, Relevance: having two first-born children of the same sex predicts a significant increase in family size.

- Assumption 4, Monotonicity: there are no parents whose preferences are such that they have more children only if their first two are of different sex, and stop having children if their first two are of the same sex.

- Assumption 5, Exclusion Restriction: once family size is taken into account, the time spent in household tasks is not affected by the sex composition of the two first-born children.

2.4.2.1 From Assumption 1 to 4

We here discuss how plausible the assumptions above are in our context. Assumption 1 is uncontroversial: our setting does not exploit within-family variations and we do not expect major general equilibrium effects from the sex-composition instrument. Except in the rare cases of multiple birth, we can also safely assume that the treatment (an increase in family size after the first two children) is the same across individuals. Sex at birth being random, Assumption 2 is also indisputable: conditional on having at least two children, the sex-composition of the first two is random. The relevance assumption can be shown to hold by looking at the first-stage regressions of model (1), as summarised by the top section of Table 2.A5. The instrument’s coefficient is positive and highly-significant in all subsamples and the Cragg-Donald Wald F-statistics (also reported at the bottom of Table 2.1) confirm the instrument's strength. Last, Assumption 4 holds if they are no defiers. Defiers in our context would be parents with a strict preference for having at least two children of the same sex, therefore choosing to have a third child or more only if the first two are of different sex. While the existence of this category is unlikely in our case (a 98%-white British sample), it would not be surprising in different cultural context, where son preference is for instance a wide-spread social norm (Lee, 2008; Rosenzweig and Wolpin, 2000). In order to rule out any systematic preference for one particular gender, we split the same-sex instrument into two dummies: one for the first two children being boys and one for them being girls. Both instruments are equally predictive of subsequent fertility behaviour, suggesting that indeed British parents in our estimation sample do not display preferences for a particular gender of their offspring.

2.4.2.2 The Exclusion Restriction: a discussion

Our identification strategy would prove problematic if the same-sex instrument were to be correlated with the dependent variable through a channel other than family size. While Angrist and Evans (1998) used mothers' labour-force participation as their main outcome, we here consider a child-level outcome that may well be correlated with the sex composition of the two first-born children. One way this could happen is via a crowding-out effect: if children of one particular gender systematically contributes more to housework, then having a sibling of such gender may reduce the residual amount of housework to be carried out by the child. If we assume ex-ante one of our key findings, that is girls contribute more to housework than boys (Gager, Cooney and Call, 1999), and keeping everything else constant, having a sister will always decrease the residual pool of housework a child can contribute to. While we partially take sibling composition into account by including a dummy for there being at least a girl and a boy in the household, we cannot fully rule out empirically the presence of such channel.

Before discussing this channel any further, we first investigate whether the data suggest in any way the presence of a direct effect of the instrument on the outcome variable. We do so by carrying out a placebo test and then looking into systematic correlations between the instrument and a number of observed covariates. The first test relies on the following intuition: if the instrument were to be correlated with the dependent variable through a channel other than family size, we would expect to find a significant 2SLS family-size estimate even for household tasks that should not be affected by family size. We here appeal to the estimates shown in the fourth and fifth rows of

Panel B of Table 2.B1. These show, respectively, the estimated family-size coefficients for household tasks that are likely increasing in family size and those that are not. None of the 2SLS family-size estimates are significant for the latter (in the fifth row) while most of the estimates in the fourth row are significant, suggesting that the instrument is unlikely to significantly affect the contribution to household tasks other than via its impact on family size.[11] We then look for the presence of any systematic correlation between the instrument and observable characteristics of the cohort members at age 16 and their parents. We follow Angrist, Lavy and Schlosser (2010) and Falck, Gold and Heblich (2014) and derive reduced-form estimates from the regression of cohort members' and their parents' characteristics at age 16 (normally used as control variables in our baseline regressions) on the same-sex instrument and all other controls. If observable characteristics were to be correlated with our instrument, we may expect the same for unobservable characteristics too. Table 2.A6 shows that the only covariate that is significantly explained by the same-sex instrument is mother's age at the birth of the cohort member. While we may worry about this correlation, the estimates are arguably small in economic terms: for example, in column (1) mothers whose two first-born children are of the same sex are on average five months younger than mothers whose first two children are of different sex. The absence of selection into consecutive same-sex pregnancies based on observable characteristics mitigates our concerns about systematic correlations with unobservable characteristics.

While reassuring, the absence of evidence suggesting the instrument has a direct effect on housework contribution alone is not sufficient to completely rule out any threat to the exclusion restriction. We then draw from Conley, Hansen and Rossi (2012) and relax the exclusion restriction, by allowing the instrument to be only 'plausibly' exogenous. The intuition behind the method in Conley, Hansen and Rossi (2012) is to allow the instrument to have a direct effect on the dependent variable in the second-stage of the 2SLS estimation. We here follow the application in Nybom (2017) and express model 2.1 as follows:

$$FamSize_i^{16} = \alpha_1 SameSex_i + \delta_1 X_i + \epsilon_i$$

$$HhTasks_i^{16} = \alpha_2 \widehat{FamSize}_i^{16} + \lambda\gamma SameSex_i + \delta_2 X_i + \mu_i. \tag{2.3}$$

[11]This is assuming that all tasks are potentially equally-affected by the instrument, a condition under which our restriction of the analysis to a set of tasks that is arguably insensitive to family size comes without loss of generality.

where γ is the instrument's reduced-form effect and λ is a parameter varying between 0 and 1. As explained in Nybom (2017), the adjusted effect of family size mechanically converges towards zero as λ converges to one. Figure 2.A1 shows the estimates of family size instrumented by the sex-composition of the first two children in the household, first for the whole sample and then separately for girls and boys. If λ is greater than 0.38 for the whole sample and greater than 0.45 for girls, the 2SLS estimates of instrumented family size are no longer significantly different from zero. This is equivalent to saying that, as long as the direct effect of the instrument is respectively smaller than 38% and 45% of the instrument's reduced form effect, the effect of family size on housework will remain significantly different from zero for the whole sample and for girls.

Conceptually, and in line with the plausible exogeneity exercise, the bias induced by a violation of the exclusion restriction would be problematic (i.e. result in an over-estimation of the effect of family size) only in case the same-sex instrument caused an increase in housework participation, through a channel other than family size. A plausible channel, as mentioned above, could be via the crowding-out effects coming from the sisters' housework participation. Table 2.A7 presents a topology of all possible sibling compositions, given the value of the same-sex instrument. Consider first the case in which the cohort member is the first- or second-born. If the cohort member is a boy, then the instrument Z will take value 0 if he has a sister (case *a*) and 1 if he has a brother (case *b*).[12] If we assume that girls contribute more than boys to household tasks, then the cohort member will be more likely to perform a larger share of housework in case *b* than in case *a*. Hence, with 2SLS we would tend to overestimate the family-size coefficient due to the positive correlation between the instrument and the dependent variable, given all other covariates. This would be problematic if we found a positive effect of family size on the share of housework performed by boys, as we would not be able to distinguish whether the effect comes from the real association between the variables or the violation of the exclusion restriction. However, since we find no statistically-significant effect of family size on the share of housework for boys, the real effect should be either zero or negative - which in either case corroborates the finding that the effect of family size is larger for girls than for boys.

Now consider the case where the cohort member is a first- or second-born girl. The instrument takes value 0 when she has a brother (case *c*) and 1 when she has a sister (case *d*). With the same assumption as above, the cohort member will be more likely to contribute more to housework in case

[12]Here, when talking about brothers and sisters in relation to first- or second-born cohort members, we refer to the brother or sister that, together with the cohort member, makes up the pool of the two first-born children.

c than in case d. Here, conditional on the controls, the instrument would be negatively correlated with the dependent variable. We again are not particularly worried about this potential violation of the exclusion restriction, as it would bias the coefficient of family size towards zero.

We finally consider the case where the cohort member is neither the first- nor the second-born. The instrument now takes the value of one in two occurrences: either the two first-born children are both boys (cases b_2 and d_2) or both girls (cases b_1 and d_1). Irrespective of his or her gender, a cohort member with two older sisters would tend to perform relatively less housework compared to the case where the instrument is zero (cases a_1 and c_1). As argued above, the 2SLS estimate of family size would then be biased toward zero. Instead, when the cohort member has two older brothers (cases b_2 and d_2) there will be comparatively more housework to do and he or she might be asked to contribute relatively more than in the case where the two eldest siblings are of opposite sexes. We may here expect the 2SLS estimate to overestimate the effect of family size. This is the most worrying case, as the effect size we estimate is potentially inflated. To check whether the sample of cohort members with this particular sibling mix is behind our results, we replicate our main results from Table 2.1 excluding the 236 cohort members who have two older brothers. Consistent with the results in Table 2.1, the 2SLS point estimate of family size is 0.061 for the whole sample and 0.088 for girls (both significant at the 1% level).

What do we empirically know about the effect of the sibship sex-mix on children's contribution to housework? While there are plenty of papers relying on the same-sex instrument, the number of studies linking it to housework contribution in childhood is scarce. One exception is the descriptive work of Schulz (2021), which assesses the influence of a variety of factors (among which, the siblings sex composition) on the time children spend performing household chores. Using the German Time Use Study, Schulz (2021) shows that, while girls spend on average more time in housework than boys, each child's contribution to housework is independent of their siblings' gender (similar results are also found in Cordero-Coma and Esping-Andersen, 2018). Assuming these conclusions would hold in our context as well (as somewhat already suggested by Tables 2.A6 and 2.B1), the direct effect of the same-sex instrument on the dependent variable would neither be positive (making the positive upper bounds derived from Figure 2.A1 implausibly high) nor negative (in contrast with what we would expect theoretically from most of the cases described in our topology above), but rather nil.[13]

[13] It is important to stress that the exclusion restriction might also be violated for infra-marginals (always-takers and never-takers). If this was the case, our 2SLS estimates would be biased (see Jones, 2015, for technical details). Based on the adjustment via calibration formula proposed by Jones (2015) and assuming no direct effect of the same-sex instrument on the outcome for never-takers, we can express the bias as the product of the effect of the instrument on the outcome for always-takers and the share of always-takers over the share of compliers, i.e. $\eta_{AT}\frac{\pi_{AT}}{\pi_C}$. Taking estimates for π_{AT} and π_C from Kowalski (2019), we can derive

2.5 Family size and contribution to household tasks at age 34

2.5.1 Main results

We now ask whether the effect of family size at age 16 persists into adulthood and affects the division of housework in households formed by BCS respondents and their partners at age 34. To do so, we replicate our 2SLS analysis using as the dependent variables the share of household tasks performed by the female partner, by the male partner, and the housework gender gap (i.e. the difference between the two shares). Before presenting our estimates, it is important to verify that growing up in a large family in childhood (instrumented by the gender composition of the first two siblings) does not influence the probability of being in a partnership at age 34 and, hence, being in our estimation sample. We rule out this concern of endogenous selection in Table 2.A8 by showing that the instrumented family size at age 16 does not affect significantly the probability of being in a partnership at age 34.

Having found no evidence of selection into cohabitation based on the same-sex instrument, we now turn to the main results in our adult sample described in Table 2.3. The 2SLS estimates for the whole sample (Panel A) confirm a persistent effect of family size at age 16 on the division of household tasks at age 34. Larger families at age 16 predict a greater share of household tasks done by women, while the male share remains unchanged. As expected, column (3) then shows that the larger the family at age 16, the greater the housework gender gap at age 34. As such, cohort members who grew up in larger families sort into couples that conform more to stereotypical gender roles and in which the housework gender gap is even larger.

In Panels B and C of Table 2.3 we then ask whether this result is stronger for men or women. It appears that only women sort into households with a significantly larger housework gender gap as family size at age 16 rises. As revealed in columns (1) and (2) of Panel B, this is mostly explained by a significantly higher share of household tasks predominantly carried out by the wife.[14] In the bottom Panel of Table 2.3, there is some evidence that male cohort members who grew up in

the minimum η_{AT} such that the true effect of family size on girls' housework is not statistically different from zero, that is $\eta_{AT} = 0.05$. In other words, as long as $\eta_{AT} < 0.05$ the true effect of family size remains positive and significant. As we argued above, most of the cases we derived from the topology in Table 2.A7 suggest that the direct effect of the same-sex instrument on the outcome should be theoretically negative. Following the evidence in Schulz (2021) and Cordero-Coma and Esping-Andersen (2018), η_{AT} has been descriptively shown to be zero in contexts similar to ours.

[14] We also find that the share of tasks to which men and women contribute equally and the share of tasks performed by someone else are unaffected by family size in childhood. These results are available upon request.

larger families have a larger housework gender gap in their adult household, although the estimated coefficient here is not statistically significant.

The previous section established that there is no significant effect of family size on the contribution to household tasks for respondents who grew-up in high-SES families or with relatively progressive mothers. We again look at this kind of heterogeneity, with the results for females appearing in the first row of Table 2.4. The pattern here is similar: larger families have no impact on the contribution to household tasks and the housework gender gap at age 34 for women from relatively well-off families and non-conservative families. Consistent with the childhood results, women raised in large low-SES families contribute significantly (at the 5% level) more to household tasks and sort into couples with a higher housework gender gap. We find no significant results for men (see Table 2.A9 in Appendix 2.A). We can replicate this analysis for cohort members at age 42, to ensure that our estimates are not a statistical artifact driven by the choice of a particular survey year: the results in Table 2.A10 are qualitatively similar.

The assumptions for a valid causal interpretation of all the estimates in this section are the same as the ones discussed in Section 4.2. While the same considerations made about the first four assumptions (SUTVA, independence, relevance and monotonicity) hold regardless of the different outcomes we consider here, it may be relevant to discuss how the exclusion restriction translates in the adulthood context. As cohort members are now living with a partner and not anymore with their childhood family members, we can rule out the presence of any sibling crowding-out effects on the pool of housework to be performed. Sibling sex-composition could however still affect individuals' contribution to housework in adulthood via peer effects: interacting with a sister (brother) who formed a household where she (he) does most (none) of the chores might set an example for our cohort members. However there is only limited evidence that this mechanism might be in place (Nicoletti, Salvanes and Tominey, 2018) and, if in place, we believe its effects to be of second order.

2.5.2 Channels

Why does family size at age 16 continue to explain the individual's contribution to housework 18 years or more later? In Table 2.4 we explore the role of family size on adult characteristics which are likely to help shaping the housework gender gap of female cohort members (we replicate the exercise in Table 2.A9 for male cohort members).

Section 2.2 suggested that we might expect children who grew up in larger families to have lower education and thus worse labour-market outcomes. We investigate this in Table 2.4 by estimating the causal effect of family size at age 16 on the school-leaving age and the probability of having at least an A-level at age 34. Most of the estimated coefficients on family size are not significantly different from zero. This finding continues to hold with other measures of educational attainment and is consistent with the results in Cáceres-Delpiano (2006), Black, Devereux and Salvanes (2005) and Angrist, Lavy and Schlosser (2010). Note that women who grew up in large families with non-conservative mothers are the only exception, as they are more likely to have an A-level and left full-time education at an older age.

We then look at the effect of age-16 family size on a set of labour-market outcomes, namely employment, the monthly wage (in logs), weekly working hours and commuting time. Commuting in BCS is measured in time bands, and we here create a dummy for commuting time of over 30 minutes. Only women from low-SES families have a significantly lower probability of being employed and, when employed, spend less time commuting. This is in line with the burgeoning literature on gendered preferences over workplace amenities (Mas and Pallais, 2017) and local labour markets (Manning and Petrongolo, 2017). Our results provide indirect evidence that the definition of local labour market might differ by gender, due to the different costs associated with distance from the workplace: women might face social constraints that confine them to 'even more local' labour markets.

Only limited information is available on the partners of BCS respondents. However, we can estimate the causal effect of family size on the probability of having an employed partner. We find here positive and significant estimates in almost all our samples: women who grew up in large families tend to sort into couples where their partner is more likely to be employed.

We now turn to life-course transitions. According to Baxter, Hewitt and Haynes (2008) the housework gender gap does not change with marriage but does increases as individuals enter parenthood, and it has been shown that fertility is transmitted across generations (Anderton *et al.*, 1987; Booth and Kee, 2009; Kolk, 2014; Fasang and Raab, 2014). The effect of family size at age 16 on the housework gender gap might therefore transit via the cohort members' own number of children. Table 2.4 asks whether family size affects the probability of being married at age 34, as well as the probability of being a parent and the number of children. There is some evidence that family size increases the probability of marriage. Our fertility results are somewhat in line with Cools and Hart (2017): only men's fertility decisions are positively influenced by their own family size in childhood, but not at conventional significance levels (see Table 2.A9). On the contrary, women's fertility decisions are not affected by their number of siblings and hence do not lie behind the effect of family size at age 16 on the housework gender gap at age 34.[15] As for the exclusion restriction, a discussion similar to that in Section 5.1 could be applied to this context as well. See Angrist, Lavy and Schlosser (2010) for an additional extensive discussion of the same-sex instrument's validity when using labour market outcomes, educational outcomes, and marriage and fertility decisions (see also Cools and Hart, 2017, for a discussion on the latter).

The results above suggest that women who grew up in large families, and particularly low-SES families, sort into partnerships with more traditional gender roles, i.e. with a lower probability of employment and shorter commuting time for the woman, a higher probability of employment for the husband, and a greater probability of being married rather than cohabiting. But does the adoption of these gender roles fully explain the persistence of the housework gender gap? In order to answer this question we here follow the decomposition approach developed by Gelbach (2016). The decomposition relies on the omitted-variables bias formula and can be used to attribute a portion of the treatment effect to potential mediators. Results from this exercise are presented in Table 2.5, first for all women in our sample and then only for women from low-SES families and families with conservative mothers. For each of the three subsamples, the Table reports three columns: the first, 'Base', reports the coefficient for family size from the baseline regression (model 2 in Section 3.3); the second, 'Full', further augments the model specifications by adding the four potential mediators (not in employment, commuting time, having an employed partner, being married); last, in column 'Expl.' the difference between the family size coefficient in the baseline specification and the full specification is decomposed into portions that can be explained by each of the potential mediators – conditional on controlling for all of them simultaneously.

The first three columns of Table 2.5 show that about half of the estimated effect of family size on the housework gender gap for women can be explained by the channels we included in the full model specification. While all channels display a positive contribution, only having an employed partner appears to be a significant mediator. Moving on to women from low-SES backgrounds, columns 4 to 6 show a similar narrative – with labour market outcomes (labour force participation and commuting time) significantly contributing to explaining the persistent effect of family size on

[15] One may worry that the significance of our estimates in Table 2.4 results from multiple hypothesis testing. Following Falck, Gold and Heblich (2014), we assume each regression to be an independent draw (where the significance at conventional levels of the family size coefficient is seen as 'success') and exploit the properties of a binomial distribution to derive how likely it is that the coefficients in Table 2.4 are statistically significant only by chance. In the Table, 14 out of 55 estimated coefficients are statistically significant at least at a 10% level. The probability that 14 or more out of 55 coefficients are significant at the 10% level by chance is only 0.08%. In the subsample of regressions based on women who grew up in low-SES families – for whom our results appear to be stronger – the probability that 4 or more coefficients are significantly different from zero at the 10% level by chance is 1.85%.

the adult housework gender gap. No change is observed instead across the two specifications in the subsample of women who grew up with conservative mothers. Although only descriptive, these results confirm that the long-lasting influence of family size in childhood on the housework gender gap can be largely attributed to the adoption of behaviours that conform to traditional gender roles.[16] Results for the subsample of women from high-SES families and women with non-conservative mothers are displayed in Table 2.A11. We also replicate this analysis for men in Tables 2.A12 and 2.A13.

2.6 Conclusion

In this paper we have assessed the impact of childhood family size on the allocation of household tasks of British Cohort Study cohort members at age 16 and then at age 34. We account for the endogeneity of fertility by exploiting parents' preferences for variety in the sex mix of their offspring, and use the sex composition of the first two children as an instrumental-variable predictor of family size.

We find that family size significantly increases the probability that adolescents contribute to housework at the age of 16. However, we show that our estimates substantially differ by gender: only girls do more housework as the family size increases. This finding is not sensitive to the measurement of housework, and girls also spend relatively less time on leisure and homework in larger families. There is also heterogeneity by father's SES, as only girls whose father has low SES do more housework as family size rises. This is consistent with high-SES parents being more likely to outsource housework and less likely to ask their children to help with the chores. We also find that the effect of family size on housework at age 16 is larger for girls whose mothers hold conservative attitudes.

The effect of family size in childhood is persistent: at age 34, female cohort members who grew up in large families are more likely to sort into couples in which the housework gender gap is significantly larger with respect to women from smaller families. We again find that women from low-SES families and with conservative mothers are behind this finding. We show that the long-term effect of childhood family size is explained by the adoption of behaviours that are in line with more conservative gender roles. First, women who grew up in large families are less likely to be employed, and when they are employed their commuting time is significantly shorter. They are also more likely

[16] Additionally including the other adult outcomes shown in Table 2.4 in our mediation analysis does not appear to further attract part of the estimated effect of family size.

to be married to employed partners who, in return, have less time to spend on household chores.

Contextualising our results in an identity formation framework à la Akerlof and Kranton (2000, 2010), it can be argued that women who grew up in large families maximise their utility by respecting the behavioural prescriptions of the traditional gender attitudes into which they were socialised. If this were the case, women would find it fair (or, at least, not sub-optimal) to do more housework than their male partners and there would be no direct cost in terms of welfare (Flèche, Lepinteur and Powdthavee, 2018, 2020). However, identity can be seen as a narrative, and as such can be interpreted as a flexible concept (Sveningsson and Alvesson, 2003; Ashforth, 2000). More specifically, it can adapt to act as a buffer against adverse life events (as in Ibarra, 2003, where changes in working identity are seen as a coping mechanism for unexpected changes in employment status).

Following the same line of thought, one may suggest that girls who grew up in larger families are more likely to adopt a conservative gender identity, in order to rationalise the fact that they are asked to contribute more to chores as the housework load increases. We have shown that these girls perform significantly more housework than their partners when they turn 34 and have worse labour market outcomes: conservative identities of women who grew up in large families, which can partly form as a childhood coping mechanism, have then the potential to develop into a set of constraining norms as in Collier (2016). This is in line with the literature showing that women in charge of most the housework load have limited opportunities for career and skills enhancements (Hirsch, 2005; Manning and Petrongolo, 2008; Russo and Hassink, 2008; Evertsson, 2013). Additionally, as argued by Mandel and Semyonov (2005) and Pettit and Hook (2005), conservative norms have the power to institutionalise economic inequality between women and men.

References

Achen, Alexandra, and Frank Stafford. 2005. *Data Quality of Housework Hours in the Panel Study of Income Dynamics: Who Really Does the Dishes?* Available at http://www.psidonline.isr.umich.edu/Publications/Papers/achenproxyreports04.pdf.

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and identity." *Quarterly Journal of Economics*, 115: 715–753.

Akerlof, George A., and Rachel E. Kranton. 2010. *Identity Economics.* Princeton University Press.

Altintas, Evrim, and Oriel Sullivan. 2016. "Fifty years of change updated: Cross-national gender convergence in housework." *Demographic Research*, 35: 455–469.

Anderton, Douglas, Noriko Tsuya, Lee Bean, and Geraldine Mineau. 1987. "Intergenerational transmission of relative fertility and life course patterns." *Demography*, 24: 467–480.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Angrist, Joshua D., and William Evans. 1998. "Children and their parents' labor supply: Evidence from exogenous variation in family size." *American Economic Review*, 88: 450–477.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of causal effects using instrumental variables." *Journal of the American Statistical Association*, 91: 444–455.

Angrist, Joshua D., Victor Lavy, and Analia Schlosser. 2010. "Multiple experiments for the causal link between the quantity and quality of children." *Journal of Labor Economics*, 28: 773–824.

Antill, John, Jacqueline Goodnow, Graeme Russell, and Sandra Cotton. 1996. "The influence of parents and family context on children's involvement in household tasks." *Sex Roles*, 34: 215–236.

Antman, Francisca. 2014. "Spousal employment and intra-household bargaining power." *Applied Economics Letters*, 21: 560–563.

Ashforth, Blake. 2000. *Role Transitions in Organizational Life: An Identity-based Perspective.* Routledge.

Åslund, Olof, and Hans Grönqvist. 2010. "Family size and child outcomes: Is there really no trade-off?" *Labour Economics*, 17: 130–139.

Bawa, Kapil, and Avijit Ghosh. 1999. "A model of household grocery shopping behavior." *Marketing Letters*, 10: 149–160.

Baxter, Janeen. 2005. "To marry or not to marry: Marital status and the household division of labor." *Journal of Family Issues*, 26: 300–321.

Baxter, Janeen, Belinda Hewitt, and Michele Haynes. 2008. "Life course transitions and housework: Marriage, parenthood, and time on housework." *Journal of Marriage and Family*, 70: 259–272.

Becker, Gary. 1960. "An economic analysis of fertility." In *Demographic and Economic Change in Developed Countries*. Ed. Universities-National Bureau, Chapter 7, 209–240. Columbia University Press.

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

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender identity and relative income within households." *Quarterly Journal of Economics*, 130: 571–614.

Bhalotra, Sonia, and Damian Clarke. 2019. "Twin birth and maternal condition." *Review of Economics and Statistics*, 101: 853–864.

Bianchi, Suzanne, and John Robinson. 1997. "What did you do today? Children's use of time, family composition, and the acquisition of social capital." *Journal of Marriage and Family*, 59: 332–344.

Bianchi, Suzanne, Liana Sayer, Melissa Milkie, and John Robinson. 2012. "Housework: Who did, does or will do it, and how much does it matter?" *Social Forces*, 91: 55–63.

Bisin, Alberto, and Thierry Verdier. 2001. "The economics of cultural transmission and the dynamics of preferences." *Journal of Economic Theory*, 97: 298–319.

Bittman, Michael, Paula England, Liana Sayer, Nancy Folbre, and George Matheson. 2003. "When does gender trump money? Bargaining and time in household work." *American Journal of Sociology*, 109: 186–214.

Björklund, Anders, and Kjell G. Salvanes. 2011. "Education and family background: Mechanisms and policies." In *Handbook of the Economics of Education*. Vol. 3, Eds. E.A. Hanushek, S. Machin and L. Woessmann, Chapter 3, 201–247. Elsevier.

Black, Sandra, Paul Devereux, and Kjell G. Salvanes. 2005. "The more the merrier? The effect of family size and birth order on children's education." *Quarterly Journal of Economics*, 120: 669–700.

Black, Sandra, Paul Devereux, and Kjell G. Salvanes. 2010. "Small family, smart family? Family size and the IQ scores of young men." *Journal of Human Resources*, 45: 33–58.

Blair, Sampson Lee. 1992*a*. "Children's participation in household labor: Child socialization versus the need for household labor." *Journal of Youth and Adolescence*, 21: 241–258.

Blair, Sampson Lee. 1992*b*. "The sex-typing of children's household labor: Parental influence on daughters' and sons' housework." *Youth and Society*, 24: 178–203.

Bloom, David, David Canning, Günther Fink, and Jocelyn Finlay. 2009. "Fertility, female labor force participation, and the demographic dividend." *Journal of Economic Growth*, 14: 79–101.

Blundell, Richard, Pierre-André Chiappori, and Costas Meghir. 2005. "Collective labor supply with children." *Journal of Political Economy*, 113: 1277–1306.

Booth, Alison, and Hiau Joo Kee. 2009. "Intergenerational transmission of fertility patterns." *Oxford Bulletin of Economics and Statistics*, 71: 183–208.

Brody, Charles, and Lala Carr Steelman. 1985. "Sibling structure and parental sex-typing of children's household tasks." *Journal of Marriage and Family*, 47: 265–273.

Cáceres-Delpiano, Julio. 2006. "The impacts of family size on investment in child quality." *Journal of Human Resources*, 41: 738–754.

Carlson, Daniel, and Jamie Lynch. 2013. "Housework: Cause and consequence of gender ideology?" *Social Science Research*, 42: 1505–1518.

Cherchye, Laurens, Bram De Rock, and Frederic Vermeulen. 2012. "Married with children: A collective labor supply model with detailed time use and intrahousehold expenditure information." *American Economic Review*, 102: 3377–3405.

Chiappori, Pierre-André. 1992. "Collective labor supply and welfare." *Journal of Political Economy*, 100: 437–467.

Chiappori, Pierre-André, Murat Iyigun, and Yoram Weiss. 2009. "Investment in schooling and the marriage market." *American Economic Review*, 99: 1689–1713.

Collier, Paul. 2016. "The cultural foundations of economic failure: A conceptual toolkit." *Journal of Economic Behavior and Organization*, 126: 5–24.

Conley, Dalton, and Rebecca Glauber. 2006. "Parental educational investment and children's academic risk estimates of the impact of sibship size and birth order from exogenous variation in fertility." *Journal of Human Resources*, 41: 722–737.

Conley, Timothy, Christian Hansen, and Peter Rossi. 2012. "Plausibly exogenous." *Review of Economics and Statistics*, 94: 260–272.

Cools, Sara, and Rannveig Kaldager Hart. 2017. "The effect of childhood family size on fertility in adulthood: New evidence from IV estimation." *Demography*, 54: 23–44.

Cordero-Coma, Julia, and Gøsta Esping-Andersen. 2018. "The intergenerational transmission of gender roles: Children's contribution to housework in Germany." *Journal of Marriage and Family*, 80: 1005–1019.

Cruces, Guillermo, and Sebastian Galiani. 2007. "Fertility and female labor supply in Latin America: New causal evidence." *Labour Economics*, 14: 565–573.

Evertsson, Marie. 2013. "The importance of work: Changing work commitment following the transition to motherhood." *Acta Sociologica*, 56: 139–153.

Falck, Oliver, Robert Gold, and Stephan Heblich. 2014. "E-lections: Voting Behavior and the Internet." *American Economic Review*, 104: 2238–65.

Farré, Lidia, and Francis Vella. 2013. "The intergenerational transmission of gender role attitudes and its implications for female labour force participation." *Economica*, 80: 219–247.

Fasang, Anette Eva, and Marcel Raab. 2014. "Beyond transmission: Intergenerational patterns of family formation among middle-class American families." *Demography*, 51: 1703–1728.

Fernández, Raquel, Alessandra Fogli, and Claudia Olivetti. 2004. "Mothers and sons: Preference formation and female labor force dynamics." *Quarterly Journal of Economics*, 119: 1249–1299.

Flèche, Sarah, Anthony Lepinteur, and Nattavudh Powdthavee. 2018. "Gender norms and relative working hours: Why do women suffer more than men from working longer hours than their partners?" *AEA Papers and Proceedings*, 108: 163–68.

Flèche, Sarah, Anthony Lepinteur, and Nattavudh Powdthavee. 2020. "Gender norms, fairness and relative working hours within households." *Labour Economics*, 65: 101866.

Gager, Constance, Teresa Cooney, and Kathleen Thiede Call. 1999. "The effects of family characteristics and time use on teenagers' household labor." *Journal of Marriage and Family*, 61: 982–994.

Geist, Claudia. 2010. "Men's and women's reports about housework." In *Dividing the Domestic: Men, Women, and Household Work in Cross-National Perspective.* Eds. J. Treas and S. Drobnič, Chapter 11, 217–240. Stanford University Press: Stanford, CA.

Gelbach, Jonah. 2016. "When do covariates matter? And which ones, and how much?" *Journal of Labor Economics*, 34: 509–543.

Giménez-Nadal, J. Ignacio, Lucia Mangiavacchi, and Luca Piccoli. 2019. "Keeping inequality at home: The genesis of gender roles in housework." *Labour Economics*, 58: 52–68.

Grunow, Daniela, Florian Schulz, and Hans-Peter Blossfeld. 2012. "What determines change in the division of housework over the course of marriage?" *International Sociology*, 27: 289–307.

Gupta, Sanjiv. 2007. "Autonomy, dependence, or display? The relationship between married women's earnings and housework." *Journal of Marriage and Family*, 69: 399–417.

Gupta, Sanjiv, and Michael Ash. 2008. "Whose money, whose time? A nonparametric approach to modeling time spent on housework in the United States." *Feminist Economics*, 14: 93–120.

Hirsch, Barry. 2005. "Why do part-time workers earn less? The role of worker and job skills." *Industrial and Labor Relations Review*, 58: 525–551.

Ibarra, Herminia. 2003. *Working Identity: Unconventional Strategies for Reinventing Your Career.* Harvard Business School Press.

International Labour Office. 2018. *World Employment and Social Outlook: Trends for Women 2018 – Global snapshot.* ILO, Geneva, Switzerland.

Johnston, David, Stefanie Schurer, and Michael Shields. 2013. "Maternal gender role attitudes, human capital investment, and labour supply of sons and daughters." *Oxford Economic Papers*, 66: 631–659.

Jones, Damon. 2015. *The Economics of Exclusion Restrictions in IV Models.* NBER Working Paper Series, 21391.

Kan, Man Yee. 2008. "Measuring housework participation: the gap between "stylised" questionnaire estimates and diary-based estimates." *Social Indicators Research*, 86: 381–400.

Kolk, Martin. 2014. "Understanding transmission of fertility across multiple generations–Socialization or socioeconomics?" *Research in Social Stratification and Mobility*, 35: 89–103.

Kowalski, Amanda E. 2019. *Counting defiers.* NBER Working Paper Series, 25671.

Lee, Jungmin. 2008. "Sibling size and investment in children's education: An Asian instrument." *Journal of Population Economics*, 21: 855–875.

Lyonette, Clare, and Rosemary Crompton. 2015. "Sharing the load? Partners' relative earnings and the division of domestic labour." *Work, Employment and Society*, 29: 23–40.

Mandel, Hadas, and Moshe Semyonov. 2005. "Family policies, wage structures, and gender gaps: Sources of earnings inequality in 20 countries." *American Sociological Review*, 70: 949–967.

Manning, Alan, and Barbara Petrongolo. 2008. "The part-time pay penalty for women in Britain." *Economic Journal*, 118: F28–F51.

Manning, Alan, and Barbara Petrongolo. 2017. "How local are labor markets? Evidence from a spatial job search model." *American Economic Review*, 107: 2877–2907.

Mas, Alexandre, and Amanda Pallais. 2017. "Valuing alternative work arrangements." *American Economic Review*, 107: 3722–3759.

McMunn, Anne, Lauren Bird, Elizabeth Webb, and Amanda Sacker. 2020. "Gender divisions of paid and unpaid work in contemporary UK couples." *Work, Employment and Society*, 34: 155–173.

Mostafa, Tarek, and Richard Wiggins. 2015. "The impact of attrition and non-response in birth cohort studies: a need to incorporate missingness strategies." *Longitudinal and Life Course Studies*, 6: 131–146.

Nicoletti, Cheti, Kjell G. Salvanes, and Emma Tominey. 2018. "The family peer effect on mothers' labor supply." *American Economic Journal: Applied Economics*, 10: 206–34.

Nybom, Martin. 2017. "The distribution of lifetime earnings returns to college." *Journal of Labor Economics*, 35: 903–952.

Peters, Jeanne, and Virginia Haldeman. 1987. "Time used for household work: A study of school-age children from single-parent, two-parent, one-earner, and two-earner families." *Journal of Family Issues*, 8: 212–225.

Pettit, Becky, and Jennifer Hook. 2005. "The structure of women's employment in comparative perspective." *Social Forces*, 84: 779–801.

Rosenzweig, Mark, and Kenneth Wolpin. 2000. "Natural "natural experiments" in economics." *Journal of Economic Literature*, 38: 827–874.

Russo, Giovanni, and Wolter Hassink. 2008. "The part-time wage gap: A career perspective." *De Economist*, 156: 145–174.

Sayer, Liana C. 2010. "Trends in housework." In *Dividing the Domestic: Men, Women, and Household Work in Cross-National Perspective*. Eds. J. Treas and S. Drobnič, Chapter 2, 19–38. Stanford University Press: Stanford, CA.

Schober, Pia. 2011. "The parenthood effect on gender inequality: Explaining the change in paid and domestic work when British couples become parents." *European Sociological Review*, 29: 74–85.

Schulz, Florian. 2021. *Mothers', Fathers' and Siblings' Housework Time Within Family Households.* Vol. 83.

Stratton, Leslie. 2015. *The Determinants of Housework Time.*

Sveningsson, Stefan, and Mats Alvesson. 2003. "Managing managerial identities: Organizational fragmentation, discourse and identity struggle." *Human Relations*, 56: 1163–1193.

Van Klaveren, Chris, Bernard van Praag, and Henriette Maassen van den Brink. 2008. "A public good version of the collective household model: an empirical approach with an application to British household data." *Review of Economics of the Household*, 6: 169–191.

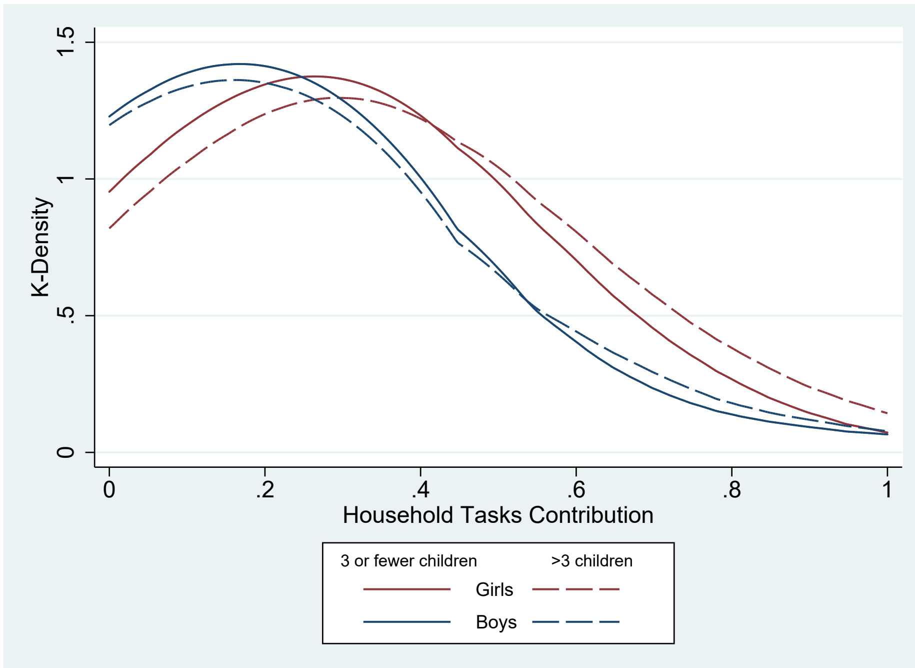
Vella, Francis. 1994. "Gender roles and human capital investment: The relationship between traditional attitudes and female labour market performance." *Economica*, 61: 191–211.

Wooldridge, Jeffrey. 2010. *Econometric Analysis of Cross Section and Panel Data.* MIT press.

World Economic Forum. 2018. *The Global Gender Gap Report.* World Economic Forum, Geneva.

Figures and Tables

Figure 2.1: Distribution of household tasks, by sex and family size at age 16



Note: These figures refer to our estimation sample. We use the Epanechnikov kernel function and bandwidth of 0.2.

Table 2.1: Family size and share of household tasks done 'Regularly' at age 16: OLS and 2SLS results

	All		Girls		Boys	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Family Size	0.014***	0.054**	0.018**	0.084***	0.008	0.014
	(0.005)	(0.023)	(0.007)	(0.029)	(0.008)	(0.036)
Female	0.088***	0.090***				
	(0.007)	(0.007)				
Observations	3389	3389	1935	1935	1454	1454
F-stat (first stage)		137		80		58

Notes: Robust standard errors in parentheses. 'Family size' indicates the number of siblings of the cohort member at age 16, plus one. All regressions control for the ethnicity of the cohort member, birth order dummies, a dummy indicating whether the cohort member's parents are still living in the same household, a dummy for the father having a high SES, years of education of the cohort member's parents, age of the parents at birth of the cohort member, an index measuring the attitude of the mother regarding maternal employment, a dummy indicating whether the gender composition of the siblings is balanced and regional dummies. Family size is instrumented by a dummy equal one if the first two children in the household are of the same sex. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Table 2.2: Family size and share of household tasks done 'Regularly' at age 16: heterogeneity - 2SLS results

	(1) Low SES	(2) High SES	(3) Conserv. Mothers	(4) Non-Conserv. Mothers
Panel A. Girls				
Family Size	0.116***	-0.017	0.116**	0.055
	(0.035)	(0.054)	(0.047)	(0.035)
Observations	1303	632	490	1445
F-stat (first stage)	61	17	38	48
Panel B. Boys				
Family Size	-0.004	0.043	0.105	-0.030
	(0.048)	(0.043)	(0.077)	(0.040)
Observations	925	529	402	1052
F-stat (first stage)	36	33	13	48

Notes: Robust standard errors in parentheses. 'Family size' indicates the number of siblings of the cohort member at age 16, plus one. All regressions control for the ethnicity of the cohort member, birth order dummies, a dummy indicating whether the cohort member's parents are still living in the same household, a dummy for the father having a high SES, years of education of the cohort member's parents, age of the parents at birth of the cohort member, an index measuring the attitude of the mother regarding maternal employment, a dummy indicating whether the gender composition of the siblings is balanced and regional dummies. Family size is instrumented by a dummy equal one if the first two children in the household are of the same sex. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Table 2.3: Family size at age 16 and household tasks at age 34 - 2SLS results

	Wife HH Tasks (1)	Husband HH Tasks (2)	Housework Gender Gap (3)
Panel A. Whole sample			
Family Size	0.052**	0.003	0.049*
	(0.025)	(0.012)	(0.029)
Female	0.125***	-0.045***	0.170***
	(0.009)	(0.004)	(0.011)
Observations	3200	3200	3200
F-stat (first stage)	140	140	140
Panel B. Women			
Family Size	0.070**	-0.001	0.071*
	(0.034)	(0.015)	(0.040)
Observations	1731	1731	1731
F-stat (first stage)	73	73	73
Panel C. Men			
Family Size	0.033	0.009	0.023
	(0.037)	(0.019)	(0.044)
Observations	1469	1469	1469
F-stat (first stage)	64	64	64

Notes: Robust standard errors in parentheses. 'Family size' indicates the number of siblings of the cohort member at age 16, plus one. All regressions control for the ethnicity of the cohort member, birth order dummies, a dummy indicating whether the cohort member's parents are still living in the same household, a dummy for the father having a high SES, years of education of the cohort member's parents, age of the parents at birth of the cohort member, an index measuring the attitude of the mother regarding maternal employment, a dummy indicating whether the gender composition of the siblings is balanced and regional dummies. Family size is instrumented by a dummy equal one if the first two children in the household are of the same sex. Statistical significance is coded following the standard notation: *** if the *p*-value is lower than 0.01, ** if the *p*-value is lower than 0.05, * if the *p*-value is lower than 0.01.

Table 2.4: FAMILY SIZE AT AGE 16 AND ADULT WOMEN'S OUTCOMES: 2SLS RESULTS

	Effect of Family Size for Women				
	All Women (1)	Low SES (2)	High SES (3)	Conserv. Mothers (4)	Non-Conserv. Mothers (5)
Housework gender gap	0.071*	0.086*	0.021	0.158	0.042
	(0.040)	(0.046)	(0.095)	(0.098)	(0.044)
Educational Attainment (age 34)					
Age left FT education	0.688	0.333	1.999	-0.713	1.096**
	(0.488)	(0.514)	(1.405)	(1.097)	(0.546)
At least A-level	0.070	0.085	0.028	-0.147	0.125*
	(0.062)	(0.070)	(0.148)	(0.144)	(0.069)
Labour Market Outcomes (age 34)					
Not employed	0.096	0.118*	-0.035	0.094	0.091
	(0.060)	(0.068)	(0.144)	(0.149)	(0.065)
Monthly wage (log)†	0.183	0.150	0.838	0.176	0.213
	(0.459)	(0.498)	(1.226)	(1.180)	(0.490)
Weekly working hours†	-3.145	-3.357	-1.007	0.443	-3.855
	(2.294)	(2.529)	(5.831)	(5.815)	(2.480)
Commuting time†	-0.053	-0.094*	0.130	0.088	-0.096*
	(0.048)	(0.048)	(0.150)	(0.131)	(0.052)
Employed partner	0.078***	0.082**	0.098**	0.047	0.079**
	(0.030)	(0.037)	(0.048)	(0.068)	(0.033)
Demographic characteristics (age 34)					
Married	0.110*	0.066	0.268*	-0.031	0.126*
	(0.061)	(0.067)	(0.152)	(0.148)	(0.067)
Having a least one child	0.017	0.047	-0.104	0.131	-0.012
	(0.055)	(0.058)	(0.163)	(0.140)	(0.060)
Number of children	0.023	0.086	-0.176	-0.123	0.054
	(0.152)	(0.156)	(0.479)	(0.336)	(0.172)

Notes: Robust standard errors in parentheses. The Table reports 2SLS estimates of the coefficient for 'Family size' (the number of siblings of the cohort member at age 16, plus one) for different dependent variables. All regressions control for the ethnicity of the cohort member, birth order dummies, a dummy indicating whether the cohort member's parents are still living in the same household, a dummy for the father having a high SES, years of education of the cohort member's parents, age of the parents at birth of the cohort member, an index measuring the attitude of the mother regarding maternal employment, a dummy indicating whether the gender composition of the siblings is balanced and regional dummies. Family size is instrumented by a dummy equal one if the first two children in the household are of the same sex. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.
†The regressions based on these outcomes are based on a subsample of employed cohort members. Results for these outcomes are similar when including also individuals who are not employed and conditioning on employment.

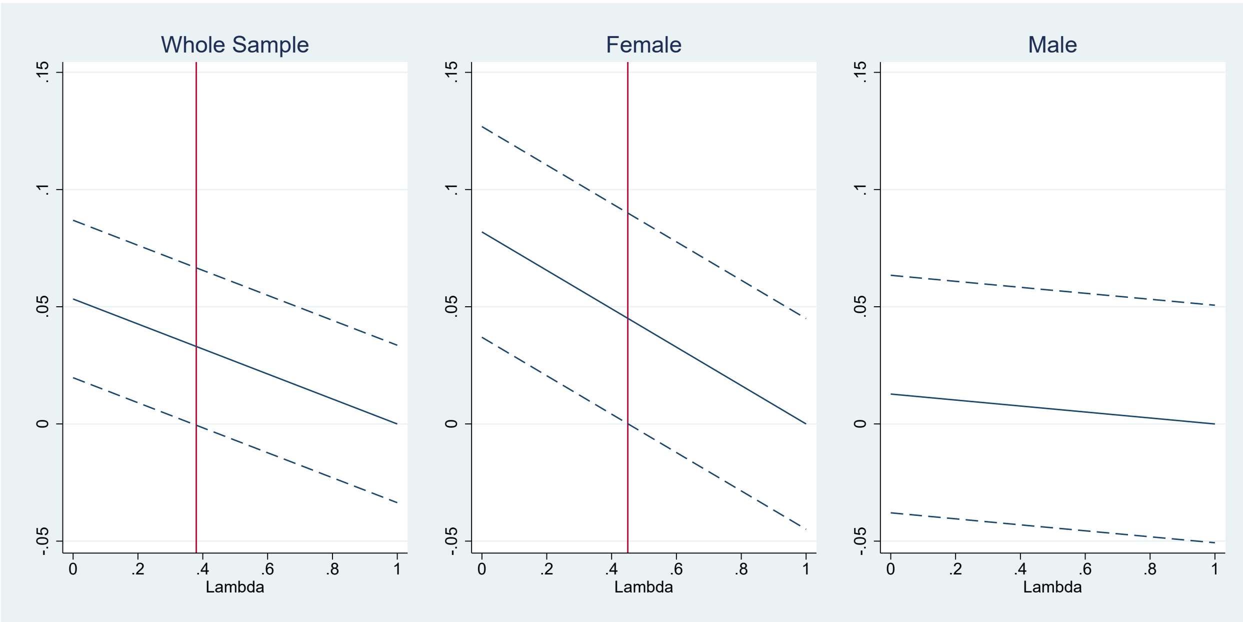
Table 2.5: Family size at age 16 and the channels: 2SLS results - Women

	Whole sample			Low SES			Conservative Mothers		
	Base	Full	Expl.	Base	Full	Expl.	Base	Full	Expl.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Family Size	0.071*	0.033	0.038***	0.086*	0.044	0.042***	0.158	0.148	0.010
	(0.040)	(0.038)	(0.012)	(0.046)	(0.043)	(0.014)	(0.098)	(0.092)	(0.023)
Contributions									
Not employed			0.011			0.014*			0.008
			(0.007)			(0.008)			(0.014)
Commuting time			0.005			0.010*			-0.003
			(0.005)			(0.005)			(0.005)
Employed partner			0.015**			0.016*			0.006
			(0.007)			(0.008)			(0.017)
Married			0.007			0.003			-0.002
			(0.004)			(0.004)			(0.009)
Observations	1731	1731	1731	1179	1179	1179	430	430	430
F-stat (first stage)	73	74		55	55		12	13	

Notes: Robust standard errors in parentheses. The Table reports 2SLS estimates of the coefficient for 'Family size' (the number of siblings of the cohort member at age 16, plus one) under different sample specifications. All regressions control for the ethnicity of the cohort member, birth order dummies, a dummy indicating whether the cohort member's parents are still living in the same household, a dummy for the father having a high SES, years of education of the cohort member's parents, age of the parents at birth of the cohort member, an index measuring the attitude of the mother regarding maternal employment, a dummy indicating whether the gender composition of the siblings is balanced and regional dummies. Family size is instrumented by a dummy equal one if the first two children in the household are of the same sex. Columns 3, 6, and 9 perform a decomposition analysis of the difference between the effect of family size in the baseline and full model specifications, following the approach of Gelbach (2016). Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Appendix 2.A: Additional figures and tables

Figure 2.A1: Relaxing Instrument Exogeneity



Note: The blue lines show the estimates of family size instrumented by the sex of the first two children as a function of the value of λ. The dashed lines indicate the 90% confidence intervals. The red lines show the minimal value of λ for which the family-size estimates are no longer significantly different from zero.

Table 2.A1: Descriptive statistics (childhood estimation sample)

	Mean	SD	Min	Max
Age 16:				
HH tasks	0.25	0.22	0	1
Homework	0.34	0.47	0	1
Leisure	0.41	0.23	0	1
Family Size	2.79	1.11	2	10
Same Sex	0.50		0	1
Balanced	0.65		0	1
Two Natural Parents in Household	0.72	0.45	0	1
Age 5:				
Mother's gender attitude	-0.08	1.02	-3	2
Age 0:				
Female	0.57		0	1
White	0.98		0	1
Birth Order	2.15	1.14	1	8
Age mother left school	15.80	2.02	0	27
Age father left school	16.13	2.59	0	33
Mother's age at cohort member's birth	26.55	5.04	18	46
Father's age at cohort member's birth	29.32	5.90	16	67
High SES	0.34		0	1
Observations	3389			

Table 2.A2: Sample composition at age 16: differences in means

	(1)	(2)	(1)-(2)	(1)	(3)	(1)-(3)
Family Size	2.785	2.964	-0.179***	2.785	2.555	0.231***
	[1.106]	[1.215]	(0.030)	[1.106]	[1.266]	(0.025)
Same Sex	0.497	0.518	-0.022	0.497	0.506	-0.010
	[0.500]	[0.500]	(0.013)	[0.500]	[0.500]	(0.011)
Balanced	0.654	0.663	-0.009	0.654	0.658	-0.004
	[0.476]	[0.473]	(0.012)	[0.476]	[0.474]	(0.010)
Two natural parents in household	0.835	0.777	0.058***	0.835	0.798	0.038***
	[0.371]	[0.416]	(0.011)	[0.371]	[0.402]	(0.009)
Mother's gender attitudes	-0.081	-0.016	-0.065	-0.081	-0.036	-0.044
	[1.019]	[1.017]	(0.026)	[1.019]	[1.026]	(0.021)
Female	0.571	0.404	0.167***	0.571	0.498	0.073***
	[0.495]	[0.491]	(0.013)	[0.495]	[0.500]	(0.010)
White	0.981	0.983	-0.001	0.981	0.983	-0.002
	[0.135]	[0.131]	(0.003)	[0.135]	[0.128]	(0.003)
Birth order	2.078	2.262	-0.185***	2.078	1.967	0.111***
	[0.95]	[0.981]	(0.025)	[0.950]	[0.984]	(0.020)
Age mother left education	15.802	15.475	0.328***	15.802	15.709	0.094
	[2.018]	[1.582]	(0.047)	[2.018]	[1.847]	(0.040)
Age father left education	16.127	15.669	0.458***	16.127	15.958	0.169***
	[2.586]	[1.971]	(0.060)	[2.586]	[2.348]	(0.050)
Mother's age at cohort member's birth	26.551	26.253	0.298	26.551	26.128	0.423***
	[5.045]	[5.230]	(0.132)	[5.045]	[5.133]	(0.106)
Father's age at cohort member's birth	29.331	28.915	0.416***	29.331	28.912	0.419***
	[6.029]	[6.200]	(0.161)	[6.029]	[6.229]	(0.131)
High SES	0.343	0.245	0.098***	0.343	0.240	0.103***
	[0.475]	[0.430]	(0.012)	[0.475]	[0.421]	(0.010)

Notes: The columns labeled '(1)' refer to the estimation sample at age 16. The column labeled '(2)' refers to the sample of cohort members living in households with at least two children but with missing information on household tasks. The column with label '(3)' instead refers to the overall BCS population with non-missing information on the covariates shown in the table. Columns '(1)-(2)' and '(1)-(3)' refer respectively to the differences in means between column (1) and column (2) and between column (1) and column (3). Standard deviations are in square brackets, while standard errors are reported in parentheses. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Table 2.A3: Descriptive statistics (adulthood estimation sample)

	Mean	SD	Min	Max
Age 34:				
Wife HH Tasks	0.47	0.26	0	1
Husband HH Tasks	0.15	0.12	0	1
Housework Gender Gap	0.32	0.32	-1	1
Married	0.73		0	1
Employed	0.84		0	1
A-level	0.44		0	1
Number of children	1.31	1.09	0	8
Age 16:				
Family Size	2.82	1.10	2	11
Same Sex	0.49		0	1
Balanced	0.66		0	1
Two natural parents in household	0.73	0.44	0	1
Age 5:				
Mother's gender attitude	-0.09	1.02	-3	2
Age 0:				
Female	0.54		0	1
White	0.99		0	1
Birth Order	2.18	1.16	1	8
Age mother left school	15.75	1.86	0	27
Age father left school	16.01	2.41	0	32
Mother's age at cohort member's birth	26.43	5.10	18	46
Father's age at cohort member's birth	29.08	5.87	16	63
High SES	0.32		0	1
Observations	3200			

Table 2.A4: Maternal attitudes at age 5 (only households where the same-sex instrument is equal one)

	Two children (1)	Three plus (2)	Difference (3)
Panel A. Whole sample			
Mother's attitudes toward:			
Maternal employment	-0.033	-0.131	0.098
	[1.010]	[1.003]	(0.063)
Sex equality	0.146	0.082	0.065
	[0.935]	[1.078]	(0.061)
Better life for women	-0.029	-0.080	0.050
	[0.950]	[1.087]	(0.062)
Anti-authoritarian child rearing	0.175	0.097	0.078
	[0.938]	[1.003]	(0.060)
Panel B. Girls			
Mother's attitudes toward:			
Maternal employment	-0.020	-0.133	0.113
	[0.976]	[0.960]	(0.082)
Sex equality	0.118	0.077	0.041
	[0.954]	[1.103]	(0.084)
Better life for women	-0.072	-0.035	-0.037
	[1.025]	[1.048]	(0.087)
Anti-authoritarian child rearing	0.142	0.098	0.044
	[0.942]	[1.043]	(0.082)
Panel C. Boys			
Mother's attitudes toward:			
Maternal employment	-0.050	-0.128	0.078
	[1.054]	[1.053]	(0.098)
Sex equality	0.184	0.087	0.097
	[0.908]	[1.051]	(0.089)
Better life for women	0.025	-0.131	0.157*
	[0.841]	[1.131]	(0.087)
Anti-authoritarian child rearing	0.218	0.096	0.122
	[0.933]	[0.958]	(0.087)

Notes: The table reports differences in means across family types. While we cannot distinguish compliers from infra-marginal individuals, we can expect compliers to populate column 2 (where the same-sex instrument is equal one and family size is above two). All variables are PCA-derived z-scores capturing maternal attitudes toward different topics (maternal employment; sex equality; needs for better life for women; anti-authoritarian child-rearing) at child age 5, computed by the data providers. High scores are to be interpreted as egalitarian/liberal views. Standard deviations are in square brackets, while standard errors are reported in parentheses. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Table 2.A5: First-stage regressions

	All	Females	Males
Child samples (age 16)			
Same sex	0.054***	0.054***	0.054***
	(0.005)	(0.006)	(0.007)
F-test	138	80	58
Observations	3389	1935	1454
Adult samples (age 34)			
Same sex	0.054***	0.054***	0.053***
	(0.005)	(0.006)	(0.007)
F-test	140	73	64
Observations	3200	1731	1469

Notes: The Table reports first-stage coefficients and Cragg-Donald Wald F-statistics for the same-sex instrument in the childhood estimation samples (age 16) and the adulthood estimation samples (age 34). Standard errors in parentheses. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Table 2.A6: INSTRUMENT AND CONTROLS IN CHILDHOOD: OLS RESULTS

	All (1)	Girls (2)	Boys (3)
A. Child characteristics			
Female	-0.025	-	-
	(0.026)		
White	-0.005	-0.018	-0.010
	(0.012)	(0.013)	(0.009)
B. Parent's characteristics			
Two natural parents in household	0.005	-0.011	0.029
	(0.018)	(0.023)	(0.027)
Mother's gender attitude	-0.080	-0.069	-0.101
	(0.053)	(0.070)	(0.079)
Age mother left school	0.137	0.194*	0.077
	(0.091)	(0.116)	(0.141)
Age father left school	0.036	-0.013	0.105
	(0.110)	(0.138)	(0.174)
Mother's age at cohort member's birth	-0.532***	-0.389*	-0.721***
	(0.160)	(0.219)	(0.242)
Father's age at cohort member's birth	-0.214	-0.094	-0.280
	(0.201)	(0.267)	(0.314)
High SES	-0.002	-0.025	0.028
	(0.014)	(0.019)	(0.022)

Notes: Robust standard errors in parentheses. Each cell shows the reduced-form effect of the instrument 'Same-sex' for each of our controls. All regressions control for the ethnicity of the cohort member, birth order dummies, the cognitive and non-cognitive skills of the cohort member at age 16, a dummy indicating whether the cohort member's parents are still living in the same household, a dummy for the father having a high SES, years of education of the cohort member's parents, age of the parents at birth of the cohort member, an index measuring the attitude of the mother regarding maternal employment, a dummy indicating whether the gender composition of the siblings is balanced and regional dummies. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01. The lowest robust F-statistics is 24.4 and belongs to the 2SLS baseline regression with at least 10 reported tasks, for the boys subsample.

Table 2.A7: A TOPOLOGY OF SIBLING COMPOSITIONS

Birth order:	First or second		Third +	
Sex of cohort member:	Boy	Girl	Boy	Girl
Instrument:				
$Z = 0$	a	c	a_1	c_1
$Z = 1$ (Two girls)	-	d	b_1	d_1
$Z = 1$ (Two boys)	b	-	b_2	d_2

Table 2.A8: Family size and probability to be partnered at age 34 - 2SLS results

	Probability to be partnered
Family Size	0.027
	(0.031)
Female	-0.024**
	(0.012)
Observations	4227
F-stat (first stage)	183

Notes: Robust standard errors in parentheses. ‘Family size’ indicates the number of siblings of the cohort member at age 16, plus one. All regressions control for the ethnicity of the cohort member, birth order dummies, a dummy indicating whether the cohort member’s parents are still living in the same household, a dummy for the father having a high SES, years of education of the cohort member’s parents, age of the parents at birth of the cohort member, an index measuring the attitude of the mother regarding maternal employment, a dummy indicating whether the gender composition of the siblings is balanced and regional dummies. Family size is instrumented by a dummy equal one if the first two children in the household are of the same sex. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Table 2.A9: Family size at age 16 and adult men's outcomes: 2SLS results

	Effect of Family Size for Men				
	Whole Sample	Low SES	High SES	Conserv. Mothers	Non-Conserv. Mothers
	(1)	(2)	(3)	(4)	(5)
Housework gender gap	0.023	-0.010	0.065	-0.046	0.059
	(0.044)	(0.060)	(0.059)	(0.079)	(0.054)
Educational Attainment (age 34)					
Age left FT education	0.158	0.028	0.745	-0.822	0.534
	(0.513)	(0.676)	(0.784)	(0.870)	(0.617)
At least A-level	-0.026	-0.053	0.058	-0.096	-0.009
	(0.063)	(0.084)	(0.096)	(0.105)	(0.078)
Labour Market Outcomes (age 34)					
Not employed	0.021	0.021	0.038	-0.047	0.052
	(0.039)	(0.054)	(0.049)	(0.050)	(0.053)
Monthly wage (log)†	-0.031	-0.029	0.033	0.099	-0.046
	(0.105)	(0.114)	(0.231)	(0.263)	(0.140)
Weekly working hours†	1.068	-0.227	1.843	4.799	-0.553
	(2.355)	(3.274)	(2.816)	(3.094)	(3.168)
Commuting time†	-0.083	-0.064	-0.147	-0.266**	0.011
	(0.071)	(0.094)	(0.104)	(0.117)	(0.089)
Employed partner	-0.091	-0.090	-0.066	-0.006	-0.162*
	(0.068)	(0.093)	(0.088)	(0.099)	(0.090)
Demographic characteristics (age 34)					
Married	0.118*	0.176*	0.007	0.145	0.122
	(0.068)	(0.097)	(0.091)	(0.104)	(0.088)
Having a least one child	0.040	0.015	0.061	0.002	0.082
	(0.070)	(0.095)	(0.098)	(0.106)	(0.089)
Number of children	0.219	0.324	0.035	0.192	0.285
	(0.154)	(0.217)	(0.196)	(0.239)	(0.194)

Notes: Robust standard errors in parentheses. The Table reports 2SLS estimates of the coefficient for 'Family size' (the number of siblings of the cohort member at age 16) for different dependent variables. All regressions control for the ethnicity of the cohort member, birth order dummies, a dummy indicating whether the cohort member's parents are still living in the same household, a dummy for the father having a high SES, years of education of the cohort member's parents, age of the parents at birth of the cohort member, an index measuring the attitude of the mother regarding maternal employment, a dummy indicating whether the gender composition of the siblings is balanced and regional dummies. Family size is instrumented by a dummy equal one if the first two children in the household are of the same sex. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

†The regressions based on these outcomes are based on a subsample of employed cohort members. Results for these outcomes are similar when including also individuals who are not employed and conditioning on employment.

Table 2.A10: Family size at age 16 and household gender gaps at age 42

	Low SES	High SES	Conserv. Mothers	Non-Conserv. Mothers
	(1)	(2)	(3)	(4)
Panel A. Women				
Family Size	0.090**	0.082	0.223*	0.049
	(0.046)	(0.106)	(0.129)	(0.043)
Observations	1179	552	430	1301
F-stat (first stage)	55	17	12	61
Panel B. Men				
Family Size	-0.056	0.043	-0.014	-0.016
	(0.049)	(0.052)	(0.061)	(0.046)
Observations	996	473	391	1078
F-stat (first stage)	34	39	24	44

Notes: Robust standard errors in parentheses. 'Family size' indicates the number of siblings of the cohort member at age 16, plus one. All regressions control for the ethnicity of the cohort member, birth order dummies, the cognitive and non-cognitive skills of the cohort member at age 16, a dummy indicating whether the cohort member's parents are still living in the same household, a dummy for the father having a high SES, years of education of the cohort member's parents, age of the parents at birth of the cohort member, an index measuring the attitude of the mother regarding maternal employment, a dummy indicating whether the gender composition of the siblings is balanced and regional dummies. Family size is instrumented by a dummy equal one if the first two children in the household are of the same sex. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Table 2.A11: Family size at age 16 and the channels: 2SLS results - Women

	High SES			Non-Conservative Mothers		
	Base (1)	Full (2)	Expl. (3)	Base (4)	Full (5)	Expl. (6)
Family Size	0.021	-0.008	0.029	0.042	-0.004	0.046***
	(0.095)	(0.094)	(0.035)	(0.044)	(0.042)	(0.015)
Contributions						
Not employed			-0.004			0.010
			(0.015)			(0.008)
Commuting time			-0.013			0.012*
			(0.015)			(0.007)
Employed partner			0.017			0.015**
			(0.014)			(0.007)
Married			0.029			0.008
			(0.019)			(0.005)
Observations	552	552	552	1301	1301	1301
F-stat (first stage)	17	17		61	60	

Notes: Robust standard errors in parentheses. The Table reports 2SLS estimates of the coefficient for 'Family size' (the number of siblings of the cohort member at age 16) under different sample specifications. All regressions control for the ethnicity of the cohort member, birth order dummies, a dummy indicating whether the cohort member's parents are still living in the same household, a dummy for the father having a high SES, years of education of the cohort member's parents, age of the parents at birth of the cohort member, an index measuring the attitude of the mother regarding maternal employment, a dummy indicating whether the gender composition of the siblings is balanced and regional dummies. Family size is instrumented by a dummy equal one if the first two children in the household are of the same sex. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Table 2.A12: Family size at age 16 and the channels: 2SLS results - Men

	Whole sample			Low SES			Conservative Mothers		
	Base	Full	Expl.	Base	Full	Expl.	Base	Full	Expl.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Family Size	0.023	0.003	0.020	-0.010	-0.034	0.024	-0.046	-0.069	0.023
	(0.044)	(0.043)	(0.013)	(0.060)	(0.059)	(0.017)	(0.079)	(0.077)	(0.028)
Contributions									
Not employed			-0.002			-0.002			0.004
			(0.004)			(0.006)			(0.005)
Commuting time			-0.004			-0.004			-0.017
			(0.003)			(0.005)			(0.011)
Employed partner			0.017			0.019			0.019
			(0.011)			(0.014)			(0.022)
Married			0.008			0.010			0.018
			(0.005)			(0.007)			(0.013)
Observations	1469	1469	1469	996	996	996	391	391	391
F-stat (first stage)	64	64		34	34		24	22	

Notes: Robust standard errors in parentheses. The Table reports 2SLS estimates of the coefficient for 'Family size' (the number of siblings of the cohort member at age 16) under different sample specifications. All regressions control for the ethnicity of the cohort member, birth order dummies, a dummy indicating whether the cohort member's parents are still living in the same household, a dummy for the father having a high SES, years of education of the cohort member's parents, age of the parents at birth of the cohort member, an index measuring the attitude of the mother regarding maternal employment, a dummy indicating whether the gender composition of the siblings is balanced and regional dummies. Family size is instrumented by a dummy equal one if the first two children in the household are of the same sex. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Table 2.A13: Family size at age 16 and the channels: 2SLS results - Men

	High SES			Non-Conservative Mothers		
	Base	Full	Expl.	Base	Full	Expl.
	(1)	(2)	(3)	(4)	(5)	(6)
Family Size	0.065	0.056	0.010	0.059	0.040	0.019
	(0.059)	(0.056)	(0.020)	(0.054)	(0.053)	(0.016)
Contributions						
Not employed			-0.001			-0.005
			(0.003)			(0.005)
Commuting time			-0.002			0.000
			(0.004)			(0.003)
Employed partner			0.012			0.017
			(0.015)			(0.014)
Married			0.001			0.007
			(0.009)			(0.006)
Observations	473	473	473	1078	1078	1078
F-stat (first stage)	39	39		44	42	

Notes: Robust standard errors in parentheses. The Table reports 2SLS estimates of the coefficient for 'Family size' (the number of siblings of the cohort member at age 16) under different sample specifications. All regressions control for the ethnicity of the cohort member, birth order dummies, a dummy indicating whether the cohort member's parents are still living in the same household, a dummy for the father having a high SES, years of education of the cohort member's parents, age of the parents at birth of the cohort member, an index measuring the attitude of the mother regarding maternal employment, a dummy indicating whether the gender composition of the siblings is balanced and regional dummies. Family size is instrumented by a dummy equal one if the first two children in the household are of the same sex. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Appendix 2.B: Robustness checks

2.B1 Sensitivity to the definition of the share of housework

We measure the contribution to housework at age 16 using the share of household tasks the cohort member helps with ‘Regularly’. Cohort members report their contribution to twelve different household tasks. Due to missing information and survey filters, the average number of reported tasks in our estimation sample is 9.3 and the median is 10. It can be argued that the number of reported tasks partially drives our estimates. We address this concern in two different ways: we first add the number of reported tasks as an additional control variable, and then re-estimate our main regressions using only cohort members reporting at least ten tasks out of twelve. The results, compared to those from our baseline estimation, appear in the first three rows of Table 2.B1. The first row shows our baseline estimates of family size for the whole sample and then for girls and boys separately. In the second row, controlling for the number of reported tasks makes no difference. We then show the estimated 2SLS coefficients of individuals reporting at least ten tasks in the third row of Table 2.B1. Here, the effect of family size for girls remains unchanged, that for the whole sample is somewhat smaller, and that for boys is negative but not statistically different from zero.

Rather than using tasks that are performed ‘Regularly’ as opposed to ‘Sometimes’ or ‘Rarely or never’, we can also look at the intermediate category ‘Sometimes’. To do so, we assign a score of 1 to tasks performed ‘Regularly’, a score of 0.5 to those performed ‘Sometimes’ and a score of 0 otherwise. As for our original dependent variable, we calculate the share as the average score across the reported tasks. Using this new dependent variable does not affect our conclusions: as revealed by the last line of the Part A of Table 2.B1, an additional family member still increases the whole sample contribution to household tasks and, once again, the result is mostly driven by girls.

2.B2 Different definitions of tasks

Our main measure of household contribution uses a set of twelve tasks that have different features. As revealed in Table 2.B2, most tasks are gender-specific. We consider a task to be 'feminine' ('masculine') if the share of girls (boys) reporting doing the task 'Regularly' is statistically larger than the share of boys (girls) at the 5% level. Girls spend significantly more time shopping, washing up, cleaning, making the bed, cooking, looking after pets, washing and ironing, and looking after younger siblings, while boys spend more time gardening, cleaning the car, and in DIY activities. The share of girls looking after older people 'regularly' is slightly larger than the share of boys, but the difference is not statistically significant at the 5% level. In the first two rows of Part B of Table 2.B1 we check whether the effect of family size affects the contribution of cohort members to 'feminine' and 'masculine' tasks differently. We find that, as family size rises, girls perform a significantly larger share of both 'feminine' and 'masculine' tasks (although their contribution to the former is larger than to the latter), while boys do not spend more time in any type of tasks. Note that this partition of housework into 'feminine' and 'masculine' almost perfectly overlaps with the intrinsic periodicity of the tasks (e.g. cooking and making the bed are daily activities, while a car needs to be cleaned less frequently), so that our results can also be interpreted in terms of frequency.

In addition, some of the tasks require the presence of particular items or person in the household. This is the case, for instance, for tasks involving care-giving or those such as cleaning the car and tending to the garden. We cannot of course assume that all households in our sample satisfy the pre-conditions for these kind of tasks to be performed. We then exclude these in row three of Part B of Table 2.B1, where we construct the share of household tasks carried out by cohort members based only on 'unconditional' tasks, i.e. tasks that can be carried out in any household. We find that the effect of family size is even stronger when using this measure of housework contribution.

By pooling together the twelve types of household tasks to create one single measure, we also implicitly assume that all tasks increase equally in family size. This is not unrealistic for some of our tasks, such as shopping, washing up, cleaning, making the bed, cooking, washing and ironing, and looking after youngsters (Bawa and Ghosh, 1999). It is however more difficult to believe that looking after the elderly and pets, gardening, cleaning the car, and painting or decorating are tasks that are more likely to be regularly performed in families with more children. We then expect our main estimates to be driven by the first set of tasks, while the second set can be seen more as a placebo test. This is confirmed in the fourth and fifth rows in Part B of Table 2.B1. A girl who grew up in a large family contributes significantly more to those tasks for which demand likely rises in family size.

However, there is no significant effect for girls when considering their contribution to the second set of tasks that we expect to be less sensitive to family size. We continue to find no effect for boys regarding either kind of tasks. The last row of Table 2.B1 excludes care-giving activities and only considers the contribution to household tasks that do not involve social interactions. Again, we find results that are in line with our baseline estimates: a larger family increases girls' contribution to housework but not that of boys.

2.B3 Other measures of time: homework and leisure

Time is a finite resource. As family size rises and girls contribute more to household tasks, we expect a reduction in the time they spend on homework and leisure. The overall time allocation of boys should instead remain unchanged. We check this in the BCS by looking at time spent on a variety of activities. We measure time spent on homework from the following question: 'How much time did you spend doing homework yesterday?'. The respondents were asked to use different time categories. Since more than two-thirds of our estimation sample reported doing no homework, we create a dummy for the cohort member having done at least some homework. Cohort members were also asked to report whether they read at least one book during the four weeks before the interview and if they were members of a sports club, a religious organisation, or any other youth organisation over the last 12 months. We construct an index of leisure activities as the share of activities a cohort member engaged in.

We re-estimate our main model using our measures of homework and leisure as the dependent variables. Table 2.B3 shows the results for the whole sample and by gender. Consistent with our main results, girls spend relatively less time doing homework and are less likely to engage into leisure activities in larger families. As expected, there is no effect for boys.[17] As in Table 2.2, the effect of family size is stronger for girls who grew up in low-SES families and with conservative mothers (these results are available upon request).

[17]The large difference in sample size between the first three columns of Table 2.B3 and Table 2.1 reflects that 'time spent in homework yesterday' and 'contribution to housework' were measured using different questionnaires. According to the data provider's documentation, the response rate of the former was much lower than that of the latter. Our measure of 'participation in activities' and 'contribution to housework' were measured using the same questionnaire, and the difference of approximately 100 observations here is due to missing information.

Table 2.B1: Family size and share of housework at age 16: Sensitivity analysis

	All (1)	Girls (2)	Boys (3)
A. Sensitivity of the share of housework			
Baseline	0.054**	0.084***	0.014
	(0.023)	(0.029)	(0.036)
Baseline (controlling for no. tasks)	0.061***	0.090***	0.022
	(0.022)	(0.029)	(0.034)
Baseline (at least 10 reported tasks)	0.020	0.082**	-0.043
	(0.022)	(0.032)	(0.029)
Share of housework, counting approach	0.028	0.040*	0.008
	(0.019)	(0.023)	(0.030)
B. Different definitions of tasks			
Feminine	0.064**	0.106***	0.011
	(0.026)	(0.036)	(0.039)
Masculine	0.034	0.051*	0.011
	(0.025)	(0.027)	(0.047)
Unconditional	0.063**	0.104***	0.011
	(0.026)	(0.036)	(0.037)
Increasing in family size	0.063**	0.108***	0.008
	(0.027)	(0.037)	(0.038)
Insensitive to family size	0.027	0.033	0.009
	(0.027)	(0.033)	(0.046)
Excluding care-giving	0.051**	0.091***	-0.001
	(0.024)	(0.032)	(0.035)

Notes: Robust standard errors in parentheses. The Table reports 2SLS estimates of the coefficient for 'Family size' (the number of siblings of the cohort member at age 16) under different definitions of the dependent variable. All regressions control for the ethnicity of the cohort member, birth order dummies, a dummy indicating whether the cohort member's parents are still living in the same household, a dummy for the father having a high SES, years of education of the cohort member's parents, age of the parents at birth of the cohort member, an index measuring the attitude of the mother regarding maternal employment, a dummy indicating whether the gender composition of the siblings is balanced and regional dummies. Family size is instrumented by a dummy equal one if the first two children in the household are of the same sex. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01. The lowest robust F-statistics is 24.4 and belongs to the 2SLS baseline regression with at least 10 reported tasks, for the boys sub-sample.

Table 2.B2: Distribution of household tasks at age 16

	Girls	Boys	Difference
1. Shopping	0.293	0.168	0.125***
	[0.455]	[0.374]	(0.016)
2. Washing up	0.503	0.332	0.171***
	[0.500]	[0.471]	(0.018)
3. Cleaning	0.302	0.133	0.170***
	[0.459]	[0.339]	(0.015)
4. Making the bed	0.493	0.319	0.174***
	[0.500]	[0.466]	(0.018)
5. Cooking	0.259	0.099	0.160***
	[0.438]	[0.299]	(0.014)
6. Looking after elders	0.084	0.060	0.025*
	[0.278]	[0.237]	(0.013)
7. Looking after pets	0.510	0.438	0.072***
	[0.500]	[0.496]	(0.021)
8. Washing and/or ironing	0.320	0.061	0.259***
	[0.467]	[0.239]	(0.015)
9. Gardening	0.035	0.147	-0.112***
	[0.184]	[0.354]	(0.011)
10. Cleaning car	0.042	0.149	-0.107***
	[0.200]	[0.356]	(0.011)
11. Painting or decorating	0.039	0.107	-0.067***
	[0.195]	[0.309]	(0.010)
12. Looking after youngsters	0.352	0.190	0.162***
	[0.478]	[0.393]	(0.024)
Observations	1935	1454	

Notes: Each household task is reduced to a dummy equal one if the task is performed regularly, zero otherwise. Standard deviations are in square brackets, while standard errors are reported in parentheses. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Table 2.B3: Family size and time use at age 16: 2SLS results

	Did homework yesterday			Participation to activities		
	All	Girls	Boys	All	Girls	Boys
	(1)	(2)	(3)	(4)	(5)	(6)
Family Size	-0.049	-0.131*	0.018	-0.017	-0.058*	0.038
	(0.053)	(0.078)	(0.072)	(0.023)	(0.030)	(0.036)
Female	0.035*			-0.001		
	(0.020)			(0.008)		
Observations	2177	1202	975	3263	1873	1390
F-stat (first stage)	103	57	49	138	81	58

Notes: Robust standard errors in parentheses. 'Family size' indicates the number of siblings of the cohort member at age 16, plus one. All regressions control for the ethnicity of the cohort member, birth order dummies, a dummy indicating whether the cohort member's parents are still living in the same household, a dummy for the father having a high SES, years of education of the cohort member's parents, age of the parents at birth of the cohort member, an index measuring the attitude of the mother regarding maternal employment, a dummy indicating whether the gender composition of the siblings is balanced and regional dummies. Family size is instrumented by a dummy equal one if the first two children in the household are of the same sex. Statistical significance is coded following the standard notation: *** if the p-value is lower than 0.01, ** if the p-value is lower than 0.05, * if the p-value is lower than 0.01.

Chapter 3

Families at a Loss: The Relationship Between Income Changes and Child Human Capital

Families at a Loss: The Relationship Between Income Changes and Child Human Capital

3.1 Introduction

The analysis of parental income in relation to human capital is no new subject in economics. A considerable number of theoretical contributions describe how income enables parents to put in place investments that will foster their kids' human capital development, which in turn will shape their later life outcomes. Extensive work from Heckman and coauthors from the early 2000s has emphasised that human capital is a multidimensional concept that cannot be equated to cognitive skills only (see, among others, Heckman and Rubinstein, 2001; Heckman, Stixrud and Urzua, 2006; Cunha, Heckman and Schennach, 2010). The existing empirical literature, however, has largely focused on how parents' socio-economic status affects children's cognitive abilities, often neglecting non-cognitive ones, despite the growing body of evidence proving their importance in determining later life outcomes (see Cunha, Heckman and Schennach, 2010; Blanden and Machin, 2010; Ermisch, 2008; Flouri, Mavroveli and Tzavidis, 2012; Flouri, Midouhas and Joshi, 2014). Further evidence from neurobiology, developmental psychology, and economics underlines the fundamental role of early age experiences and environment in shaping brain functions and future development. While there is an increasing consensus on the importance of non-cognitive skills, the evidence surrounding its determinants and, in particular, on the impact of economic shocks on the social and behavioural outcomes of children is still scarce.

Transitory economic conditions, such as income shocks, are likely to affect child human capital development. While there is an increasingly large body of evidence on the relationship between parental socio-economic status or permanent income on child human capital, relatively little is known on the role of transitory income changes. The causal evidence on the expansion of tax credit policies and child benefits shows positive effects on a range of child cognitive outcomes (Dahl and Lochner, 2012; Evans and Garthwaite, 2014; Milligan and Stabile, 2011), while evidence on non-cognitive outcomes is scarce. However, positive economic shocks might differ in nature from negative ones: insights from prospect theory (Kahneman and Tversky, 1979) suggest that individuals overweight utility losses over gains. Similarly, the realisation of income losses, together with their potential interaction with market conditions (e.g. credit constraints), may well have an asymmetric effect on the parents' ability to foster their children's human capital, either via the provision of material inputs or via the quality and quantity of their time inputs.

The main scope of this paper is to assess the relationship and pathways that link gains and losses in family income to the cognitive and non-cognitive development of children. I do so using a longitudinal dataset from the UK, the Millennium Cohort Study (MCS), which follows the lives of around 19,000 children born at the turn of the millennium and their families. An almost unique feature of the dataset is that it contains measures of both cognitive and non-cognitive development of children aged 3 to 15. The relationship between parental income and child human capital in MCS has already been the object of attention of some papers: Kelly *et al.* (2011), using cognitive and non-cognitive measures of child development from waves 2 and 3 of MCS, find evidence of an income gradient, consistently with the previous literature. Noonan, Burns and Violato (2018) links family income to health and non-cognitive outcomes of children, finding that permanent income has a protective effect against the probability of experiencing behavioural problems at age 11. Other papers use the MCS to document a gradient between parental economic background and children's cognitive (Dearden, Sibieta and Sylva, 2011) and non-cognitive (Tamura, Morrison and Pikhart, 2020) development.

I here use information from the six available waves of MCS to investigate the relationship between cognitive and non-cognitive skills formation and family income changes. The outcomes of interest, namely cognitive and non-cognitive skills, are respectively measured through age-adjusted reading test scores and the Strengths and Difficulties Questionnaire (SDQ), a widely recognised behavioural screening tool for children and adolescents (Goodman, Lamping and Ploubidis, 2010). Using a value added model to assess the impact of income gains and losses on child human capital, I find that income losses are correlated with lower residualised measures of cognitive and non-cognitive skills, while gains only predict better cognitive performance. Consistent with the literature, results suggest that about one third of the effect of losses on non-cognitive outcomes transits via maternal measures of well-being. Similar to Bruckauf and Chzhen (2016), I then explore mobility in and out of the bottom of the reading test-scores and SDQ distributions. I find that income losses (gains) are positively (negatively) correlated with the probability of entering the bottom quintile of the distribution of all outcomes, and that the bottom of the distribution is stickier for non-cognitive outcomes rather than cognitive ones. This paper contributes to the literature in at least three ways: it is the first study to use data on measures of both cognitive and non-cognitive development for all currently available waves of the MCS in relationship to movements across the income distribution; it uses a value added model approach to assess the contribution of income changes on the year-to-year formation of cognitive and non-cognitive skills; lastly, it relaxes the assumption underlying most of the empirical literature in this field, which is that income gains and losses have a symmetric effect on children's outcomes.

Here follows an outline of the remainder of the paper. Section 3.2 reviews part of the relevant literature in the field. Section 3.3 describes the dataset and the main variables of interest, and presents the empirical strategy. Section 3.4 describes the main results. Robustness checks are conducted in Section 3.5, to test for the sensitivity of the estimates to changes in the specification and measurement issues. Before concluding with Section 3.7, Section 3.6 shows some additional results addressing persistence and transition dynamics.

3.2 Literature review

There is a large literature addressing the relationship between family income and child human capital (see Dahl and Lochner, 2012, for a review). Part of this literature addresses the causal impact of income, exploiting the exogenous variations coming from policy changes (e.g. income transfer programs). With US data, Dahl and Lochner (2012) exploit discontinuities in the Earned Income Tax Credit to identify the effect of income on test scores, finding that a 1,000 dollars increase in family income raises combined math and reading test scores by 6% of a standard deviation. Using the same policy discontinuities, Evans and Garthwaite (2014) find that higher income causes lower levels of both self-reported maternal stress and biological markers associated with stress. Milligan and Stabile (2011) look at variations in income induced by child benefit policy expansion in Canada and find significant positive effects on child and mother's mental health. Blau (1999) performs a fixed effect analysis of the NLSY cohort, finding little to no effect of current income on cognitive, social, and emotional development of kids; however, he does not control for potentially endogenous transitory shocks. Dahl and Lochner (2012) improve Blau's identification strategy with an instrumental variables approach, finding larger effects. Kuehnle (2014) explores the link between income and self-reported health on the 1970 British Cohort Study. Using local unemployment rates as an instrument, he identifies a small positive causal effect of family income on children's health.

Other studies adopt a descriptive approach to document a positive association between family income and child human capital, the effect being mostly larger for cognitive rather than non-cognitive outcomes (Duncan and Brooks-Gunn, 1997). While some focus on the net effect of family income on human capital accumulation (Shea, 2000), other studies explore the channels mediating this relationship (Washbrook, Gregg and Propper, 2014; Yeung, Linver and Brooks-Gunn, 2002). Income, for instance, is known to be a determinant of individual well-being, with several studies establishing a causal link between the two (Frijters, Haisken-DeNew and Shields, 2004; Gardner and Oswald, 2007; Powdthavee, 2010). Parental well-being, in turn, determines parenting practices: higher well-being is associated with warmer and responsive parenting (McLoyd *et al.*, 1994; Sampson and Laub, 1994; Smith and Brooks-Gunn, 1997), with positive spillovers on children's development (Conger *et al.*, 1992; McLoyd, 1990). Looking at the correlation between a permanent and a transitory measure of income on preschool children's outcomes, Yeung, Linver and Brooks-Gunn (2002) test for the presence of two main set of mediating channels, respectively linked to the 'family stress' theory and the 'investment' theory. They find that mothers' emotional affect and parenting style play a significant role in explaining the effect of income on preschool children's externalising behaviour; on the other hand, the effect of income on children's cognitive skills runs mostly through the setting up of material investments. Despite the important role of mediating factors, the authors find that a direct effect of income on cognitive skills and externalising behaviour still persists. Washbrook, Gregg and Propper (2014) find consistent results on the mediating role of parents, using a broader set of measures of maternal psychosocial functioning. Frank and Meara (2009) find that maternal depression has a large negative effect on child development and the accumulation of non-cognitive skills, while it does not seem to affect math and reading test scores. However, these papers, using only cross sectional variations in income, fail to capture the dynamics between income changes, the short-term reaction of parents in terms of well-being, and children's behavioural and cognitive response. In this sense, a paper that comes closer to this objective is Clark, D'Ambrosio and Barazzetta (2021), who use the same cross-sectional approach to estimate the effect of mothers' financial problems (a variable capturing financial distress rather than plain income) on a variety of childhood outcomes and find that only one quarter of the effect is captured by mothers' mental health.

In this paper, I use a value added model to address the relationship between changes in income and the accumulation of child human capital over time. Value added models are an established tool in the field of economics of education and are typically used to assess the impact of teachers on kids' performance in school. In general, they can be used to evaluate the contribution of an input in the accumulation of human capital from a given point in time to a subsequent one (Todd and Wolpin, 2003, 2007; Koedel, Mihaly and Rockoff, 2015). With respect to regressions to the mean, value added models offer the advantage of assessing the average year-to-year contribution of factors of interest to the trajectories of fairly persistent outcomes. In a way, they provide a life-event approach to the short-term evolution of human capital that, under certain assumptions, allows to control for latent factors contributing to the human capital production function. Although widely used in relationship to teachers' quality, there are only few examples of their application to different contexts. With the same dataset used in this paper, Del Bono *et al.* (2016) adopt a cumulative value-added model to show the importance of early childhood maternal time investments on child cognitive skills. Other papers use value-added models to address, for example, the effect of private schools on learning achievements (Andrabi *et al.*, 2011), the role of obesity in child non-cognitive development (Black and Kassenboehmer, 2017), the persistence of mental health issues (Roy and Schurer, 2013), or the relationship between income changes and changes in life satisfaction Boyce *et al.* (2013).

3.3 Data and Methodology

3.3.1 Data description

This paper uses data from five waves of the UK Millennium Cohort Study (MCS). MCS is a longitudinal birth cohort study following the lives of around $19,000$ children born in the UK between 2000 and 2001. Six waves of the survey have been conducted so far, at age 9 months, 3 years, 5 years, 7 years, 11 years, and 15 years. The study collects a variety of socio-economic and demographic characteristics of the cohort members and their families, as well as information on parenting and childcare. From age 3 onward, data on cognitive and non-cognitive development are also available.

As far as cognitive outcomes are concerned, reading and word assessment tests are consistently available throughout waves 2 to 6. Numerical skills, on the other hand, are measured less frequently and have limited cross-wave comparability. Cognitive skills are assessed through age-appropriate standardised tests from the British Ability Scales (BAS) from waves 2 to 5. In order to capture reading and vocabulary skills, I rely on the BAS Naming Vocabulary scale for waves 2 and 3, the BAS Word Reading scale for wave 4, and BAS Verbal Similarity for wave 5 (see Hansen, 2014, for further details on the tests available for each wave). In wave 6 the only available word assessment is devised on the basis of standardised vocabulary tests developed by the Applied Psychology Unit at the University of Edinburgh in 1976 (this measure was already used to evaluate children in the same age range in the 1970 British Cohort Study). The measure of cognitive ability I use here is derived from the standardisation of the age-adjusted standardised t-scores from each of the tests described above (henceforth, referred to as 'reading test-scores' for simplicity).[1]

[1]The only exception is constituted by the vocabulary test at wave 6, for which only a raw score is available;

The measure of non-cognitive outcomes available for most waves of the MCS is the Strength and Development Questionnaire (SDQ). The SDQ is a screening test consisting of a set of age-appropriate questions assessing the behavioural and emotional health of children aged 3 to 16. The questionnaire is compiled by the cohort member's main caregiver in waves 2 to 6. Additionally, teacher-reported SDQ is available in waves 4 and 5 of MCS. The questionnaire is made of 25 items, which can be divided between five different scales: emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behaviour. Emotional symptoms and peer problems make up the category 'internalising problems', while conduct problems and hyperactivity/inattention constitute the 'externalising problems' category. Both categories are measured on a scale going from 0 to 20, which I reverse so that high values of SDQ correspond to better behavioural outcomes. As argued by Goodman et al. (2010), in low-risk samples, using these two broader categories yields better cross-sectional discriminant validity with respect to using the five SDQ scales. See Table 3.A1 in the Appendix for more details on measurement and on the items that make up internalising and externalising SDQ.

As it is often the case in cohort studies, reported family income in MCS is not continuous, but instead limited to a discrete number of bands that vary from wave to wave. Respecting the limits imposed by the extremes of each income band, the data providers developed a measure of imputed income using interval regression. Among the predictors of income were respondent's age, housing tenure, region of residence, education, and labour market status (see Millennium Cohort Study, 2020, for a full list of predictors and more details on the imputation procedure).The measure of imputed income was then equivalised in order to account for economies of scale within the family, using the OECD household equivalence scale. While this measure allows to have a continuous income variable in the dataset, it is likely to be affected by measurement error and to only partly reflect the latent income of the families in the survey. In order to limit the sensitivity of the results to this measurement issue, I build my main explanatory variables (that is, income gains and losses between consecutive periods) based on the quintiles of the equilvalised imputed income. This approach has the advantage of closely reflecting self-reported banded income, without suffering from the cross-wave differences in the definition of the bands.[2] Furthermore, it allows me to capture relatively larger variations in family income, as transitions from one income quintile to another will arguably be observed only for sufficiently large income gains or losses (I formally test whether this is indeed

I standardise it beforehand to match the same range of the standardised reading scores of the previous waves.

[2]The income bands extremes and the number of bands changing from wave to wave, it is a difficult task to harmonize such categories. Refer to the MCS data documentation and questionnaires for further details on the definition of income bands for each wave.

the case in Section 3.5). However, as shown more in detail in the robustness checks section, results are qualitatively similar when using the broader range of information coming from the continuous measure of imputed income provided in MCS. Transition matrices showing the raw probability of moving across quintiles of the distributions of income, reading test-scores, and SDQ from one wave to the next in the estimation sample are reported in the Appendix (Figures 3.A1 to 3.A4).

3.3.2 Empirical strategy

In this paper, I describe the evolution of children's cognitive and non-cognitive outcomes between two consecutive periods as a function of changes in household income. For this scope, I adopt a 'lagged score' value added model (Koedel, Mihaly and Rockoff, 2015), which can be read as a model generating from an autoregressive process of order one. This method explores the dynamics of human capital formation by capturing the residualised changes in the measures of cognitive and non-cognitive skills described in Section 3.3.1, while accounting for their unobserved time-invariant determinants. For each of the outcomes of interest (i.e. internalising SDQ, externalising SDQ, and reading test-scores), I estimate the following regression using pooled OLS:

$$Y_{i,t} = \beta_0 + \beta_1 Y_{i,t-1} + \beta_2 L_{i,t} + \beta_3 G_{i,t} + \sum_{s=2}^{5} \gamma_s I^s_{i,t-1} + X'\delta + \zeta_t + \varepsilon_{i,t} \tag{3.1}$$

where $Y_{i,t}$ is one of the three outcomes of interest for individual i at time t, all of which are standardised. $L_{i,t}$ and $G_{i,t}$ are dummy variables indicating respectively whether there was a loss or a gain in household income between period $t-1$ and period t. As discussed above, income is coded as quintiles of equivalised imputed income and a loss (gain) is realised when the household the cohort member belongs to is in a lower (higher) income quintile with respect to the previous wave. By separately controlling for gains and losses in household income, income changes are allowed to have an asymmetric effect on the accumulation of cognitive and non-cognitive skills. $\{I^2_{i,t-1}, ..., I^5_{i,t-1}\}$ is a set of four dummies indicating the income band reported by the carer of child i in wave $t-1$ ($I^1_{i,t-1}$, i.e. the dummy indicating the first income band, is omitted and used as the reference category). X is a vector of standard controls, including child and household's time-invariant characteristics such as sex, mother's age at birth, and child ethnicity; lagged characteristics and their variation between $t-1$ and t (housing tenure and its variation); covariates at time t, such as single-parent household, whether both parents participate to the labour market, and the square root of household size (see Table 3.1 for a full list of controls). Finally, ζ_t is a set of wave dummies. Standard errors are clustered at the child level.

Thanks to the richness of the dataset, I am able to test whether the effect of income changes on cognitive and non-cognitive outcomes is at least partly mediated by channels pertaining to the well-being of the parents. As it is often the case in cohort studies, parental variables are measured more accurately for mothers than they are for their spouse. This is because mothers tend to be the main caregiver and, hence, the main survey respondent. Furthermore, fathers might not always be present in the household at all waves and might not always coincide with the mother's partner or spouse. Because of this I focus on maternal well-being as a potential mediator of the effect of income changes on the accumulation of children's cognitive and non-cognitive skills. In order to capture mothers' physical well-being, I rely on a measure of self-assessed general health derived from the question "How would you describe your health generally?". Potential answers are "Excellent", "Very good", "Good", "Fair", and "Poor". As for psychological well-being, I use two measures to capture both the affects and the cognitive dimensions of well-being. The Kessler Psychological Distress Scale (K6), measuring affects, is a 6-items scale assessing mood and anxiety disorders in a short-term horizon. The question is introduced by the sentence "During the past 30 days, about how often did you feel...", followed by the items: "...nervous?", "...hopeless?", "...restless or fidgety?", "...so depressed that nothing could cheer you up?", "...that everything was an effort?", "...worthless?". Answers range from 1, meaning "all of the time", to 5, meaning "none of the time". Life satisfaction is used to measure cognitive well-being: respondents are faced with a scale going from 1, meaning "that you are completely dissatisfied" and 10, meaning "that you are completely satisfied" and asked to choose a number indicating how satisfied or dissatisfied they are about the way their life has turned out up to that moment.

The new specification mirrors the one described above, allowing for mothers' physical and psychological well-being to act as mediators:

$$Y_{i,t} = \beta_0 + \beta_1 Y_{i,t-1} + \beta_2 L_{i,t} + \beta_3 G_{i,t} + \sum_{s=2}^{5} \gamma_s I_{i,t-1}^{s} + (\Delta C_i)' \mu_1 + C_{i,t-1}' \mu_2 + \mathbf{X}' \delta + \zeta_t + \varepsilon_{i,t} \quad (3.2)$$

where $C_{i,t-1}$ is a vector containing the aforementioned measures of maternal well-being at time $t-1$: the Kessler K6 score, life satisfaction, and a dummy equal one if self-assessed health is either "fair" or "poor". All measures are coded in such a way that higher values reflect better outcomes. ΔC_i is a vector capturing the changes in the maternal channels, containing the standardised differences of the levels of psychological well-being between time $t-1$ and time t, and a dummy equal one if there was a worsening in the mother's self-assessed general health between the two same periods.

Conditioning on the availability of the dependent variables, the final estimation sample consists of 40,189 observations (14,394 cohort members, each observed on average for 3.8 waves).[3] Missing values of the explanatory variables were imputed using mean imputation; thus all regressions control for dummies indicating the position of the missing values for each variable.[4] Sampling weights and non-response weights provided by MCS are used throughout the analysis. Table 3.A2 in the Appendix describes the features of the estimation sample. Around 22% of children experience a drop in their family income that moves them to a lower quintile of the income distribution; gains in family income quintile are instead experienced by around 27% of the estimation sample.

3.4 Results

3.4.1 Main regressions

Table 3.1 presents estimates of different specifications of the baseline model of equation (3.1): for each of the three dependent variables, the first column (i.e. columns 1, 3, 6) reports pooled OLS estimates of the baseline model without the lagged outcome; in columns 2, 4, and 6 instead the lagged value of the outcome variable is added to the model. Irrespective of the specification used, income losses seem to be systematically associated with lower levels of both reading test-scores and the two dimensions of SDQ. While income gains appear to foster cognitive skills, their effect on non-cognitive outcomes cannot be distinguished from zero. Comparing the first two columns for each outcome, it seems that the adoption a value added specification improves the fit of the model without qualitatively affecting the estimated coefficients.

Overall, the effect of moving to a lower income quintile is associated with a loss of about 3 to 4% of a SD of both externalising and internalising SDQ, and a loss of 3.5% of a SD in the standardised reading t-scores distribution. Although the effect sizes might look modest at first sight, the contribution of an income loss to the residualised internalising and externalising SDQ is comparable to about half the effect of being born with a weight lower than 2.5 kg, and for internalising SDQ it is not statistically different from the magnitude of the effect of being the first-born. While losses appear to play a larger role than gains in explaining residualised SDQ,

[3]Note that information on the first wave a cohort member is observed are only used as lagged values in relationship to the second wave of observation. So in practice, the estimation is conducted on average on 2.8 waves per cohort member.

[4]Missingness is not a big problem in MCS: the percentage of imputed missing values is never above 5% for the main explanatory variables. Predictably, results are not sensitive to the imputation of missing values and hold also when the correspondent observations are dropped from the sample.

pairwise Wald tests fail to reject the equality (in absolute value) of the coefficients for income gains and losses for all outcomes (the p-values of the tests are, respectively, 0.15 for externalising SDQ and 0.34 for internalising SDQ).[5]

3.4.2 Channels

The literature in economics and developmental psychology suggest that family income changes can affect children's human capital accumulation directly, through the provision of material inputs, and/or indirectly, through changes in parents' well-being, which can in turn affect the process of skills formation. While income gains and losses arguably reflect changes in parents' ability to provide material inputs to their children (e.g. piano lessons, books), specification (3.1) does not take into account other mechanisms. Table 3.2 uses the value added model described in specification (3.2) to explore the presence of mediators of the effect of income losses and gains reported in Table 3.1. The magnitude of the coefficients estimated in Table 3.1 might in fact reflect the presence of channels, such as mothers' well-being, that are likely positively correlated with income changes. As expected, the variables capturing the changes between $t-1$ and t, as well as the levels in $t-1$, of the mother's psychological and physical health explain a significant portion of the child's human capital formation trajectories and their introduction in the specification reduces on average the magnitude of the coefficients for both gains and losses. For internalising and externalising SDQ, about one third of the effect of income losses appears to transit through these channels - although the estimates are not precise enough to rule out the equality of the coefficients across specifications.

The coefficients of income gains and losses for reading test-scores are instead more robust to the introduction of potential mediators, perhaps suggesting that income changes have a stronger direct effect on school performance rather than internalising or externalising behaviour. The values of the adjusted R-squared across each pair of specifications in Table 3.2 further shows that the introduction of channels marginally improves the model's prediction in the case of internalising and externalising

[5]Note that the effect of income quintile gains and losses can reflect both pure mobility effects and positional effects deriving from the new family income quintile at time t. As controlling for income quintiles at time t would introduce an identification problem, a possible alternative suggested by sociologists is recurring to a diagonal reference model (DRM), typically used to study social mobility (Sobel, 1981, 1985). Under a set of assumptions, DRM provide a way to disentangle origin, destination, and mobility effects. Results from this model are presented in Table 3.A3, where the coefficients for upwards and downwards mobility along the income distribution are only statistically meaningful (and of similar magnitude to the effects shown in Table 3.1) for reading test-scores. However, estimates of mobility from DRM are known to suffer from a high chance of type-II error, as shown by a body of null or weak evidence on mobility effects (see, for instance, Chan, 2018; Houle and Martin, 2011; Kaiser and Trinh, Forthcoming; Schuck and Steiber, 2018; Tolsma, De Graaf and Quillian, 2009; Zang and Dirk de Graaf, 2016) − in contradiction with the predominant sociological theoretical frameworks.

SDQ, but not for test-scores. This is also consistent both with Duncan and Brooks-Gunn (1997), who suggest that cognitive skills, with respect to non-cognitive ones, rely more heavily on material inputs. Yeung, Linver and Brooks-Gunn (2002)'s findings further corroborate the results presented in Table 3.2, in at least two ways: first, their paper shows that the effect of income instability on non-cognitive skills is mostly conveyed through maternal affects; secondly, they show that the effect on cognitive skills is in larger part mediated by material investments, rather than mothers' emotional health. Qualitatively similar predictions are also supported by Washbrook, Gregg and Propper (2014).

3.4.3 Dynamic panel data analysis

While the value added model accounts for unobserved time-invariant factors explaining the dependent variable, there might still be some unobserved time-invariant factors affecting the residualised outcome, that is the portion of the outcome that is not explained by its past value. Such residual unobserved between-individuals heterogeneity can be addressed thanks to the panel structure of the data, by including individual fixed effects and thus isolating within variation only. However, the naive combination of a value added model with fixed effects would lead to a form of dynamic panel bias known as the Nickell bias (Nickell, 1981): through the demeaning process of fixed effects regression, the demeaned lagged value of the outcome (now the endogenous regressor) can no longer be distributed independently of the error term. The deriving endogeneity produces a bias that Nickell shows to be larger in samples with "small T and large N" - situation mirroring the MCS sample. A solution to this problem is the adoption of a system generalised method of moments (GMM) estimator (Arellano and Bover, 1995; Blundell and Bond, 1998). The estimator is derived from a system of two simultaneous equations (the regression model specified in first-differences and in levels), in which the endogenous variables are instrumented with suitable lags of, respectively, their own levels and their first differences (under the assumption that the changes in the instrumenting variables are uncorrelated with the fixed effects; see Roodman, 2009). Table 3.A4 compares the performance of pooled OLS (same as in Table 3.1) and system GMM.[6]. The first and second columns of each GMM specification differ for the number of GMM-style instruments used for the endogenous

[6] I here implement the system GMM estimator in Stata v 16.0 using the `xtabond2` command developed by David Roodman (see Roodman, 2009, for an introduction to difference and system GMM and the use of `xtabond2`). All variables are considered as included instruments, except for the lag of the dependent variable. This is instead instrumented GMM-style using its own lags of order two or higher. Standard errors are estimated with a two-step procedure, with a finite-sample correction (Windmeijer, 2005). Instead of first-differences, orthogonal deviations are used in order to minimise the loss of information due to the presence of gaps in the panel (Arellano and Bover, 1995).

regressor (the lag of the outcome variable): columns 2, 5, and 8 use only the outcome's lags of order two or greater to build the instruments, while columns 3, 6, and 9 use the same lags for all the available outcomes (i.e. Externalising and Internalising SDQ, standardised reading test-scores). The size of the autoregressive coefficient for the lagged value of each outcome in the GMM columns constitutes an indirect validity test for the specification of the model, as the coefficient lays between the FE (not shown in the table) and the OLS estimates (as shown by Hsiao, 2014, these are, respectively, a lower and an upper bound for the true value of the coefficient). The GMM estimates of gains and losses appear to be qualitatively similar to (where not of significantly larger magnitude than) the OLS ones for all outcomes. This suggests that the omission of time-invariant factors that are potentially correlated with the residualised outcome might translate into an attenuation bias at worst; as such, the coefficients from baseline value-added model without individual fixed effects can be interpreted as lower bounds of the real effect of income gains and losses. Differently from the OLS estimates in columns 1 and 4, gains appear to be statistically meaningful in explaining part of the residualised outcome in all GMM specifications of Internalising and Externalising SDQ, although their magnitude is lower than that of losses (the difference between the two absolute coefficients being statistically different from zero at the 5% level in the case of Externalising SDQ).

The use of system GMM does however come with a set of stringent assumptions. A crucial one is of course that the instruments should be exogenous (that is, uncorrelated with the error term). However, the Hansen J-statistic testing for over-identifying restrictions rejects the null hypothesis of joint validity of the instruments, no matter which combination of lagged outcomes is used as GMM-style instruments. Additionally, the use of the in-levels equation in system GMM require an extra assumption to hold, that is the first differences of the instrumenting variables should be uncorrelated with the time-invariant component of the error term (i.e. the fixed effects). This is equivalent to saying that, conditionally on all other covariates, the observed deviations in the instruments from one period to the next should be taken as deviations from a sort of stationary state and, as such, they do not depend on intrinsic individual characteristics (Roodman, 2009). Given of the absence of convincing evidence in support of the identifying assumptions required by system GMM and the conservative size of the OLS estimates with respect to the dynamic panel data ones, a pooled OLS estimator of the value-added model illustrated by equation 3.1 will be used throughout the remainder of the paper.

3.5 Robustness Checks

3.5.1 The measurement of income

One important concern with the analysis conducted above is linked to the interpretation of the coefficients of gains (losses) for individuals at the top (bottom) income quintile in $t-1$. Due to the discrete nature of the income variable used, these individuals cannot transition upwards (downwards) the income distribution, hence gains (losses) are not defined for them. A way of getting around the issue is to replicate the estimates above using only cohort members who can potentially transition both upwards and downwards the income quintile scale, that is, excluding in each wave those individuals who were either in the top or in the bottom quintile of the household income distribution in the previous wave. Columns 1, 3, and 5 of Table 3.A5 in the Appendix replicate the baseline value-added model for a sub-sample of cohort members whose family income is neither in the top nor in the bottom quintile around waves 2 to 5. Although the coefficients for reading test-scores are less precisely estimated, the same considerations made for Table 3.1 qualitatively hold.

One could take a step further and exclude from the estimation sample not only individuals whose upwards or downwards movements across the income quintile distribution are made impossible because of their position in either one of the its extremes, but also those for whom the *size* of the jump is constrained because of their position. As an example, keeping all other things constant, a cohort member who finds herself in the fourth income quintile and experience a family income gain in the next period can only transition to the fifth quintile, no matter how large the gain her family experienced. On the contrary, the gain experienced by someone going from the third to the fourth quintile is less limited by the scale of the income variable (had the relative gain been larger, such person could have potentially transitioned to the top quintile). Results for cohort members whose movements are not constrained to one-quintile jumps across the income distribution can be found in columns 2, 4, and 6 of Table 3.A5. Although of larger magnitude, the estimated coefficients of income gains and losses are overall consistent with results in Table 3.1.

So far I only considered income as measured by quintiles. Despite the issues linked to its measurement (see discussion in Section 3.3.1), the MCS imputed measure of continuous family income has the potential to provide extra layers of information that could be useful in disentangling the effect of more sophisticated categories of gains and losses. Arguably, gains and losses based on income quintiles will likely capture larger changes in family income, while changes that are not large enough to drive a family out of their income quintile are considered as an absence of change (I formally test that this is indeed the case at the end of this section). Additionally, an analysis based on the continuous imputed measure of income would not depend on the relative position of individuals across the income distribution, but would be based on their absolute income status instead. As income in the MCS is imputed using not only banded income, but also information on educational status, age, geography and a variety of other covariates (see Millennium Cohort Study, 2020, for more details on the imputation procedure), it can be interpreted as a broader measure of socio-economic status.

First, I computed the growth rate of imputed equivalised income between one period and the next, splitting it into two variables: one, 'positive income growth', reflecting its positive values (and equal zero for all negative values) and the other, 'negative income growth', reflecting the absolute value of its negative values (and equal to zero for all positive values). I then substituted the loss and gain dummies in equation 3.1 with positive and negative income growth. The distribution of the income growth rate is roughly normal, centered around zero, with a long right tail. Results for this specification are illustrated in Table 3.A6, trimming any income growth rate larger than 10 (top 0.5% of its distribution). [7] The story shown by the Table is consistent with that implied by Table 3.1: negative income growth hinders both cognitive and non-cognitive outcomes (although the effect is not always precisely estimated). Differently from the baseline, a positive income growth rate between one period and the next is now significantly associated with better measures of non-cognitive skills, although the absolute effect size is roughly one third of that of negative income growth. Another difference with the baseline results is the difference in magnitude between the gains and losses coefficients for reading test-scores: while the baseline estimates suggested symmetry of gains and losses, here the absolute values of the positive and negative income growth estimates are statistically different from each other at the 10% level – with gains affecting learning outcomes to a lesser extent than losses.

An assumption implied so far is that income gains and losses (defined by transitions across the income quintile distribution) are 'large'. However, changes in a family's relative income position could well occur even in response to relatively small changes in imputed income. I here explore the composition of income changes involved in the gain and loss dummies and their relative role in shaping human capital accumulation. In Table 3.A7 the income quintile gain (loss) indicator is decomposed into four dummies, based on the magnitude of the continuous income growth rate

[7] Note that the sample size is smaller than the baseline, because of missing values of imputed income and trimming of the right tail. Baseline results still hold in this smaller sample.

associated driving the underlying upwards (downwards) quintile movement.[8] While we can almost never reject the equality of all the losses (gains) coefficients in each column, Table 3.A7 suggests that the baseline results from Table 3.1 are not primarily driven by gains and losses induced by small income changes: income quintile losses (gains) associated with a -25% (25%) income growth rate or smaller (greater) are the ones to attract the most statistically significant estimates. This is somewhat unsurprising, as about 54% (80%) of all downwards (upwards) movements in the income quintile distribution involve an income growth rate of -25% or lower (25% or greater).

3.5.2 Omitted variables

One question that might emerge at this point concerns what are the drivers of these upwards and downwards movements across the household income distribution. Income changes are indeed likely to depend on factors such as changes in the country's social security system, in the labour market status of the parents, in the household's demographic structure, in housing tenure. However, is the process of human capital formation affected by these changes per-se, or does income have a direct way of affecting cognitive and non-cognitive outcomes? In other words, do these factors affect the outcomes only through changes in income or are they omitted variables threatening to confound its effect?

Table 3.3 is an attempt to clarify the matter. Columns 1, 3, and 5 replicate columns 2, 5, and 8 of Table 3.1. Columns 2, 4, and 6 introduce a list of life events between $t-1$ and t that are likely to be correlated with changes in quintiles of equivalised income. Since housing tenure and its changes are already controlled for in all specifications, remaining determinants of income changes I could control for are separations, job losses and job changes, and additional changes in household composition driven by siblings. The coefficients of gains and losses are overall robust to the introduction of these potential confounders, suggesting that their omission does not contribute to the creation of an omitted variables bias. This evidence is partly in contrast with Washbrook, Gregg and Propper (2014), who find that the income gradient of non-cognitive skills and health is completely shut out by distal factors such as socio-demographic and labour market outcomes, with only one fifth of the effect of income on cognitive skills surviving the introduction of these covariates. Conditional on current employment status, changes in the parents' labour force status from one period to the next do not appear to explain changes in the residualised cognitive and

[8]Here I chose 25%, 10%, and 5% (and their negative equivalents) as arbitrary thresholds to distinguish between different categories of income growth. Results are however robust to a battery of other thresholds and number of intervals.

non-cognitive outcomes. A parent leaving the household appears to be negatively associated with the residualised measures of non-cognitive outcomes (the association being statistically significant at the 10% level only for Internalising SDQ), while no effect is found on reading test-scores. Changes in the siblings pool composition appear to have a negative effect on child human capital accumulation, especially in the case of socio-emotional development. Externalising problems increase with the presence of new siblings, consistent with children engaging in disruptive behaviours to capture the parents' attention. The results for internalising symptoms instead hide substantial heterogeneity across gender: while boys have lower residualised internalising SDQ when younger siblings are born, girls are only significantly affected by an older sibling leaving the household.

3.6 Additional results

3.6.1 Persistence

As shown by results in Tables 3.1 and 3.2, income gains between $t-1$ and t do not seem to be statistically significant in explaining changes in non-cognitive outcomes, while income losses have a significant negative impact. One may wonder whether the same is true for past movements across family income quintiles. Table 3.4 investigates the role of past gains and losses, as well as current ones, and their interactions over time. The Table shows a picture similar to that of Table 3.1 for recent gains and losses ($Gain_t$ and $Loss_t$). While there is some evidence that past income losses decrease residualised Internalising SDQ and reading test-scores, these effects are not statistically significant. Similarly, past income gains appear to foster human capital, significantly so only for reading test scores. As household income losses seem to affect cohort members partly through parents' well-being, it seems plausible that their effect on child human capital be mostly immediate, driven by affects. As shown by Boyce *et al.* (2013), income gains typically have a positive impact on subjective well-being of a lower magnitude with respect to that of losses. An income gain between time $t-2$ and $t-1$ might not have a strong enough impact on parents' well-being to justify a positive effect on non-cognitive human capital formation at time $t-1$, but it might still enable parents to put in place material investments fostering their children's cognitive skills that will still have an effect at time t, thus explaining the positive effect of past income gains on reading test scores. There is however no evidence of complementarity between income gains in two consecutive periods: if anything they appear to have a certain degree of substitutability, as shown by the negative coefficient for the interaction between two consecutive gains. On the other hand, old income losses seem to matter only in relationship to current income losses, exacerbating their negative relationship with reading test-scores.

Results from Table 3.4 can be interpreted in relationship to the literature on homeostatic well-being (Cummins, 2016). As about one third of the effect of income losses on Internalising and Externalising SDQ is mediated by mothers' well-being, one might wonder whether the absence of persistence of past income losses is due to an adaptation mechanism that pushes mothers' well-being back towards its homeostatic level. I test for this possibility, by replicating Table 3.4 for the two outcomes reflecting mothers' psychological well-being, namely life satisfaction and the Kessler (K6) scale of affects (results available on request). As expected, I find evidence of mothers adapting to income changes both in terms of affects and cognitive well-being, with the measure of affects adapting at a faster rate than the cognitive one. Since the effect of income changes on reading test-scores does not seem to be mediated by any parental well-being channel, the well-being adaptation mechanism does not affect the persistence of past losses and gains, which matter both in absolute terms and in conjunction with current income changes.

3.6.2 Transition dynamics

The results presented so far are just average effects across all income quintiles. However, following the approach of Bruckauf and Chzhen (2016), it might be interesting to focus on the risk factors that predict the entry to and exit from the bottom quintile of the income distribution.[9] Table 3.5 reports average marginal effects derived from logistic regressions predicting the probability of entering or exiting the bottom quintile of the cognitive or non-cognitive skills distributions. Note that the estimation samples here are different: by construction, cohort members who are already at the bottom quintile of an outcome's distribution are dropped from the estimation sample of the column tagged "entry" (unless they transition into a higher quintile and then back again into the lowest one). For exit instead, the estimation sample is made up only by cohort members who already were in the bottom quintile of the outcome's distribution in $t-1$.

Controlling for the position in the income distribution in period $t-1$, a movement down the income distribution quintiles is associated with a 2 pp increase of the probability of entering the

[9] Income quintile changes may not be independent of the child's position in the distribution of cognitive and non-cognitive outcomes. I empirically test whether that is the case in the estimation sample and find little differences in the probability of experiencing income gains or losses between individuals at the bottom quintile of any outcome's distribution and those in higher quintiles. The likelihood of experiencing income gains (losses) is 0.3 pp higher (0.9* pp lower) for those at the bottom quintile of the Externalising SDQ distribution; 1** pp higher (0.4 pp lower) for those at the bottom quintile of the Internalising SDQ distribution; and 0.8 pp higher (1** pp lower) for those at the bottom quintile of the reading test-scores distribution.

bottom quintile of the externalising SDQ distribution. While losses seem to predict the probability of entering in the bottom quintile of the SDQ distributions, gains are associated with a lower likelihood of entering the bottom quintile of internalising SDQ and reading test-scores. Neither income losses nor gains seem to contribute to explaining transition dynamics out of the bottom quintile of non-cognitive outcomes (with the exception of losses for externalising SDQ). On the other hand, for reading test-scores, income gains are associated with a higher probability of exiting the outcome's bottom quintile.

3.7 Conclusion

This paper explores the relationship between changes in family income and the accumulation of children's cognitive and non-cognitive skills. By relaxing the assumption of a symmetric impact of losses and gains, I find that losses matter more than gains in explaining changes in non-cognitive outcomes between one wave and the next. Movements downwards the distribution of family income are associated with a decrease of 3 to 4% of a standard deviation for both SDQ and reading test-scores, an effect size comparable to that of a parent leaving the household.

The effect of losses is mediated for one third by channels reflecting mothers' well-being. Losses also predict the probability of transitioning into the bottom quintile of the distribution of both non-cognitive and cognitive abilities; for the latter, experiencing a loss hinders the probability of moving out of the bottom of the distribution. Moving upwards the family income distribution, on the contrary, is correlated with both higher probability of exiting and lower probability of entering the bottom quintile of the reading test-scores distribution. The effect of gains on reading test scores is also persistent in time: past income gains still matter for today's cognitive trajectories, consistently with the theory of family investment.

Despite the robustness of the results presented above to a battery of sensitivity tests, the empirical strategies used throughout the paper remain exposed to potential endogeneity issues. However, results are consistent with the established literature in economics and developmental psychology and contribute to uncovering some novel mechanism. From a policy perspective, the findings of this paper suggest that income transfers, while fostering cognitive skills, might not have the same effect on non-cognitive skills. The fact that human capital accumulation appears to be more sensitive to income losses might provide yet another piece of evidence in support of insurance and welfare policies to limit the negative impact of adverse economic conditions, paying particular attention to the effects on the psychological well-being of adults.

References

Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja, and Tristan Zajonc. 2011. "Do value-added estimates add value? Accounting for learning dynamics." *American Economic Journal: Applied Economics*, 3: 29–54.

Arellano, Manuel, and Olympia Bover. 1995. "Another look at the instrumental variable estimation of error-components models." *Journal of Econometrics*, 68: 29–51.

Black, Nicole, and Sonja C. Kassenboehmer. 2017. "Getting weighed down: the effect of childhood obesity on the development of socioemotional skills." *Journal of Human Capital*, 11: 263–295.

Blanden, Jo, and Stephen Machin. 2010. "Intergenerational inequality in early years assessments." In *Children of the 21st century: The first five years.* Vol. 2, Eds. K. Hansen, H. Joshi and S. Dex, Chapter 9, 153–168. Policy Press.

Blau, David M. 1999. "The effect of income on child development." *Review of Economics and Statistics*, 81: 261–276.

Blundell, Richard, and Stephen Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics*, 87: 115–143.

Boyce, Christopher J., Alex M. Wood, James Banks, Andrew E. Clark, and Gordon D.A. Brown. 2013. "Money, well-being, and loss aversion: Does an income loss have a greater effect on well-being than an equivalent income gain?" *Psychological Science*, 24: 2557–2562.

Bruckauf, Zlata, and Yekaterina Chzhen. 2016. *Poverty and Children's Cognitive Trajectories: Evidence from the United Kingdom Millennium Cohort Study.* UNICEF - Innocenti Working Paper, 14.

Chan, Tak Wing. 2018. "Social mobility and the well-being of individuals." *British Journal of Sociology*, 69: 183–206.

Clark, Andrew E., Conchita D'Ambrosio, and Marta Barazzetta. 2021. "Childhood circumstances and young adulthood outcomes: The role of mothers' financial problems." *Health Economics*, 30: 342–357.

Conger, Rand D., Katherine J. Conger, Glen H. Elder Jr., Frederick O. Lorenz, Ronald L. Simons, and Les B. Whitbeck. 1992. "A family process model of economic hardship and adjustment of early adolescent boys." *Child Development*, 63: 526–541.

Cummins, Robert A. 2016. "The theory of subjective wellbeing homeostasis: A contribution to understanding life quality." In *A Life Devoted to Quality of Life.* Vol. 60, Ed. F. Maggino, Chapter 4, 61–79. Springer.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the technology of cognitive and noncognitive skill formation." *Econometrica*, 78: 883–931.

Dahl, Gordon B., and Lance Lochner. 2012. "The impact of family income on child achievement: Evidence from the earned income tax credit." *American Economic Review*, 102: 1927–56.

Dearden, Lorraine, Luke Sibieta, and Kathy Sylva. 2011. "The socio-economic gradient in early child outcomes: evidence from the Millennium Cohort Study." *Longitudinal and Life Course Studies: International Journal*, 2: 19–40.

Del Bono, Emilia, Marco Francesconi, Yvonne Kelly, and Amanda Sacker. 2016. "Early maternal time investment and early child outcomes." *Economic Journal*, 126: F96–F135.

Duncan, Greg J., and Jeanne Brooks-Gunn. 1997. *Consequences of Growing Up Poor.* Russell Sage Foundation.

Ermisch, John. 2008. "Origins of social immobility and inequality: parenting and early child development." *National Institute Economic Review*, 205: 62–71.

Evans, William N., and Craig L. Garthwaite. 2014. "Giving mom a break: The impact of higher EITC payments on maternal health." *American Economic Journal: Economic Policy*, 6: 258–90.

Flouri, Eirini, Emily Midouhas, and Heather Joshi. 2014. "The role of urban neighbourhood green space in children's emotional and behavioural resilience." *Journal of Environmental Psychology*, 40: 179–186.

Flouri, Eirini, Stella Mavroveli, and Nikos Tzavidis. 2012. "Cognitive ability, neighborhood deprivation, and young children's emotional and behavioral problems." *Social Psychiatry and Psychiatric Epidemiology*, 47: 985–992.

Frank, Richard G., and Ellen Meara. 2009. *The Effect of Maternal Depression and Substance Abuse on Child Human Capital Development.* NBER Working Paper Series, 15314.

Frijters, Paul, John P. Haisken-DeNew, and Michael A. Shields. 2004. "Money does matter! Evidence from increasing real income and life satisfaction in East Germany following reunification." *American Economic Review*, 94: 730–740.

Gardner, Jonathan, and Andrew J. Oswald. 2007. "Money and mental wellbeing: A longitudinal study of medium-sized lottery wins." *Journal of Health Economics*, 26: 49–60.

Goodman, Anna, Donna L. Lamping, and George B. Ploubidis. 2010. "When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers and children." *Journal of Abnormal Child Psychology*, 38: 1179–1191.

Heckman, James J., and Yona Rubinstein. 2001. "The importance of noncognitive skills: Lessons from the GED testing program." *American Economic Review*, 91: 145–149.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor Economics*, 24: 411–482.

Houle, Jason N., and Molly A. Martin. 2011. "Does intergenerational mobility shape psychological distress? Sorokin revisited." *Research in Social Stratification and Mobility*, 29: 193–203.

Hsiao, Cheng. 2014. *Analysis of Panel Data.* Cambridge University Press.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect theory: An analysis of decision under risk." *Econometrica*, 47: 263–292.

Kaiser, Caspar, and Nhat An Trinh. Forthcoming. "Positional, mobility and reference effects: How does social class affect life satisfaction in Europe?" *European Sociological Review*.

Kelly, Y., A. Sacker, E. Del Bono, M. Francesconi, and M. Marmot. 2011. "What role for the home learning environment and parenting in reducing the socioeconomic gradient in child development? Findings from the Millennium Cohort Study." *Archives of Disease in Childhood*, 96: 832–837.

Koedel, Cory, Kata Mihaly, and Jonah E. Rockoff. 2015. "Value-added modeling: A review." *Economics of Education Review*, 47: 180–195.

Kuehnle, Daniel. 2014. "The causal effect of family income on child health in the UK." *Journal of Health Economics*, 36: 137–150.

McLoyd, Vonnie C. 1990. "The impact of economic hardship on Black families and children: Psychological distress, parenting, and socioemotional development." *Child Development*, 61: 311–346.

McLoyd, Vonnie C., Toby Epstein Jayaratne, Rosario Ceballo, and Julio Borquez. 1994. "Unemployment and work interruption among African American single mothers: Effects on parenting and adolescent socioemotional functioning." *Child Development*, 65: 562–589.

Millennium Cohort Study. 2020. *Millennium Cohort Study - User Guide (Surveys 1 to 5).* 9^{th} ed., Centre for Longitudinal Studies.

Milligan, Kevin, and Mark Stabile. 2011. "Do child tax benefits affect the well-being of children? Evidence from Canadian child benefit expansions." *American Economic Journal: Economic Policy*, 3: 175–205.

Nickell, Stephen. 1981. "Biases in dynamic models with fixed effects." *Econometrica*, 1417–1426.

Noonan, Katharine, Richéal Burns, and Mara Violato. 2018. "Family income, maternal psychological distress and child socio-emotional behaviour: Longitudinal findings from the UK Millennium Cohort Study." *SSM - Population Health*, 4: 280–290.

Powdthavee, Nattavudh. 2010. "How much does money really matter? Estimating the causal effects of income on happiness." *Empirical Economics*, 39: 77–92.

Roodman, David. 2009. "How to do xtabond2: An introduction to difference and system GMM in Stata." *Stata Journal*, 9: 86–136.

Roy, John, and Stefanie Schurer. 2013. "Getting stuck in the blues: persistence of mental health problems in Australia." *Health Economics*, 22: 1139–1157.

Sampson, Robert J., and John H. Laub. 1994. "Urban poverty and the family context of delinquency: A new look at structure and process in a classic study." *Child Development*, 65: 523–540.

Schuck, Bettina, and Nadia Steiber. 2018. "Does intergenerational educational mobility shape the well-being of young Europeans? Evidence from the European Social Survey." *Social Indicators Research*, 139: 1237–1255.

Shea, John. 2000. "Does parents' money matter?" *Journal of Public Economics*, 77: 155–184.

Smith, Judith R., and Jeanne Brooks-Gunn. 1997. "Correlates and consequences of harsh discipline for young children." *Archives of Pediatrics and Adolescent Medicine*, 151: 777–786.

Sobel, Michael E. 1981. "Diagonal mobility models: A substantively motivated class of designs for the analysis of mobility effects." *American Sociological Review*, 46: 893–906.

Sobel, Michael E. 1985. "Social mobility and fertility revisited: Some new models for the analysis of the mobility effects hypothesis." *American Sociological Review*, 50: 699–712.

Tamura, K., J. Morrison, and H. Pikhart. 2020. "Children's behavioural problems and its associations with socioeconomic position and early parenting environment: findings from the UK Millennium Cohort Study." *Epidemiology and Psychiatric Sciences*, 29.

Todd, Petra E., and Kenneth I. Wolpin. 2003. "On the specification and estimation of the production function for cognitive achievement." *Economic Journal*, 113: F3–F33.

Todd, Petra E., and Kenneth I. Wolpin. 2007. "The production of cognitive achievement in children: Home, school, and racial test score gaps." *Journal of Human Capital*, 1: 91–136.

Tolsma, Jochem, Nan Dirk De Graaf, and Lincoln Quillian. 2009. "Does intergenerational social mobility affect antagonistic attitudes towards ethnic minorities?" *British Journal of Sociology*, 60: 257–277.

Washbrook, Elizabeth, Paul Gregg, and Carol Propper. 2014. "A decomposition analysis of the relationship between parental income and multiple child outcomes." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 177: 757–782.

Windmeijer, Frank. 2005. "A finite sample correction for the variance of linear efficient two-step GMM estimators." *Journal of Econometrics*, 126: 25–51.

Yeung, W. Jean, Miriam R. Linver, and Jeanne Brooks-Gunn. 2002. "How money matters for young children's development: Parental investment and family processes." *Child Development*, 73: 1861–1879.

Zang, Emma, and Nan Dirk de Graaf. 2016. "Frustrated achievers or satisfied losers? Inter-and intragenerational social mobility and happiness in China." *Sociological Science*, 3: 779–800.

Figures and Tables

Table 3.1: The effect of income changes on child human capital

	Externalising SDQ		Internalising SDQ		Reading test-scores	
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome ($t-1$)		0.644***		0.518***		0.287***
		(0.007)		(0.008)		(0.006)
Gain	0.019	0.005	0.016	0.020	0.049***	0.041***
	(0.016)	(0.012)	(0.016)	(0.014)	(0.013)	(0.012)
Loss	-0.027*	-0.033***	-0.052***	-0.039***	-0.035***	-0.035***
	(0.016)	(0.012)	(0.016)	(0.014)	(0.013)	(0.012)
2nd income quintile ($t-1$)	0.034	0.029	0.053**	0.031	0.072***	0.056***
	(0.026)	(0.020)	(0.026)	(0.022)	(0.019)	(0.018)
3rd income quintile ($t-1$)	0.083***	0.057***	0.110***	0.056**	0.128***	0.100***
	(0.030)	(0.021)	(0.029)	(0.023)	(0.022)	(0.020)
4th income quintile ($t-1$)	0.136***	0.079***	0.171***	0.097***	0.173***	0.132***
	(0.033)	(0.023)	(0.031)	(0.024)	(0.026)	(0.022)
5th income quintile ($t-1$)	0.188***	0.106***	0.281***	0.153***	0.247***	0.198***
	(0.038)	(0.025)	(0.036)	(0.027)	(0.029)	(0.025)
Observations	40,189	40,189	40,189	40,189	40,189	40,189
Adjusted R-squared	0.135	0.469	0.121	0.357	0.361	0.424

Notes: Outcome ($t-1$) represents the standardised lagged value of the dependent variable. All regressions control for dummies indicating the child's gender, having a twin or being part of a triplet, having low birth-weight (< 2.5 kg), being a first-born, being in a single-parent household, and whether both parents are employed; dummies reflecting the child's ethnic background (mixed, Indian, Pakistani or Bangladeshi, Black, other), and the mother's highest educational level; mother's age at birth of the cohort member, parental involvement at age 3 (i.e. frequency of reading to the child, regular bedtime, hours spent in front of the TV), squared root of household size; dummies for house-tenure at time $t-1$ (ownership, mortgage, rent, other) and its variation between $t-1$ and t; survey wave dummies, and country dummies (England, Wales, Scotland, NI). Robust standard errors, clustered at the cohort member level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.2: THE EFFECT OF INCOME CHANGES ON CHILD HUMAN CAPITAL: VALUE ADDED MODELS WITH CHANNELS

	Externalising SDQ		Internalising SDQ		Reading test-scores	
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome $(t-1)$	0.644***	0.624***	0.518***	0.484***	0.287***	0.286***
	(0.007)	(0.007)	(0.008)	(0.008)	(0.006)	(0.006)
Gain	0.005	-0.001	0.020	0.010	0.041***	0.039***
	(0.012)	(0.012)	(0.014)	(0.013)	(0.012)	(0.012)
Loss	-0.033***	-0.022*	-0.039***	-0.026*	-0.035***	-0.034***
	(0.012)	(0.012)	(0.014)	(0.014)	(0.012)	(0.012)
2nd income quintile $(t-1)$	0.029	0.023	0.031	0.021	0.056***	0.055***
	(0.020)	(0.019)	(0.022)	(0.021)	(0.018)	(0.018)
3rd income quintile $(t-1)$	0.057***	0.042**	0.056**	0.033	0.100***	0.097***
	(0.021)	(0.021)	(0.023)	(0.023)	(0.020)	(0.020)
4th income quintile $(t-1)$	0.079***	0.058**	0.097***	0.065***	0.132***	0.128***
	(0.023)	(0.022)	(0.024)	(0.024)	(0.022)	(0.022)
5th income quintile $(t-1)$	0.106***	0.072***	0.153***	0.104***	0.198***	0.193***
	(0.025)	(0.025)	(0.027)	(0.026)	(0.025)	(0.025)
Δ(Kessler scale)$_{t-1,t}$		-0.115***		-0.153***		-0.020***
		(0.007)		(0.008)		(0.006)
Kessler scale $(t-1)$		-0.026***		-0.037***		-0.004**
		(0.002)		(0.002)		(0.002)
Δ(Life satisfaction)$_{t-1,t}$		0.028***		0.036***		0.002
		(0.006)		(0.007)		(0.006)
Life satisfaction $(t-1)$		0.013***		0.016***		-0.002
		(0.004)		(0.004)		(0.003)
Worsening in mother's health		-0.045**		-0.068***		-0.020
		(0.021)		(0.023)		(0.019)
Mother had poor health in $t-1$		0.006		-0.049***		-0.028**
		(0.015)		(0.017)		(0.014)
Observations	40,189	40,189	40,189	40,189	40,189	40,189
Adjusted R-squared	0.469	0.483	0.357	0.385	0.424	0.425

Notes: Outcome $(t-1)$ represents the standardised lagged value of the dependent variable. Δ(Kessler scale)$_{t-1,t}$ is the standardized difference of the mother's Kessler Psychological Distress Scale (K6) score between wave $t-1$ and t. Similarly, Δ(Life satisfaction)$_{t-1,t}$ is the standardized difference in the mother's life satisfaction between two consecutive waves and Life satisfaction $(t-1)$ is the level of her life satisfaction in wave $t-1$. Worsening in mother's health is a dummy equal 1 if there was a worsening in the self-reported mother's general health between wave $t-1$ and t. Mother had poor health in $t-1$ is a dummy equal 1 if the mother had either "fair" or "poor" self-reported general health in wave $t-1$. All regressions control for dummies indicating the child's gender, having a twin or being part of a triplet, having low birth-weight (<2.5 kg), being a first-born, being in a single-parent household, and whether both parents are employed; dummies reflecting the child's ethnic background (mixed, Indian, Pakistani or Bangladeshi, Black, other), and the mother's highest educational level; mother's age at birth of the cohort member, parental involvement at age 3 (i.e. frequency of reading to the child, regular bedtime, hours spent in front of the TV), squared root of household size; dummies for house-tenure at time $t-1$ (ownership, mortgage, rent, other) and its variation between $t-1$ and t; survey wave dummies, and country dummies (England, Wales, Scotland, NI). Sampling and non-response weights used. Robust standard errors, clustered at the cohort member level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.3: Robustness checks: Other life events as potential confounders

	Externalising SDQ		Internalising SDQ		Reading test-scores	
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome ($t-1$)	0.644***	0.644***	0.518***	0.517***	0.287***	0.287***
	(0.007)	(0.007)	(0.008)	(0.008)	(0.006)	(0.006)
Gain	0.005	0.005	0.020	0.019	0.041***	0.042***
	(0.012)	(0.013)	(0.014)	(0.014)	(0.012)	(0.012)
Loss	-0.033***	-0.029**	-0.039***	-0.033**	-0.035***	-0.039***
	(0.012)	(0.013)	(0.014)	(0.014)	(0.012)	(0.012)
Life events between $t-1$ and t						
Parent left		-0.029		-0.058*		0.016
		(0.027)		(0.033)		(0.023)
Mother lost job		0.002		0.021		0.019
		(0.022)		(0.026)		(0.022)
Father lost job		0.032		0.038		0.021
		(0.036)		(0.041)		(0.032)
Mother changed job		-0.004		0.020		-0.006
		(0.012)		(0.013)		(0.011)
Father changed job		0.006		-0.004		-0.011
		(0.012)		(0.013)		(0.011)
1 new sibling		-0.078***		-0.034**		0.002
		(0.016)		(0.017)		(0.015)
2+ new siblings		-0.144**		-0.114*		-0.049
		(0.062)		(0.068)		(0.056)
Any siblings left		-0.012		-0.062**		-0.016
		(0.027)		(0.030)		(0.022)
Observations	40,189	40,189	40,189	40,189	40,189	40,189
Adjusted R-squared	0.469	0.470	0.357	0.358	0.424	0.425

Notes: Outcome ($t-1$) represents the standardized dependent variable at $t-1$ for SDQ, while for Reading test-scores it is the quintile rank at $t-1$. Dummies indicating the child's family's income quintile in $t-1$ are controlled for in all columns. Additionally, all regressions control for dummies indicating the child's gender, having a twin or being part of a triplet, having low birth-weight (< 2.5 kg), being a first-born, being in a single-parent household, and whether both parents are employed; dummies reflecting the child's ethnic background (mixed, Indian, Pakistani or Bangladeshi, Black, other), and the mother's highest educational level; mother's age at birth of the cohort member, parental involvement at age 3 (i.e. frequency of reading to the child, regular bedtime, hours spent in front of the TV), squared root of household size; dummies for house-tenure at time $t-1$ (ownership, mortgage, rent, other) and its variation between $t-1$ and t; survey wave dummies, and country dummies (England, Wales, Scotland, NI). Sampling and non-response weights used. Robust standard errors, clustered at the cohort member level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.4: PERSISTENCE OF THE EFFECT OF GAINS AND LOSSES

	Externalising SDQ (1)	Internalising SDQ (2)	Reading test-scores (3)
Outcome_{t-2}	0.535***	0.451***	0.224***
	(0.009)	(0.012)	(0.007)
Gain_t	0.007	0.006	0.052***
	(0.020)	(0.022)	(0.017)
Loss_t	-0.038*	-0.057***	-0.055***
	(0.020)	(0.022)	(0.017)
Gain_{t-1}	0.034	0.021	0.042*
	(0.027)	(0.029)	(0.023)
Loss_{t-1}	0.003	-0.045	-0.021
	(0.026)	(0.029)	(0.023)
$\text{Gain}_t \times \text{Gain}_{t-1}$	0.000	0.004	-0.087**
	(0.044)	(0.046)	(0.038)
$\text{Gain}_t \times \text{Loss}_{t-1}$	-0.037	-0.004	0.000
	(0.037)	(0.044)	(0.033)
$\text{Loss}_t \times \text{Gain}_{t-1}$	0.021	0.020	-0.047
	(0.037)	(0.043)	(0.033)
$\text{Loss}_t \times \text{Loss}_{t-1}$	0.012	0.071	-0.087*
	(0.053)	(0.062)	(0.047)
Observations	25,377	25,377	25,377
Adjusted R-squared	0.362	0.253	0.462

Notes: Outcome_{t-2} represents the standardized dependent variable at $t-2$. All regressions control for income quintile dummies for waves $t-2$ and $t-1$. Additionally, all regressions control for dummies indicating the child's gender, having a twin or being part of a triplet, having low birth-weight (< 2.5 kg), being a first-born, being in a single-parent household, and whether both parents are employed; dummies reflecting the child's ethnic background (mixed, Indian, Pakistani or Bangladeshi, Black, other), and the mother's highest educational level; mother's age at birth of the cohort member, parental involvement at age 3 (i.e. frequency of reading to the child, regular bedtime, hours spent in front of the TV), squared root of household size; dummies for house-tenure at time $t-1$ (ownership, mortgage, rent, other) and its variation between $t-1$ and t; survey wave dummies, and country dummies (England, Wales, Scotland, NI). With respect to the tables above, the levels of time varying controls refer to wave $t-2$ and both the changes between $t-2$ and $t-1$, and between $t-1$ and t were controlled for. Sampling and non-response weights used. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.5: Transitions in and out of the outcomes bottom quintiles

	Externalising SDQ		Internalising SDQ		Reading test-scores	
	Entry	Exit	Entry	Exit	Entry	Exit
Gain	-0.002	-0.017	-0.015**	0.004	-0.014**	0.053***
	(0.005)	(0.016)	(0.006)	(0.016)	(0.006)	(0.017)
Loss	0.019***	0.037**	0.015**	0.009	0.007	-0.013
	(0.005)	(0.018)	(0.007)	(0.017)	(0.006)	(0.019)
Outcome quintiles in $t-1$ (reference: 5th quintile)						
2nd quintile	0.227***		0.214***		0.124***	
	(0.011)		(0.012)		(0.007)	
3rd quintile	0.127***		0.126***		0.093***	
	(0.012)		(0.012)		(0.007)	
4th quintile	0.067***		0.066***		0.050***	
	(0.013)		(0.013)		(0.008)	
Observations	31,955	8,230	30,925	9,256	32,603	7,584
Pseudo R-squared	0.147	0.147	0.0926	0.0926	0.103	0.103

Notes: The coefficients shown in the Table are average marginal effects derived from logistic regressions. The "Entry" and "Exit" columns represent respectively logistic regression where the outcome variable is the probability of moving in and out of the bottom quintile of the distribution of the outcome of reference. All regressions control for dummies indicating the child's gender, having a twin or being part of a triplet, having low birth-weight (< 2.5 kg), being a first-born, being in a single-parent household, and whether both parents are employed; dummies reflecting the child's ethnic background (mixed, Indian, Pakistani or Bangladeshi, Black, other), and the mother's highest educational level; mother's age at birth of the cohort member, parental involvement at age 3 (i.e. frequency of reading to the child, regular bedtime, hours spent in front of the TV), squared root of household size; dummies for house-tenure at time $t-1$ (ownership, mortgage, rent, other) and its variation between $t-1$ and t; survey wave dummies, and country dummies (England, Wales, Scotland, NI). Sampling and non-response weights used. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix 3.A: Additional Figures and Tables

Figure 3.A1: Transitions along the quintiles of Externalising SDQ

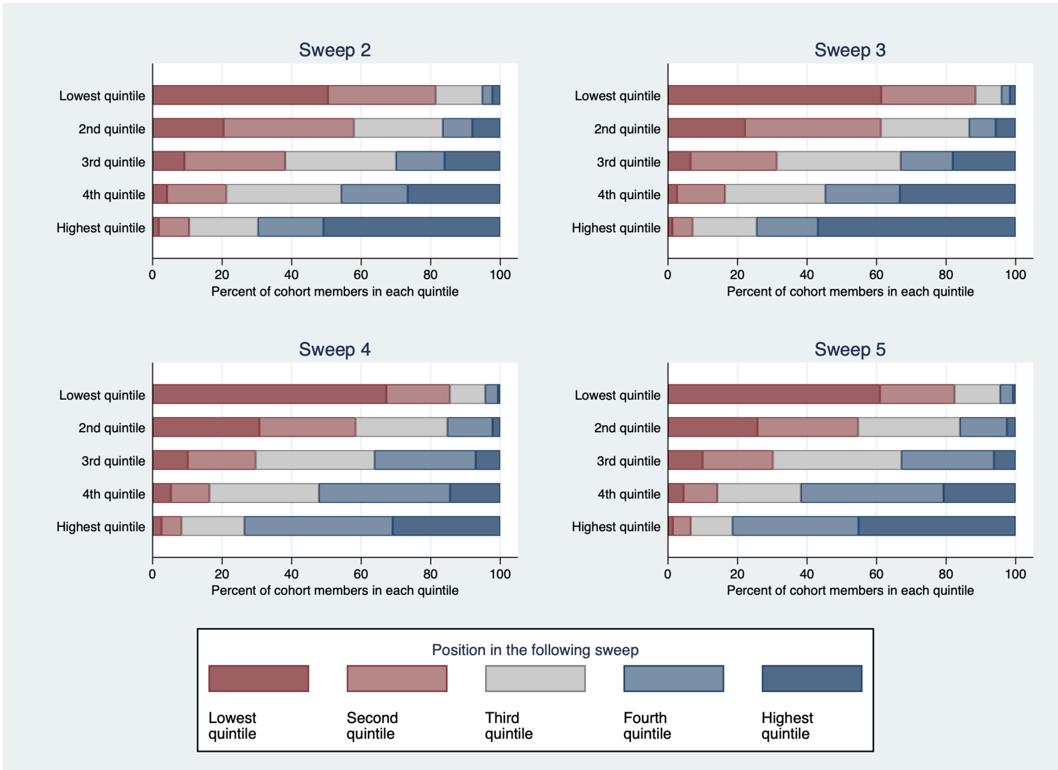


Figure 3.A2: Transitions along the quintiles of Internalising SDQ

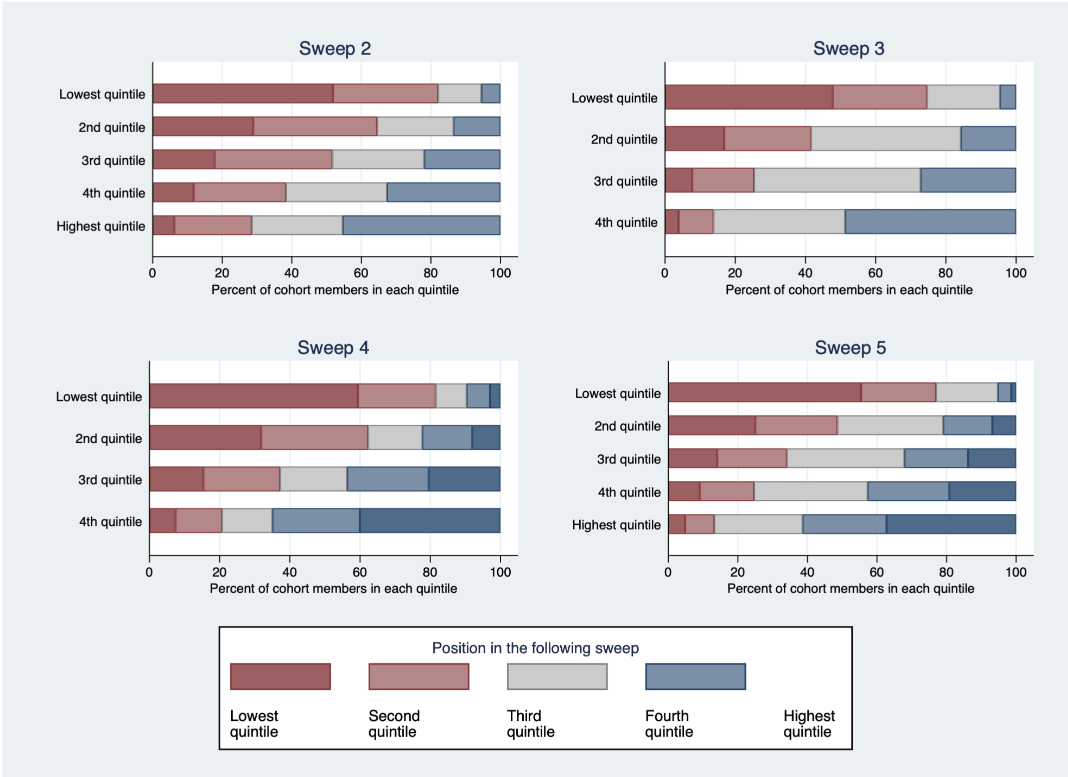


Figure 3.A3: Transitions along the quintiles of reading test-scores

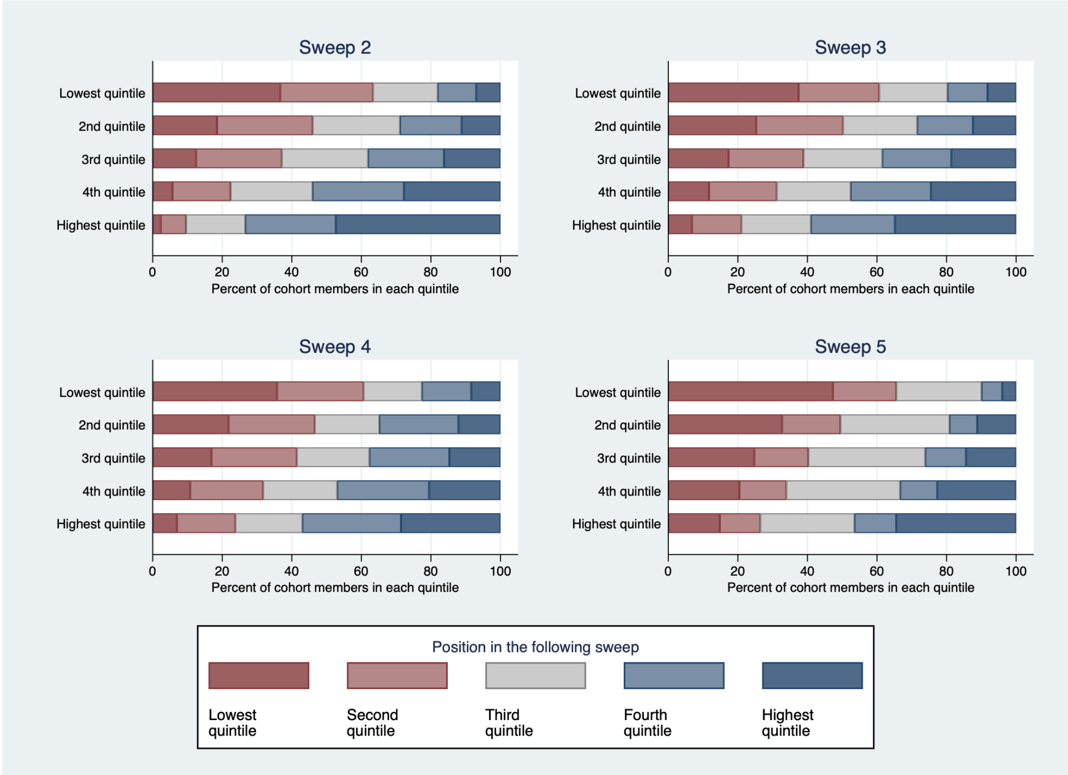


Figure 3.A4: Transitions along the quintiles of household income

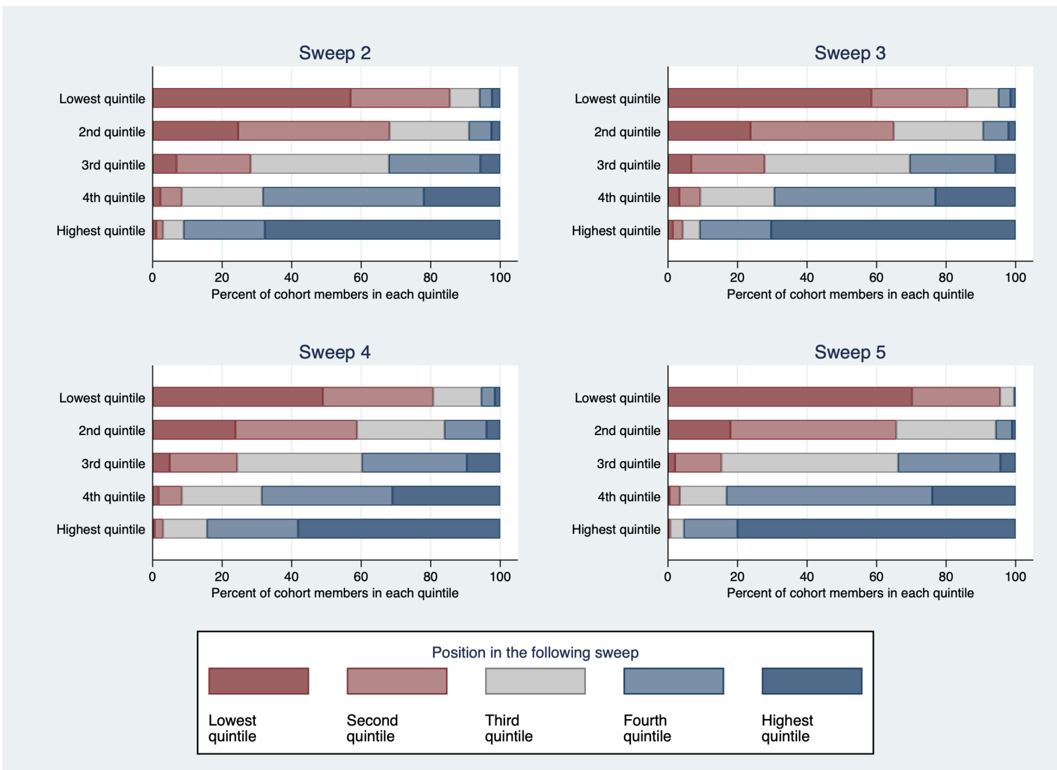


Table 3.A1: THE STRENGTHS AND DIFFICULTIES QUESTIONNAIRE

Please think about this child's behaviour over the last 6 months if you can:

This child:	NOT TRUE	SOMEWHAT TRUE	CERTAINLY TRUE
Emotional health:			
Often complains of headaches, stomachaches or sickness	0	1	2
Has many worries, often seems worried	0	1	2
Is often unhappy, down-hearted or tearful	0	1	2
Is nervous or clingy in new situations, easily loses confidence	0	1	2
Has many fears, is easily scared	0	1	2
Total emotional health score: 0-10			
Conduct problems:			
Has temper tantrums or hot tempers	0	1	2
Is generally obedient, usually does what adults request	2	1	0
Often fights with other children or bullies them	0	1	2
[a]Often lies or cheats	0	1	2
[b]Steals from home/school/elsewhere	0	1	2
Total conduct problems score: 0-10			
Hyperactity/Inattention:			
Is restless, overactive, cannot stay still for long	0	1	2
Constantly fidgets or squirms	0	1	2
Is easily distracted, concentration wandered	0	1	2
[c]Thinks things out before acting	2	1	0
Sees tasks through to the end, good attention span	2	1	0
Total hyperactivity score: 0-10			
Peer relationship problems:			
Is rather solitary, tends to play alone	0	1	2
Has at least one good friend	2	1	0
Is generally liked by other children	2	1	0
Is picked on or bullied by other children	0	1	2
Gets on better with adults than with other children	0	1	2
Total peer relationship problems score: 0-10			
Total internalising behaviour = emotional + peer relationship (0-20)			
Total externalising behaviour = behaviour + hyperactivity (0-20)			

[a] Changed to "Often argumentative with adults" in the questionnaire for 3-4 years old.

[b] Changed to "Can be spiteful to others" in the questionnaire for 3-4 years old.

[c] Changed to "Can stop and think things out before acting" in the questionnaire for 3-4 years old.

Table 3.A2: Descriptive statistics

Variables	Mean	SD	Min	Max
Outcomes				
Externalising SDQ	15.347	3.539	0	20
Internalising SDQ	17.029	2.991	1	20
Reading test-scores	54.669	11.904	20	80
Lagged outcomes				
Externalising SDQ	14.694	3.677	0	20
Internalising SDQ	17.206	2.702	1	20
Reading test-scores	55.261	11.150	20	80
Income changes				
Loss in income quintile between $t-1$ and t	0.218	.	0	1
Gain in income quintile between $t-1$ and t	0.271	.	0	1
Child characteristics				
Low birthweight (<2.5 kg)	0.069	.	0	1
First born	0.414	.	0	1
Twin or triplet	0.023	.	0	1
White	0.880	.	0	1
Mixed	0.035	.	0	1
Indian	0.017	.	0	1
Pakistani or Bangladeshi	0.029	.	0	1
Black	0.026	.	0	1
Other ethnicity	0.012	.	0	1
Female	0.496	.	0	1
Household characteristics				
Single parent	0.225	.	0	1
One working parent	0.352	.	0	1
Two working parents	0.508	.	0	1
Square root of household size	2.113	0.296	1.414	4
England	0.821	.	0	1
Wales	0.048	.	0	1
Scotland	0.090	.	0	1
Northern Ireland	0.040	.	0	1
Ownership ($t-1$)	0.052	.	0	1
Mortgage ($t-1$)	0.605	.	0	1
Rented ($t-1$)	0.318	.	0	1
Other ($t-1$)	0.025	.	0	1
No ownership/mortgage between $t-1$ and t	0.342	.	0	1
Lost house ownership between $t-1$ and t	0.025	.	0	1
Parental investment at age 3				
Up to one hour of TV per day	0.217	.	0	1
More than 1 hour of TV, less than 3 hours	0.623	.	0	1
More than 3 hours of TV per day	0.159	.	0	1
Regular bedtime	0.816	.	0	1
Read every day to the child	0.630	.	0	1

Read more than once per week, not every day	0.312	.	0	1
Read less than twice per month	0.058	.	0	1
Mother's characteristics				
Mother's age at birth	28.977	5.709	18	58
No educational qualifications	0.138	.	0	1
Less than O-level	0.019	.	0	1
GCSE or O-level	0.466	.	0	1
A-level or equivalent	0.099	.	0	1
Diploma of higher education	0.095	.	0	1
University degree or higher	0.182	.	0	1
Mother's well-being				
Δ(Kessler scale)$_{t-1,t}$	0.221	3.743	-24	24
Kessler scale $(t-1)$	3.430	3.825	0	24
Δ(Life satisfaction)$_{t-1,t}$	-0.125	2.030	-10	9
Life satisfaction $(t-1)$	7.599	1.866	1	10
Mother's health worsened	0.068	.	0	1
Mother has poor health $(t-1)$	0.143	.	0	1
Life events between $t-1$ and t				
One additional sibling	0.128	.	0	1
Two or more additional siblings	0.011	.	0	1
One or more siblings left household	0.052	.	0	1
No change in siblings composition	0.809	.	0	1
One parent left	0.063	.	0	1
Mother lost job	0.055	.	0	1
Father lost job	0.026	.	0	1

All descriptive statistics refer to the main estimation sample of 40,189 observations.

Table 3.A3: Estimates of income mobility on cognitive and non-cognitive outcomes from a Diagonal Reference Model

	Externalising SDQ (1)	Internalising SDQ (2)	Reading test-scores (3)
Outcome ($t-1$)	0.644***	0.517***	0.287***
	(0.007)	(0.008)	(0.006)
Gain	-0.017	-0.010	0.038*
	(0.017)	(0.024)	(0.021)
Loss	-0.011	-0.009	-0.032
	(0.018)	(0.025)	(0.021)
Estimated effects for the immobile			
Bottom income quintile	-0.055***	-0.073***	-0.098***
	(0.021)	(0.022)	(0.017)
2nd income quintile	-0.037***	-0.048***	-0.042***
	(0.014)	(0.016)	(0.013)
3rd income quintile	0.005	-0.004	0.004
	(0.012)	(0.016)	(0.012)
4th income quintile	0.037***	0.032**	0.035***
	(0.012)	(0.013)	(0.010)
Top income quintile	0.050***	0.094***	0.102***
	(0.013)	(0.013)	(0.013)
Weight for income in t	0.664*	0.620*	0.061
	(0.387)	(0.353)	(0.309)
Weight for income in $t-1$	0.336	0.380	0.939***
	(0.387)	(0.353)	(0.309)
Observations	40,189	40,189	40,189
AIC	91,452.44	99,253.61	92,443.78

Notes: The table displays maximum likelihood estimates for a diagonal reference model (Sobel, 1981), with the 'origin' variable being the family income quintile at time t and the 'destination' variable the income quintile in $t-1$. The dummies 'Gain' and 'Loss', defined as in the empirical strategy section, here can be interpreted, respectively, as indicators of upward and downward mobility. The table reports coefficients for the immobile categories, i.e. those individuals whose family income quintile does not change between $t-1$ and t (note that the sum of these coefficients is constrained to zero). Estimated weights are one the inverse of the other, and represent the relative importance of the origin vs destination variables. Outcome ($t-1$) represents the standardised lagged value of the dependent variable. All regressions control for dummies indicating the child's gender, having a twin or being part of a triplet, having low birth-weight (< 2.5 kg), being a first-born, being in a single-parent household, and whether both parents are employed; dummies reflecting the child's ethnic background (mixed, Indian, Pakistani or Bangladeshi, Black, other), and the mother's highest educational level; mother's age at birth of the cohort member, parental involvement at age 3 (i.e. frequency of reading to the child, regular bedtime, hours spent in front of the TV), squared root of household size; dummies for house-tenure at time $t-1$ (ownership, mortgage, rent, other) and its variation between $t-1$ and t; survey wave dummies, and country dummies (England, Wales, Scotland, NI). Sampling and non-response weights used in all columns. Robust standard errors, clustered at the cohort member level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.A4: Income changes and child human capital: Pooled OLS and GMM regressions

	Externalising SDQ			Internalising SDQ			Reading test-scores		
	OLS	GMM		OLS	GMM		OLS	GMM	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Outcome ($t-1$)	0.644***	0.301***	0.332***	0.518***	0.330***	0.331***	0.287***	0.104***	0.090***
	(0.007)	(0.014)	(0.014)	(0.008)	(0.014)	(0.013)	(0.006)	(0.010)	(0.009)
Gain	0.005	0.031**	0.025**	0.020	0.027**	0.026*	0.041***	0.066***	0.067***
	(0.012)	(0.012)	(0.012)	(0.014)	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)
Loss	-0.033***	-0.063***	-0.057***	-0.039***	-0.045***	-0.043***	-0.035***	-0.073***	-0.079***
	(0.012)	(0.012)	(0.012)	(0.014)	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)
2nd income quintile ($t-1$)	0.029	0.074***	0.076***	0.031	0.067***	0.069***	0.056***	0.121***	0.110***
	(0.020)	(0.020)	(0.019)	(0.022)	(0.021)	(0.021)	(0.018)	(0.018)	(0.018)
3rd income quintile ($t-1$)	0.057***	0.139***	0.145***	0.056**	0.117***	0.120***	0.100***	0.222***	0.221***
	(0.021)	(0.021)	(0.021)	(0.023)	(0.023)	(0.023)	(0.020)	(0.020)	(0.020)
4th income quintile ($t-1$)	0.079***	0.224***	0.222***	0.097***	0.180***	0.180***	0.132***	0.313***	0.312***
	(0.023)	(0.023)	(0.023)	(0.024)	(0.024)	(0.023)	(0.022)	(0.023)	(0.023)
5th income quintile ($t-1$)	0.106***	0.323***	0.316***	0.153***	0.255***	0.257***	0.198***	0.463***	0.467***
	(0.025)	(0.025)	(0.025)	(0.027)	(0.026)	(0.026)	(0.025)	(0.025)	(0.025)
Observations	40,189	40,189	40,189	40,189	40,189	40,189	40,189	40,189	40,189

Notes: Outcome ($t-1$) represents the standardised lagged value of the dependent variable. Columns 2, 5, and, 8 use a system GMM estimator where the lagged outcome is instrumented with its own lags of order two or greater. The remaining GMM columns additionally use the lags of order two or greater of the other two outcomes as instruments (e.g. Internalisng and Externalising SDQ in column 9). As the panel presents gaps, orthogonal deviations are used instead of first differences for all GMM estimations. All regressions control for dummies indicating the child's gender, having a twin or being part of a triplet, having low birth-weight (< 2.5 kg), being a first-born, being in a single-parent household, and whether both parents are employed; dummies reflecting the child's ethnic background (mixed, Indian, Pakistani or Bangladeshi, Black, other), and the mother's highest educational level; mother's age at birth of the cohort member, parental involvement at age 3 (i.e. frequency of reading to the child, regular bedtime, hours spent in front of the TV), squared root of household size; dummies for house-tenure at time $t-1$ (ownership, mortgage, rent, other) and its variation between $t-1$ and t; survey wave dummies, and country dummies (England, Wales, Scotland, NI). Sampling and non-response weights used only in columns (1), (4), and (7). Robust standard errors, clustered at the cohort member level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.A5: THE MEASUREMENT OF INCOME: RELAXING THE SCALE CONSTRAINTS OF THE INCOME QUINTILES DISTRIBUTION

	Externalising SDQ		Internalising SDQ		Reading test-scores	
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome $(t-1)$	0.652***	0.658***	0.518***	0.528***	0.286***	0.289***
	(0.008)	(0.009)	(0.009)	(0.010)	(0.007)	(0.008)
Gain	0.009	0.011	0.028*	0.031	0.025*	0.039**
	(0.014)	(0.018)	(0.015)	(0.019)	(0.014)	(0.018)
Loss	-0.041**	-0.048***	-0.041**	-0.044**	-0.028*	-0.051***
	(0.017)	(0.018)	(0.019)	(0.021)	(0.016)	(0.018)
3rd income quintile $(t-1)$	0.027	0.026	0.036*	0.032	0.037**	0.062***
	(0.017)	(0.019)	(0.020)	(0.021)	(0.016)	(0.018)
4th income quintile $(t-1)$	0.048**	0.051**	0.085***	0.097***	0.066***	0.090***
	(0.019)	(0.024)	(0.021)	(0.026)	(0.019)	(0.024)
Observations	25,326	19,948	25,326	19,948	25,326	19,948
Adjusted R-squared	0.459	0.458	0.341	0.349	0.406	0.412

Notes: Outcome $(t-1)$ represents the standardised lagged value of the dependent variable. All regressions control for dummies indicating the child's gender, having a twin or being part of a triplet, having low birth-weight (< 2.5 kg), being a first-born, being in a single-parent household, and whether both parents are employed; dummies reflecting the child's ethnic background (mixed, Indian, Pakistani or Bangladeshi, Black, other), and the mother's highest educational level; mother's age at birth of the cohort member, parental involvement at age 3 (i.e. frequency of reading to the child, regular bedtime, hours spent in front of the TV), squared root of household size; dummies for house-tenure at time $t-1$ (ownership, mortgage, rent, other) and its variation between $t-1$ and t; survey wave dummies, and country dummies (England, Wales, Scotland, NI). Sampling and non-response weights used in all columns. Robust standard errors, clustered at the cohort member level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.A6: The measurement of income: continuous income growth rate

	Externalising SDQ (1)	Internalising SDQ (2)	Reading test-scores (3)
Outcome ($t-1$)	0.644***	0.518***	0.287***
	(0.007)	(0.008)	(0.006)
Positive income growth$_{t-1,t}$	0.019**	0.019**	0.016**
	(0.008)	(0.008)	(0.008)
Negative income growth$_{t-1,t}$	-0.047	-0.063*	-0.079**
	(0.034)	(0.036)	(0.034)
Observations	39,722	39,722	39,722
Adjusted R-squared	0.469	0.357	0.423

Notes: Outcome ($t-1$) represents the standardised lagged value of the dependent variable. Positive income growth$_{t-1,t}$ is a continuous variable taking the positive values of the MCS imputed income growth rate between $t-1$ and t, and is set to zero for negative values. Similarly, Negative income growth$_{t-1,t}$ reflects the absolute value of negative income growth rates, and is set to zero for positive income growth. Note that the estimation sample here is smaller than the main one due to conditioning on the availability of the continuous measure of income and trimming values of income growth above 10 (around the top 0.5%). Dummies indicating the child's family's income quintile in $t-1$ are controlled for in all columns. Additionally, all regressions control for dummies indicating the child's gender, having a twin or being part of a triplet, having low birth-weight (< 2.5 kg), being a first-born, being in a single-parent household, and whether both parents are employed; dummies reflecting the child's ethnic background (mixed, Indian, Pakistani or Bangladeshi, Black, other), and the mother's highest educational level; mother's age at birth of the cohort member, parental involvement at age 3 (i.e. frequency of reading to the child, regular bedtime, hours spent in front of the TV), squared root of household size; dummies for house-tenure at time $t-1$ (ownership, mortgage, rent, other) and its variation between $t-1$ and t; survey wave dummies, and country dummies (England, Wales, Scotland, NI). Sampling and non-response weights used in all columns. Robust standard errors, clustered at the cohort member level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.A7: The measurement of income: gains and losses by continuous income growth rate

	Externalising SDQ (1)	Internalising SDQ (2)	Reading test-scores (3)
Outcome $(t-1)$	0.645***	0.518***	0.288***
	(0.007)	(0.008)	(0.006)
Gain			
$\geq 25\%$	0.008	0.018	0.035***
	(0.013)	(0.015)	(0.013)
$[10\%, 25\%)$	-0.006	0.015	0.074***
	(0.029)	(0.031)	(0.022)
$[5\%, 10\%)$	-0.045	0.048	-0.019
	(0.053)	(0.071)	(0.047)
$< 5\%$	0.107	0.102	0.089
	(0.082)	(0.082)	(0.063)
Loss			
$\leq -25\%$	-0.042***	-0.050**	-0.035**
	(0.016)	(0.019)	(0.016)
$(-25\%, -10\%]$	-0.013	-0.030	-0.024
	(0.020)	(0.021)	(0.018)
$(-10\%, -5\%]$	-0.042	-0.001	-0.080**
	(0.036)	(0.040)	(0.036)
$> -5\%$	-0.059	-0.054	-0.017
	(0.045)	(0.050)	(0.046)
Observations	39,825	39,825	39,825
Adjusted R-squared	0.469	0.357	0.423

Notes: Outcome $(t-1)$ represents the standardised lagged value of the dependent variable. The income quintile gain and loss dummies are here decomposed into a set of dummies based on the value of the continuous income growth rate between $t-1$ and t. The reference category (no change) here includes also small income changes (income growth rates between -5% and +5%). Note that the estimation sample here is smaller than the main one (loss of 364 observations) due to conditioning on the availability of the continuous measure of income. Dummies indicating the child's family's income quintile in $t-1$ are controlled for in all columns. Additionally, all regressions control for dummies indicating the child's gender, having a twin or being part of a triplet, having low birth-weight (< 2.5 kg), being a first-born, being in a single-parent household, and whether both parents are employed; dummies reflecting the child's ethnic background (mixed, Indian, Pakistani or Bangladeshi, Black, other), and the mother's highest educational level; mother's age at birth of the cohort member, parental involvement at age 3 (i.e. frequency of reading to the child, regular bedtime, hours spent in front of the TV), squared root of household size; dummies for house-tenure at time $t-1$ (ownership, mortgage, rent, other) and its variation between $t-1$ and t; survey wave dummies, and country dummies (England, Wales, Scotland, NI). Sampling and non-response weights used in all columns. Robust standard errors, clustered at the cohort member level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Chapter 4

Income and Wealth Volatility: Evidence from Italy and the U.S. in the Past Two Decades

Income and Wealth Volatility: Evidence from Italy and the U.S. in the Past Two Decades

with Edward N. Wolff (New York University),
and Conchita D'Ambrosio (University of Luxembourg).

Published in the *Journal of Economic Inequality*

4.1 Introduction

Income volatility has risen in a number of OECD countries in the recent past (Bartels and Bönke, 2010, 2013; Daly and Valletta, 2008; Dynan, Elmendorf and Sichel, 2012; Jappelli and Pistaferri, 2010). Most of the literature has focussed on the U.S., documenting a moderate to large increase in household income volatility from the 1970s to the 2000s, with a variety of different data sources and methods (DeBacker *et al.*, 2013; Dynan, Elmendorf and Sichel, 2012; Hacker and Jacobs, 2008; Hacker, 2019; Shin and Solon, 2011; Winship, 2009). The financial lives of individuals and households is increasingly subject to instability and unpredictability, due to changes in labour earnings, access to welfare, and family composition. However, less is known about individual wealth volatility in a comparative perspective, especially in relationship to income instability. Wealth inequality is known to be typically higher than income inequality in the U.S. (Conley and Glauber, 2008). These findings are confirmed by Fisher *et al.* (2016) and Johnson and Fisher (2020), who look at the relationship between inequality and mobility in income, wealth, and consumption for the same individuals. However, volatility is hardly ever addressed. One exception is the work of Whalley and Yue (2009), who use Chinese data to investigate rural income inequality and argue that higher income volatility exacerbates income inequality, as well as poverty concerns.

In this paper we investigate the relationship between income and wealth volatility in Italy and in the United States, the only two countries for which, to the best of our knowledge, data is available for more than a decade at the household and individual level. In order to do so, we adopt a range of descriptive measures typically used in the realm of income and earnings volatility, and apply them to both equivalised income and equivalised wealth following the tradition of the literature on the measurement of individual wellbeing. In particular, we apply a battery of variance decomposition methods developed by Gottschalk *et al.* (1994) and Moffitt and Gottschalk (2002, 2012), in order to disentangle a transitory component from a permanent component of the variances of income and wealth. We base our empirical analysis by calculating individual equivalised income and wealth from household longitudinal data from the United States (the Panel Study of Income Dynamics, from here onwards PSID) and Italy (Banca d'Italia's Survey on Household Income and Wealth, from here onwards SHIW). The panel dimension of these datasets, together with their focus on income processes and assets distribution, make them the perfect candidates for our purpose.

Several papers have documented an increase in household income volatility using PSID over the last few decades. While most studies unanimously report an increasing trend in income volatility in the U.S., there is no consensus on the magnitude of the effect – the differences between estimates being mostly driven by measurement issues and sample selection. Using the transitory component of the variance of log income, Hacker and Jacobs (2008) find that household income volatility doubled between the late 1960s and the early 2000s; Winship (2009), on the other hand, reports a more modest 30% increase in income volatility, measured as the standard deviation of the two-year percentage change of income, around the same period. See Dynan, Elmendorf and Sichel (2012) for a thorough review of studies on earnings and household income volatility in the U.S.. While wealth in PSID has been linked to macroeconomic volatility (Huang *et al.*, 2015; Heathcote and Perri, 2018; Stiglitz, 2012), little to nothing has been said on household wealth volatility over time. One exception is Conley and Glauber (2008), who measure wealth volatility in PSID as changes in average and median wealth across time and find that more than one-third of adults experience at least one $1,000 drop in their inflation-adjusted wealth before retirement.

With regard to Italy, Boeri, Brandolini and Rossi (2004) use SHIW to investigate whether the impoverishment of Italian households was partly attributable to higher income volatility. They perform their analysis at the individual level, using equivalised household income. Although referring to income volatility, the authors focus their attention on income mobility measures and report that mobility in Italy increased noticeably from the mid-1990s to the early 2000s. Diaz-Serrano (2005), using the SHIW rotatory panel from 1986 to 2000, estimates transitory shocks in labour income as the residuals from a Mincerian equation and finds a level of labour income uncertainty (measured as the variance of individual level residuals over time) of 0.264 across the whole sample of male earners. Using more recent waves of the panel component of the SHIW, Jappelli and Pistaferri (2010) highlight that in 2006 the variance of earnings per adult equivalent is almost 0.5, higher than the variance of raw earnings. As for the analysis of wealth, Brandolini *et al.* (2006) find a steady increase in wealth inequality in SHIW during the 1990s, mostly due to a larger concentration of financial wealth.

There are several papers that compare income and wealth variance (or inequality) on a theoretical level. One of the most notable is Deaton and Paxson (1994). They conclude that "Assets are the sum of previous saving and so will be an I(2) process, whose cross-sectional variance will therefore expand more rapidly than that of either consumption or disposable income" [p. 460]. However, the paper does not directly address the issue of whether wealth volatility (over time) is greater than income volatility.

De Nardi and Fella (2017) survey the savings mechanisms generated by a number of factors, concluding that the transmission of bequests and human capital, entrepreneurship, and medical-expense risk are crucial determinants of savings and wealth inequality. Their analysis begins with the basic Bewley (1977) model which features an incomplete market environment in which people save to self-insure against earnings shocks. In this framework, precautionary savings in the face of earnings risk is the key force driving wealth concentration. However, since the ability to self-insure improves as wealth increases relative to earnings, the model implies that the saving rate decreases with net worth relative to earnings, which is inconsistent with empirical findings. They then consider several other factors. The first is the intergenerational transmission of bequests and human capital. While introducing voluntary bequests generates more wealth concentration at the top, it also happens that, when calibrated using a standard earning process, this resultant economy has too many poor households. The second is heterogeneous preferences. However, when calibrated, this factor has limited success in generating realistic inequality throughout the entire wealth distribution. The third force is earnings dynamics. With the key assumption that that earning shocks are log-normally distributed, the main result is that the model predicts well the savings of the bottom 60% of the wealth distribution but does not generate the kind of saving behaviour at the top that is necessary to lead to a high concentration of wealth among the very rich. The fourth set of forces are medical expense risk and heterogeneity in life expectancy. On the basis of the well-established finding that richer people live significantly longer, the introduction of medical-expense risk and heterogeneous longevity into a model of savings after retirement helps match wealth by age and income quintile during retirement. The fifth force is idiosyncratic random shocks to the rate of return to wealth. This process is capable of generating a long right tail in the wealth distribution. However, the introduction of a bequest motive is found to be quantitatively more important than heterogeneous rates of return to generate the observed degree of wealth concentration. The sixth force is entrepreneurship, which is an important way to endogenize rates of return by explicitly modelling their production function. The survey shows that in a model with a simple life-cycle structure, entrepreneurship does generate a realistic level of wealth concentration.

More to the point is a paper by Heathcote and Perri (2018) as noted above. They indicate that between 2007 and 2013, U.S. households experienced a large decline in net worth. The main objective of this article is to study the macroeconomic implications of this decrease. However, they do find that during the Great Recession, wealth-poor households increased saving more than richer households, pointing towards the importance of the precautionary motive over this period. As they argue when wealth is high, the precautionary motive to save is weak but when wealth is low, the precautionary motive to save is strong. If precautionary savings played an important role during the Great Recession, then one should expect low wealth households to have reduced consumption especially sharply, since their precautionary savings should be most sensitive to increased risk. They do, in fact, find that low net worth households systematically increased savings rates by much more than high net worth households around the onset of the recession.

They find from the PSID that over the period 2006-2008, poor households reduced their consumption rate by about 4 percentage points more than rich households. One implication of these findings (which the authors do not directly specify) is that wealth volatility should be greater during economic downturns than normal times or booms, which is consistent with our findings reported below. However, as far as we are aware, there are no papers that directly investigate the issue of whether wealth volatility exceeds income volatility.

We analyse individual income and wealth volatility in PSID and SHIW across the years 2002 to 2014. We find that in both countries wealth volatility takes significantly higher values than income volatility. We then investigate the determinants of wealth volatility by exploring the dynamics of assets prices in Italy and the U.S., using data on rates of return to various components of wealth from the Jordà-Schularick-Taylor Macrohistory Database. In particular, we decompose the variance of wealth in order to disentangle the share of volatility that is due to changes in asset prices from a residual component, and find that changes in the market values of stocks and real estate assets drive most of the wealth volatility in our data. We also show that income and wealth volatility are higher in the United States and that the overall trend in both countries is increasing over time. Furthermore, we find evidence that the volatility of both income and wealth is higher during the years of the Great Recession, more so for the U.S. than for Italy. We conclude by exploring volatility in consumption and find that it predictably behaves in line with income volatility in both countries.

Our paper contributes to the literature in several ways. We are the first, to the best of our knowledge, to describe the evolution of income and wealth volatility for the same individuals over time. We do so adopting a comparative approach for two countries and a unified framework for each of the monetary variables. We further explore the channels that are likely to drive our findings and, finally, we exploit sources of heterogeneity across households in order to identify which groups are more vulnerable to income and wealth instability.

The remainder of the paper is organised as follows: Section 2 provides a review of the measures of volatility that will be used in the paper, while Section 3 describes the data. Section 4 presents results on income, wealth and consumption, and explores the role of rates of return in explaining wealth volatility. Heterogeneity analysis is conducted in Section 5. Finally, Section 6 concludes.

4.2 Measuring Volatility

A large strand of the literature has developed sophisticated econometric methods to estimate variance components models. However, as argued by Shin and Solon (2011), these methods rely on many assumptions and results are very sensitive to parametric specifications. This is one of the reasons behind the popularity of a simpler class of descriptive measures, developed by Gottschalk and Moffitt across the last few decades (see below for the references). Relying on the literature on permanent income and permanent wealth, we can think of the logarithm of each of our monetary variables (say log of income, y_it) as the following:

$$y_{it} = p_i + \epsilon_{it}$$

where p_i is a fixed permanent component with variance σ_p^2 (with mean zero and common across all individuals) and ϵ_{it} is a transitory component analogous to an idiosyncratic shock with variance $\sigma_{\epsilon t}^2$. The total variance of the observed monetary variable can be decomposed into:

$$\sigma_t^2 = \sigma_p^2 + \sigma_{\epsilon t}^2.$$

Based on this underlying model, we decompose the variance of income and wealth into a transitory and a permanent component. We here use two of the descriptive methods proposed in the literature: the first, which we call MG1, is based on Moffitt and Gottschalk (2002, 2012); the second, from here onwards MG2, was introduced by Gottschalk *et al.* (1994) and subsequently applied in Gottschalk and Moffitt (2009) and Moffitt and Gottschalk (2012), among others. See Chapter 6 in Jenkins (2011) for a thorough review of the econometric and descriptive methods for the estimation of the transitory variance.

The first method, MG1, offers a straightforward way of decomposing the variance. Given a long enough time interval s (based on data availability), it is possible to estimate the permanent component of the variance as the covariance between income (wealth) at time t and income (wealth) at time t-s. Subtracting the permanent variance to the observed variance yields an estimate of the transitory component of the variance.

On the other hand, MG2 uses a window averaging method: instead of considering a time period with respect to a number of lags, it requires the creation of a symmetric time window around a given year. Then individual averages are computed across that interval, which gives an estimate of the individual permanent income (wealth). The permanent variance is computed on the basis of deviations of the permanent income (wealth) from the sample average, while the transitory variance can be estimated as the average of individual variances of the difference between observed income (wealth) and permanent income (wealth).

We then take what Moffitt and Zhang (2018) refer to as a measure of 'gross volatility', i.e. the standard deviation of the individual differences in log income between one period and the next. Although using a measure of dispersion of income changes such as the standard deviation or the variance does not allow one to distinguish between a transitory and a permanent component, Shin and Solon (2011) argue that the standard deviation is less sensitive to calendar changes over time and that, under certain assumptions, it can provide less biased estimates of the transitory variance than MG1. Hence we conduct sensitivity analysis using the standard deviation of the two-year percentage changes in equivalent income and wealth as a measure of volatility. This measure is systematically used in the literature to analyse the dynamics and volatility individual earnings over time (see Dynan, Elmendorf and Sichel, 2012, for a review of the relevant literature and methodology).

4.3 Data Description

For our empirical application, we use individual panel data from the U.S. and Italy. For the first country we rely on the Panel Study of Income Dynamics (PSID), while for Italy we use data from the Banca d'Italia's Survey on Household Income and Wealth (SHIW).

The latter began in 1965, with microdata available from 1977 onwards. It currently covers a nationally representative sample of 8,000 Italian families (about 20,000 people), with a variety of information on economic and financial behaviour of individuals, both at the individual and family level. The SHIW was a repeated cross-section until 1989, when a randomly selected sub-sample of about 4,000 previously interviewed families was selected to be part of the panel component of the study. From 1989 onwards data were collected biannually (with the exception of a three-year gap between 1995 and 1998), with the latest available wave dating 2016. The year associated with the wave in SHIW is not the year when the interview took place, but the year to which all variables refer to. For example, the 2016 wave refers to income and wealth of year 2016, but was collected in 2017.

The PSID is a longitudinal study collecting measures of income and other socio-economic information for individuals living in the U.S.. It is currently the longest running panel study in the world: starting with a nationally representative sample of 18,000 individuals surveyed in 1968, the 2017-released 40th wave of the study covers around 26,000 people, of which 3,500 from the original sample. While income in PSID is collected both at the individual and family level, a wealth supplement is only available at the family level, for years 1984, 1989, 1994, and biannually from 1999 to 2017. For this reason we restrict the analysis to years 1999 to 2017. This choice is also consistent with the SHIW, as data are available biannually from 1998 to 2016 (after a 3-years discontinuity from 1995 to 1998). Since income - in PSID refers to the calendar year before the year in which the interview took place, the time period we consider goes from 1998 to 2016, with biannual observations for both countries. Note that measures of wealth in PSID are instead observed in the interview year. See Appendix 4.A for more information on income and wealth components in the two surveys.

In order to have a consistent time window and interview spells in the two countries, we focus on biannual observations from 1998 to 2016, with the caveat that wealth observations in the U.S. refer to the calendar year after.

Unlike in SHIW, measures of income in PSID are not net of taxes and transfers. Since 1992, PSID stopped providing estimates on federal income tax payments, making it impossible to directly derive a measure of net income from the available data for more recent years. However, the National Bureau of Economic Research made available the Internet TAXSIM program, a simulation tool aimed at calculating tax liabilities in the U.S. by assigning individuals to tax units and tax filing statuses (see Feenberg and Coutts, 1993, for a thorough description of the TAXSIM module). In particular, we rely on the method developed by Kimberlin, Kim and Shaefer (2014) to compute federal and state income taxes between year 1999 and 2011 and extend their procedure so to include subsequent years in our sample.[1]

Most of the literature looking at earnings or income volatility using PSID focuses only on earnings of the male household head. As we are interested in individual wellbeing, we prefer to keep the individual as our main unit of analysis without restricting our study to male earners only. However, both in PSID and SHIW, wealth is only available at the household level (contrary to income, which can be traced back to individuals). To reconcile an individual-based analysis with the data restrictions on wealth, we decided to attribute equivalised measures of household income and wealth to individuals older than 15. Although we consider a variety of equivalence scale parameters,

[1]The process is straightforward, since Federal laws from 1960 to 2023 and State laws from 1977 to 2016 are already coded within the program.

we here present our analysis using the square-root equivalence scale. Appendix 4.B discusses how the choice of the scale parameter affects the volatility measures we use throughout the paper.

As standard in the literature, we convert euros to dollars using PPP from the OECD data portal and deflate all monetary measures with 2010 constant prices. Additionally, we trim the top 1% and the bottom 1% of the observations in our samples (separately for each year in each of the two countries). We perform our analysis using a logarithmic transformation of equivalised income and an inverse hyperbolic sine transformation of equivalised wealth. The latter allows us to work with negative values without dropping them from the dataset, as would happen with a logarithmic transformation. We are therefore able to take an unrestricted sample for wealth, while for income we drop negative values and attribute the value 1 to zeros. Note that in our estimation sample, after trimming, we only have eight individuals in the U.S. with negative income and none in Italy. As for the zeroes, to which we attribute value 1, we have 414 cases in the U.S. (less than 0.5% of the American estimation sample) and again zero for Italy.

For MG1 we use 6-years lags to compute the permanent and transitory variance of income and wealth, whereas for MG2 we use time window averaging 5 years.[2]

Because of the longitudinal nature of the measures described above, we further restrict the sample to individuals who are observed in at least two waves before and one wave after the current interview. The final estimation sample spans from year 2002 to 2014, retrospectively and prospectively using information collected in years 1998, 2000, and 2016. It covers 11,458 individuals from Italy and 20,975 Americans, with non-missing information on income and wealth for at least three consecutive periods. Table 4.1 summarises the general characteristics of the estimation sample for the two countries.

4.4 Results

4.4.1 Income and Wealth

We here look at the evolution of the trends in the permanent and transitory component of the variance of income and wealth, as well as the standard deviation of individual changes from one period to the next.

Figure 4.1 shows the evolution of the two descriptive Moffitt and Gottschalk variance decom-

[2]We also use a 9-year centered window as a robustness check and find that results are qualitatively similar to the ones derived from the 5-year window MG2.

position methods (MG1 and MG2) for incomes in the U.S. and in Italy. For each of the methods used, the variance of income is decomposed into a transitory and into a permanent component, the sum of which gives the total variance. The two methods seem to describe similarly the evolution of the permanent component of income in both countries, while the same cannot be said for the transitory component. The latter appears to be less smooth when using MG1 – the more so for the U.S.. Regardless of the method used, Figure 4.1 shows that income volatility has been increasing since 2006, mostly due to increases in its transitory component, more steeply for the U.S. than for Italy. In the latter country, in fact, income volatility appears to be at most half of the U.S. levels.

To our knowledge, there are no other papers applying MG1 and MG2 to income data after the mid-2000s to which we can compare our results. Moffitt and Gottschalk (2012) use MG1 and MG2 to decompose the earnings variance of males aged 30 to 39, with PSID data, but their series ends in 2004, the period after ours begins. Their estimates of the transitory variance for the latest years in their sample have a magnitude of around 0.2, about half of the effect we find at the beginning of our series. The trend in the transitory component of MG1 also seems to mirror that in income inequality for the U.S., with a big jump over the Great Recession from 2006 to 2009, an abatement from 2009 to 2012, and then a strong upward trend from 2012 to 2015 (see Wolff, 2017, for income inequality trends based on the Survey of Consumer Finances). The permanent component of MG1 shows a similar pattern over time, though with smaller slopes (more attenuated changes). Both the permanent and transitory components of MG2 show a more or less continuous increase over time. When it comes to the Italian case, the trends in both MG1 and MG2 follow more of a U-shaped pattern, slowly increasing around the Great Recession. With regard to the magnitude of income volatility in Italy, in the early 2000s, we find our transitory income variance estimates to be in line with the estimates size of Diaz-Serrano (2005). The overall variance of log income, i.e. the sum of the permanent and transitory component of the MG1 variance, appears to be slightly lower than the figure of 0.45 found by Jappelli and Pistaferri (2010) using SHIW between 1995 and 2005. However, with respect to these authors, our sample is selected differently (e.g. we do not exclude retirees, whose stable pension income might partly explain our lower figures), and we use net equivalised income instead of earnings.

Figure 4.2 mirrors Figure 4.1, describing the evolution of wealth volatility in Italy and the U.S.. In both countries we can see that wealth volatility seems to have increased more steeply in concurrence of the Great Recession, more so in the U.S. than in Italy. Although, to the best of our knowledge, no other paper applies MG1 and MG2 to wealth, we can still draw a parallel between the evolution of the variance of wealth and wealth inequality in the U.S. and in Italy. Wealth inequality in the U.S. was flat from 2004 to 2007, spiked upward from 2007 to 2010, and then rose modestly after that (see Wolff, 2017). Here the transitory component of MG1 tracks well with this pattern from 2006 to 2012 but then shows a decline from 2012 to 2014. In contrast, the permanent component of MG1 as well as both the permanent and transitory components of MG2 shows a more or less continuous rise over the whole period. As for Italy, Dagnes, Filandri and Storti (2018) find a modest decline in wealth inequality between 2000 and 2004, followed by a steep upward trend peaking in 2012 and a sharp decline in 2014. These movements appear to be mirrored by both the permanent and transitory MG1 components of the variance of wealth in Italy, while the MG2 components show a flatter trend.

Two remarks can be made when comparing Figure 4.1 and Figure 4.2. First, wealth volatility appears to be strikingly higher than income volatility, irrespectively of the time period or the country considered. This is true not only for Moffitt and Gottschalk's descriptive measures, but also for the standard deviation of the two-year percent change in income and wealth (see Figure 4.C1 in Appendix 4.C). The second remark is a methodological one: consistent with Jenkins (2011), it appears that MG1 systematically overestimates the magnitude of the transitory component of the variances of both income and wealth with respect to MG2, whereas the contribution of the permanent component of the variances is robustly estimated across the two methods.

So far, our results suggest that wealth is more volatile than income, at least in the countries and years considered and the trend is increasing over time; in addition, both income and wealth volatility are much higher in the U.S. than in Italy.

4.4.2 Consumption

We now extend this exercise to include consumption volatility. The definition of consumption is very different in the two datasets undermining the comparability of the results by country. Still, we decided to report our findings and maximize the use of the available information. We did our best to harmonize the variables with only partial success. Similarly to the analysis for income and wealth volatility, we here attribute equivalent household consumption to each individual.

In the SHIW, there are already variables coded as "durable consumption" (DC) and "non-durable consumption" (NDC). We have to use them as they are, since it's not possible to break them down into their components in the dataset. In particular, DC is the total value of cars or other vehicles bought in the last calendar year, net of the value of cars and vehicles sold; furniture, furnishings, household appliances, sundry equipment. NDC is the value of food eaten at home and outside, utilities, holidays, clothing, education, leisure, medical expenses, rent.

PSID is more focused on expenditure rather than on consumption. We tried as much as possible to apply the SHIW definitions and arrive to comparable DC and NDC values. For DC we built a measure of the value of vehicles/cars owned net of vehicles/cars sold. Questions on furniture and household appliances were introduced only in 2005. For NDC we included food eaten at home and outside, utilities, repairs and maintenance of house and cars, transportation, holidays, clothing, education, leisure, medical expenses, childcare, rent. PSID collects information on insurance expenses (on house and vehicles), loans and car leases. We decided to neglect this since there is no equivalent in SHIW.

Following the literature on consumption inequality (see Jappelli and Pistaferri, 2010), we expect consumption volatility to be lower than income volatility: while the latter is subject to both transitory and permanent shocks, the former tends to be more stable, as transitory shocks can be typically smoothed out through the credit market, dissaving, and insurance to maintain a stable living standard. As consumption is theoretically expected to mostly reflect permanent shocks, we use the same variance decompositions methods used in the figures above to test empirically whether consumption volatility is mostly driven by its permanent component. Figure 4.3 shows results using MG2 with a 5-year moving average window (the other methods yield qualitatively and quantitatively similar results and the results are available upon request). The figure shows that, as predicted, when it comes to non-durable consumption, the permanent component of consumption volatility is higher than the transitory one in both countries. Looking at consumption volatility of durable goods, instead, we find the opposite results. This is still quite reasonable, as durable goods can be seen as less essential and easier to renounce in case of adverse economic conditions. What is more surprising is the net effect on total consumption volatility: while in Italy the narrative of non-durable goods seem to prevail, in the U.S. the transitory component of consumption volatility matters the most.

4.4.3 The Effects of the Rate of Return

On the surface, it seems surprising that wealth volatility is so much greater than income and consumption volatility, at least for the U.S.. The reason is that an individual's wealth in year t depends directly on the person's wealth in year $t-1$. The actual equation (for individual i) is:

$$W_{it} = W_{i(t-1)} + r_{it}W_{i(t-1)} + s_{it}Y_{it} + G_{it}.$$

where W_{it} is the net worth at time t, r is the rate of return on wealth, Y_{it} is income, s represents the savings rate out of income Y_{it}, and G_{it} is the net inheritances and gifts received. Changes in r, s, Y_{it}, or G_{it} could lead to volatility in wealth over time. However, it is unlikely that s or G_{it} varies too much over time (G_{it}, in any case, is relatively small). Y_{it}, on the other hand, does show some volatility over time, as is evident in Figure 4.5, though it is smaller than that in W_{it}. Perhaps, the most volatile component is the rate of return r. The rate of return faced by an individual over time depends on both the rate of return for individual assets and the portfolio composition of assets. The latter is relatively stable over time while rates of return on individual assets do show a great deal of variation (see Table 4.C1). To test whether the changes in asset prices do actually explain a significant portion of wealth volatility, we use rates of return for equity and housing to simulate how these individual wealth components would evolve if they perfectly followed the market rates of return of the antecedent period. In order to do so we use country-year data for Italy and the U.S. from the Jordà-Schularick-Taylor Macrohistory Database,[3] which collects a wide range of macroeconomic variables capturing, among others, asset price dynamics for 17 developed economies between years 1870 to 2016. We build two-year real return rates based on the annual nominal rates in the database and use them to compute the "explained" part of real estate and financial equity. This allows us to compute an individual "residual" component based on the difference between the actual value of real estate (equity), as reported in PSID and SHIW, and the explained value of real estate (equity), based on asset price changes. The decomposition of the levels of housing and equity into an explained and a residual component for the U.S. and Italy is illustrated respectively in Figures 4.C2 and 4.C3 of Appendix 4.C. The figures show that housing seems to be very closely predicted by asset price changes, while changes in equity from year to year are not explained quite as accurately. This is more so for Italy, partly because the only available measure of equity in SHIW also includes other financial instruments (such as bills and bonds) which are impossible to disentangle from financial equity alone. We then test our hypothesis that wealth volatility is in great part driven by the volatility of returns. Based on the aforementioned decomposition, we assess the contribution of each component (i.e. explained and residual) to the yearly variances of equity and real estate. We do so by following the Shorrocks (1982) decomposition of the variance of income into the contributions of its factor components. Figures 4.4 and 4.5 illustrate the results of this exercise. In both countries, asset price changes not only fully explain the variance of the value of real estate, but they tend to systematically overestimate it, such that the residual component of the variance is negative for almost all years. This appears to hold also for equity, although the relationship is less stable across years, especially for Italy (probably due to the measurement issue mentioned in the

[3] Accessed on July 23^{rd} 2019.

paragraph above). Furthermore, both in the U.S. and in Italy (albeit in the latter only for housing), the portion of the variance explained by asset price changes tends to reflect more closely the actual variance during the Great Recession with respect to other years.

4.5 Heterogeneity Analysis

In this section we analyze whether our results on income and wealth volatility are driven by particular groups of individuals in our samples. In order to do so, we use the transitory component of the variance derived from the MG1 method. This can in fact be interpreted not only as the difference between the cross-sectional variance of income and the covariance of current income and one of its past levels, but also as the covariance between current income and the difference between current income and one of its past levels (the same holds for wealth). Put more simply,

$$\sigma^2_{\epsilon t} = Var(y_t) - Cov(y_t, y_{(t-s)}) = Cov(y_t, y_t^*)$$

where $\sigma^2_{\epsilon t}$ is the MG1 transitory component of the variance and $y_t^* := y_t - y_{(t-s)}$, $s < t$. In order to assess the contribution of different groups of individuals to the transitory component of the variances of income and wealth, we decompose the latter by population sub-groups. Let G_1, G_2, ..., G_k be k groups such that every individual in a population of size N belongs to one (and only one) of the groups, with $k \leq N$. Let n_j be the size of group G_j and $\pi_j := \frac{n_j}{N}$ the corresponding population share. It is straightforward then to decompose the covariance between y_t and y_t^* as follows:

$$Cov(y_t, y_t^*) := \frac{1}{N}\sum_{i=1}^{N}(y_{it} - \overline{y}_t)(y_{it}^* - \overline{y}_t^*) = \sum_{j=1}^{k}\pi_j \left(\frac{1}{n_j}\sum_{i \in G_j}(y_{it} - \overline{y}_t)(y_{it}^* - \overline{y}_t^*)\right)$$

Where $\overline{y}_t$ and $\overline{y}_t^*$ are the population averages of y_t and y_t^* respectively. We apply this decomposition to households in our sample on the basis of available characteristics of the household itself and household heads. In particular, we use household head's relationship status (single or in a cohabiting relationship), gender, age, education, and age. In the U.S. we are also able to observe the racial group of the household head. We further decompose volatility on the basis of household size. Results are illustrated in Table 4.2. For each of the two countries, the table reports the percentage contribution of each sub-group to the overall volatility of income and wealth, as well as the number of households in each sub-group.[4] When it comes to relationship status, cohabitation seems to have

[4] Here the overall volatility of income and wealth is computed as the MG1 transitory component of the variance of income and wealth at the household level over the years 2002-2014. Volatility levels for the two

an insulating effect against income and wealth volatility in the U.S., while the opposite is true for Italy. Differences between the two countries appear also when looking at age of household head: while the share of income and wealth volatility is the highest for young individuals in the U.S., in Italy it is middle-aged household heads who appear to be the most vulnerable to income and wealth volatility. Furthermore, individuals in retirement age in the U.S. do not seem to account for a large share of volatility, which is not the case in Italy – especially for wealth. Female headed households are less subject to wealth volatility with respect to male headed and having a graduate or post-graduate degree also seems to have a dampening effect on both income and wealth volatility. Finally, in the U.S. most of income volatility can be attributed to households where the head is African-American. Heterogeneity results shown in Table 4.2 are robust to the use of other measures of volatility, such as the variance of the two-year difference in the natural logarithm of income or in the hyperbolic sine transformation of wealth. The reason we use the variance as a robustness check instead of the standard deviation is that the former is sub-group decomposable – what we need in order to disentangle the contribution of different groups of individuals to the sample volatility of income and wealth. Results for heterogeneity analysis using this other measure of volatility are available upon request.

4.6 Conclusion

In this paper we look at the recent trends in income, wealth, and consumption volatility in Italy and the U.S.. Income volatility is systematically lower than wealth volatility in both countries. All measures of income and wealth volatility appear to increase in the aftermath of the financial crisis, with wealth being the most affected, and their levels are always higher in the U.S. than in Italy. In particular, income volatility in Italy reaches at most half of the U.S. levels: while partly driven by higher earning inequality in the U.S., this result could also suggest that the system of tax and transfers in Italy is more efficient in protecting individuals against income shocks. Consistently with the literature, we also find that consumption volatility is lower than income volatility and substantially driven by permanent changes in consumption patterns, although the volatility of durable consumption shows a larger sensitivity to transitory fluctuations rather than permanent ones. We explain our findings on wealth volatility by looking at how changes in asset prices predict the evolution of individual wealth in our sample. We find that most of the fluctuations in housing and equity can be explained by changes in market return rates of these assets. Our results show that

countries are reported in the last row of Table 4.2.

individual wealth in Italy and the U.S. is highly sensitive to year-to-year fluctuations, largely more so than income. If wealth acts as a buffer to ensure consumption smoothing over time, protecting individuals against income shocks, then our results are indeed worrisome – especially in the light of the increasing trend in income volatility. While in this paper we offer an explanation based on changes in the price of real estate assets and stocks, other concurrent phenomena are likely to be in place as well. Conley and Glauber (2008) argue that a reason behind the increased wealth volatility in the U.S. could come from the liberalization of credit laws in 1978. In fact, between that period and 2004, there was an over 400 percent increase in personal bankruptcy filings in the U.S., most of which were due to unexpected medical expenses – with individuals covered by medical insurance being affected as well. The authors argue that other cases were potentially likely to be triggered by trends in demographic transition, such as increases in family dissolutions. We hope that this paper will contribute to the debate on income and wealth volatility and that it will stimulate further research on their interplay, as well as the mechanisms driving these two forces.

References

Bartels, Charlotte, and Timm Bönke. 2010. *German Male Income Volatility 1984 to 2008: Trends in Permanent and Transitory Income Components and the Role of the Welfare State.* SOEPpapers on Multidisciplinary Panel Data Research, 325.

Bartels, Charlotte, and Timm Bönke. 2013. "Can households and welfare states mitigate rising earnings instability?" *Review of Income and Wealth*, 59: 250–282.

Bewley, Truman. 1977. "The permanent income hypothesis: A theoretical formulation." *Journal of Economic Theory*, 16: 252–292.

Boeri, Tito, Andrea Brandolini, and Nicola Rossi. 2004. "The age of discontent: Italian households at the beginning of the decade." *Giornale degli Economisti e Annali di Economia*, 63: 449–490.

Brandolini, Andrea, Luigi Cannari, Giovanni d'Alessio, and Ivan Faiella. 2006. "Household wealth distribution in Italy in the 1990s." In *International Perspectives on Household Wealth.* Ed. E. N. Wolff, Chapter 7, 225–245. Edward Elgar Publishing.

Conley, Dalton, and Rebecca Glauber. 2008. *Wealth Mobility and Volatility in Black and White.* Center for American Progress. Retrieved from: www.cdn.americanprogress.org.

Cowell, Frank A, and Magda Mercader-Prats. 1999. "Equivalence scales and inequality." In *Handbook of Income Inequality Measurement.* Ed. J. Silber, Chapter 14, 405–435. Springer.

Dagnes, Joselle, Marianna Filandri, and Luca Storti. 2018. "Social class and wealth inequality in Italy over 20 years, 1993–2014." *Journal of Modern Italian Studies*, 23: 176–198.

Daly, Mary C., and Robert G. Valletta. 2008. "Cross-national trends in earnings inequality and instability." *Economics Letters*, 99: 215–219.

Deaton, Angus, and Christina Paxson. 1994. "Intertemporal choice and inequality." *Journal of Political Economy*, 102: 437–467.

DeBacker, Jason, Bradley Heim, Vasia Panousi, Shanthi Ramnath, and Ivan Vidangos. 2013. "Rising inequality: transitory or persistent? New evidence from a panel of US tax returns." *Brookings Papers on Economic Activity*, 2013: 67–142.

De Nardi, Mariacristina, and Giulio Fella. 2017. "Saving and wealth inequality." *Review of Economic Dynamics*, 26: 280–300.

Diaz-Serrano, Luis. 2005. "On the negative relationship between labor income uncertainty and homeownership: Risk-aversion vs. credit constraints." *Journal of Housing Economics*, 14: 109–126.

Dynan, Karen, Douglas Elmendorf, and Daniel Sichel. 2012. "The evolution of household income volatility." *BE Journal of Economic Analysis and Policy*, 12: 1–42.

Feenberg, Daniel, and Elisabeth Coutts. 1993. "An introduction to the TAXSIM model." *Journal of Policy Analysis and Management*, 12: 189–194.

Fisher, Jonathan D., David S. Johnson, Jonathan P. Latner, Timothy Smeeding, and Jeffrey Thompson. 2016. "Inequality and mobility using income, consumption, and wealth for the same individuals." *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 2: 44–58.

Gottschalk, Peter, and Robert A. Moffitt. 2009. "The rising instability of US earnings." *Journal of Economic Perspectives*, 23: 3–24.

Gottschalk, Peter, Robert A. Moffitt, Lawrence F. Katz, and William T. Dickens. 1994. "The growth of earnings instability in the US labor market." *Brookings Papers on Economic Activity*, 1994: 217–272.

Hacker, Jacob S. 2019. *The Great Risk Shift: The New Economic Insecurity and the Decline of the American Dream.* Oxford University Press.

Hacker, Jacob S., and Elisabeth Jacobs. 2008. *The rising instability of American family incomes, 1969-2004: Evidence from the Panel Study of Income Dynamics.* Economic Policy Institute Briefing Paper, 213.

Heathcote, Jonathan, and Fabrizio Perri. 2018. "Wealth and volatility." *Review of Economic Studies*, 85: 2173–2213.

Huang, Ho-Chuan, WenShwo Fang, Stephen M. Miller, and Chih-Chuan Yeh. 2015. "The effect of growth volatility on income inequality." *Economic Modelling*, 45: 212–222.

Jappelli, Tullio, and Luigi Pistaferri. 2010. "Does consumption inequality track income inequality in Italy?" *Review of Economic Dynamics*, 13: 133–153.

Jenkins, Stephen P. 2011. *Changing Fortunes: Income Mobility and Poverty Dynamics in Britain.* Oxford University Press.

Johnson, David S., and Jonathan D. Fisher. 2020. "Inequality and mobility over the past half century using income, consumption and wealth." In *Measuring and Understanding the Distribution and Intra/Inter-Generational Mobility of Income and Wealth.* Eds. R. Chetty, J. N. Friedman, J. Gornick, B. Johnson and A. Kennickell. University of Chicago Press.

Kimberlin, Sara, Jiyoun Kim, and Luke Shaefer. 2014. *An Updated Method for Calculating Income and Payroll Taxes from PSID Data Using the NBER's TAXSIM, for PSID Survey Years 1999 through 2011.* University of Michigan. Retrieved from: https://simba.isr.umich.edu/help/psid_taxsim.pdf.

Moffitt, Robert A., and Peter Gottschalk. 2002. "Trends in the transitory variance of earnings in the United States." *Economic Journal*, 112: C68–C73.

Moffitt, Robert A., and Peter Gottschalk. 2012. "Trends in the transitory variance of male earnings methods and evidence." *Journal of Human Resources*, 47: 204–236.

Moffitt, Robert A., and Sisi Zhang. 2018. "Income volatility and the PSID: Past research and new results." *AEA Papers and Proceedings*, 108: 277–80.

PSID. 2019. *Main Interview User Manual: Release 2019.* Institute for Social Research, University of Michigan.

Shin, Donggyun, and Gary Solon. 2011. "Trends in men's earnings volatility: What does the Panel Study of Income Dynamics show?" *Journal of Public Economics*, 95: 973–982.

Shorrocks, Anthony F. 1982. "Inequality decomposition by factor components." *Econometrica: Journal of the Econometric Society*, 193–211.

Stiglitz, Joseph E. 2012. "Macroeconomic fluctuations, inequality, and human development." *Journal of Human Development and Capabilities*, 13: 31–58.

Whalley, John, and Ximing Yue. 2009. "Rural income volatility and inequality in China." *CESifo Economic Studies*, 55: 648–668.

Winship, Scott Robert. 2009. "Has There Been a Great Risk Shift? Trends in Economic Instability Among Working-age Adults." PhD diss. Harvard University.

Wolff, Edward N. 2017. *Household Wealth Trends in the United States, 1962 to 2016: Has Middle Class Wealth Recovered?* NBER Working Paper Series, 24085.

Figures and Tables

Figure 4.1: Trends in Transitory and Permanent Income Variance

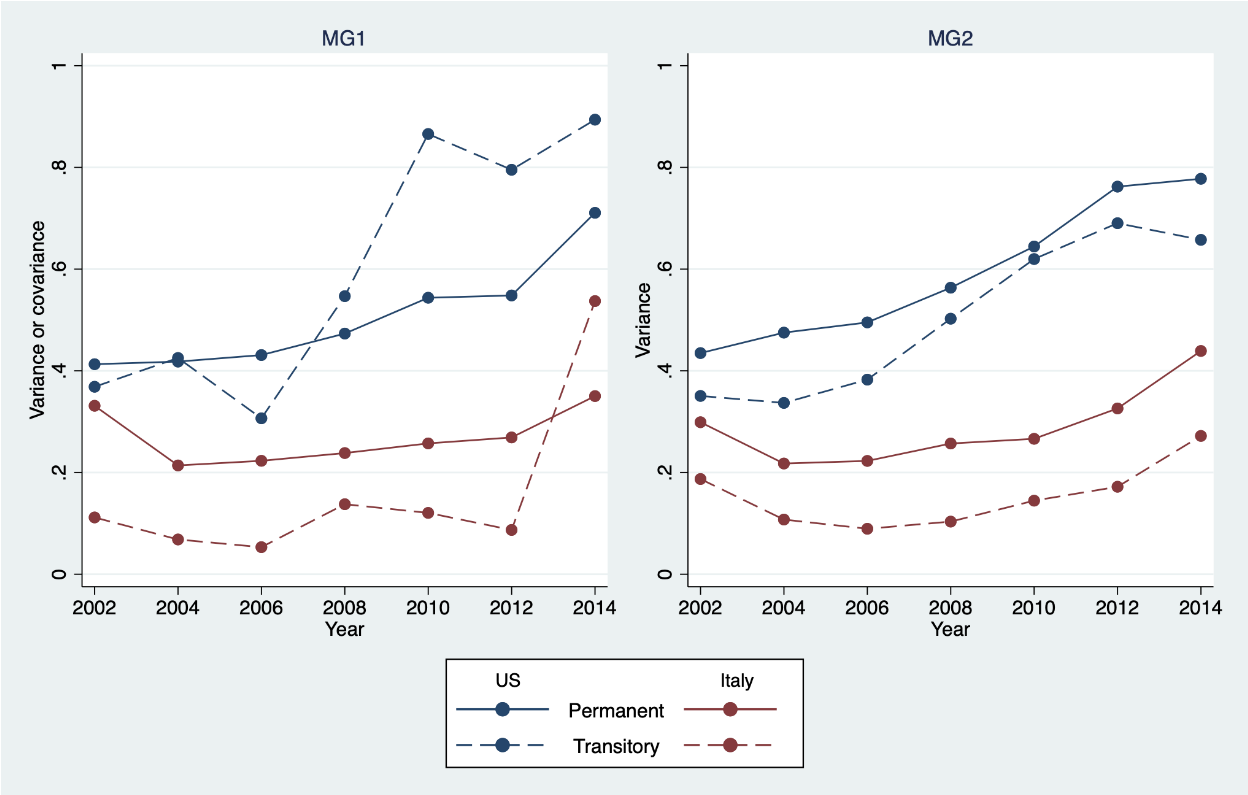


Figure 4.2: Trends in Transitory and Permanent Wealth Variance

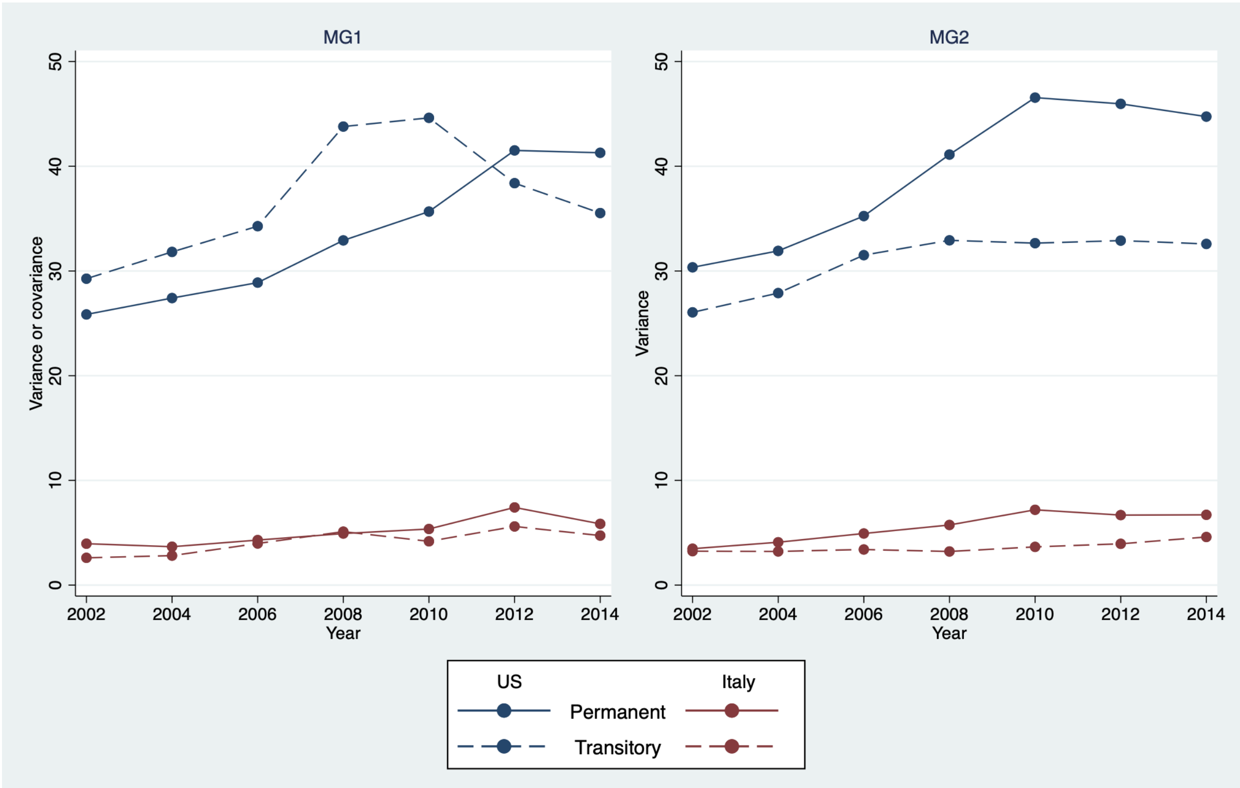


Figure 4.3: Trends in Consumption Variance

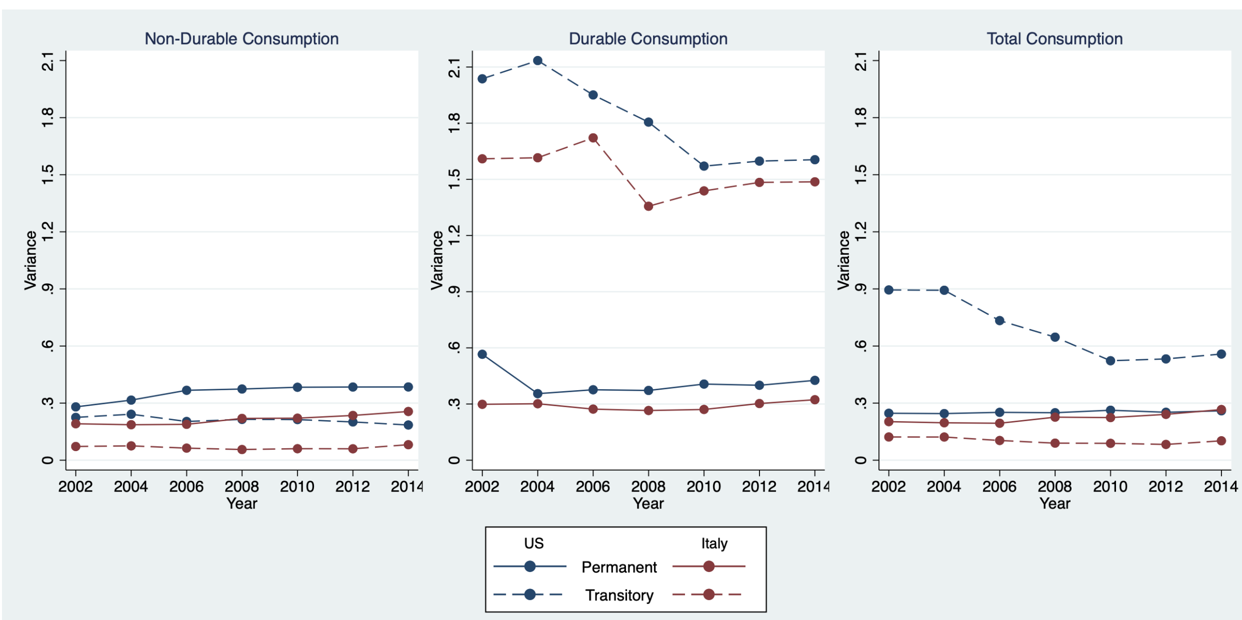


Figure 4.4: Variance Decomposition into its Explained and Residual Components (PSID)

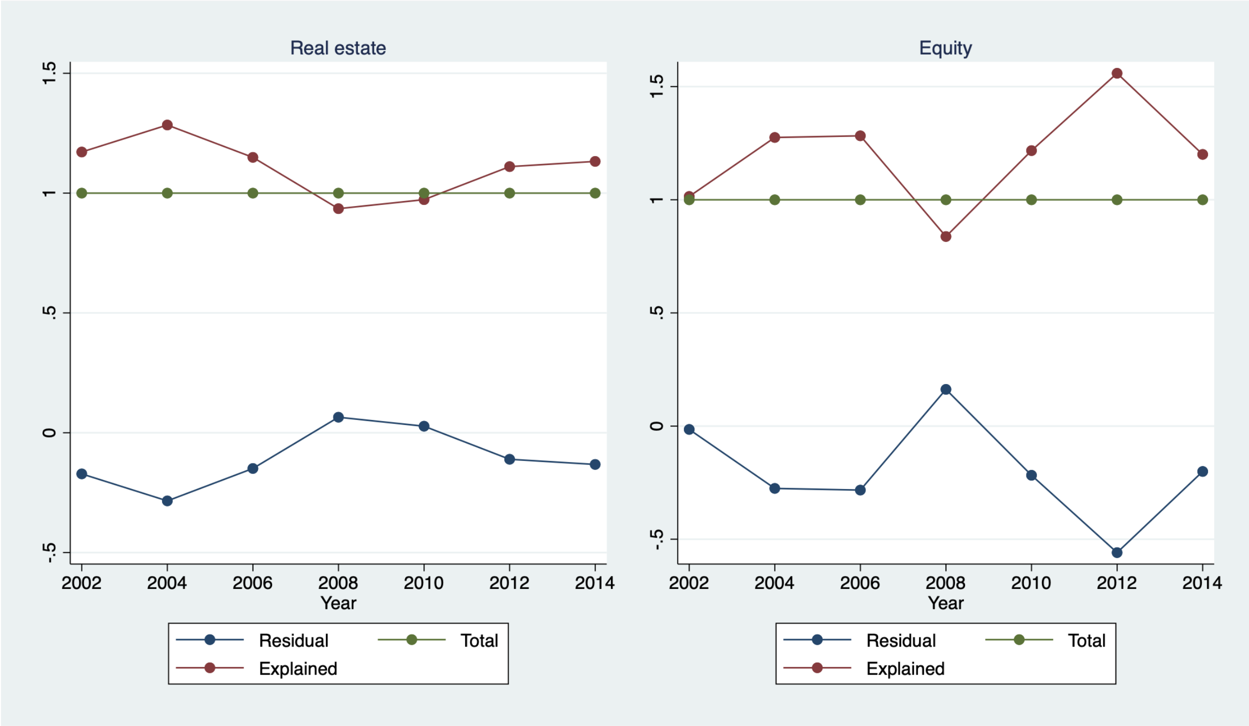


Figure 4.5: Variance Decomposition into its Explained and Residual Components (SHIW)

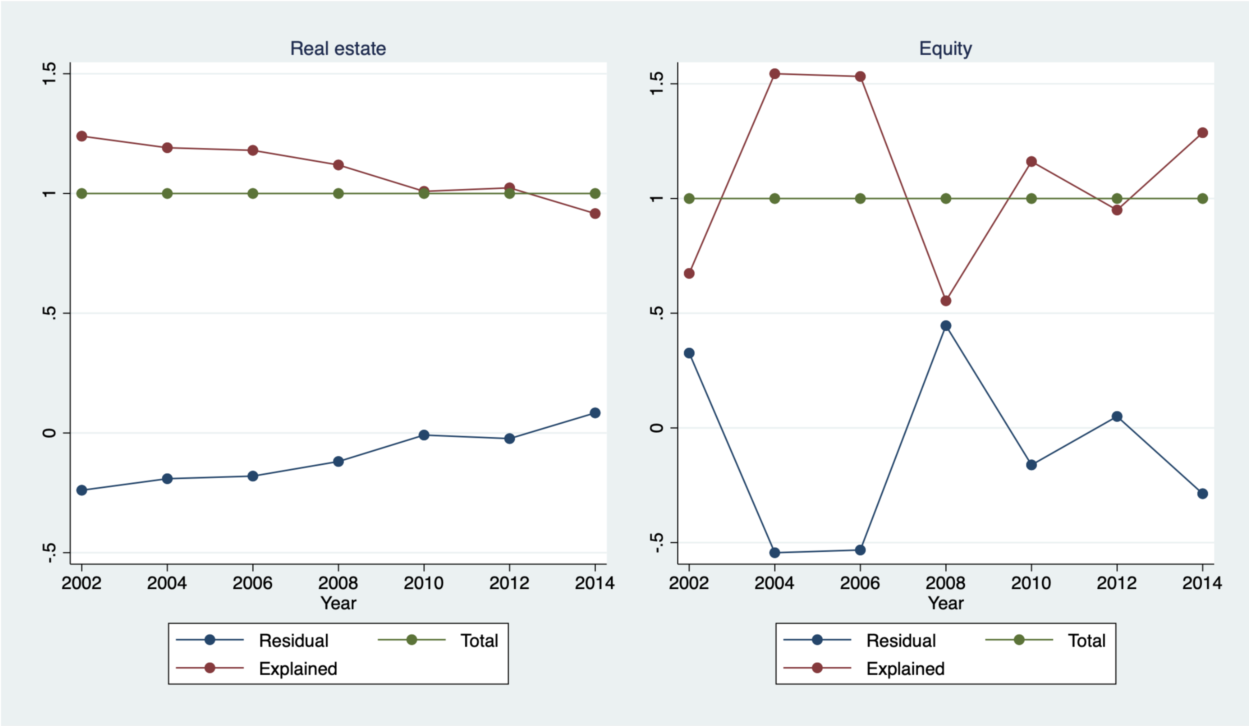


Table 4.1: Descriptive Statistics for the Estimation Sample

Variable	Obs.	Mean	Std. Dev.	Min	Max
Italy					
Income	32,804	26,639	13,855	1	128,601
Wealth	32,804	194,492	200,362	-6,800	2,250,115
Non-Durable Consumption	32,804	11,779	6,718	719	115,358
Durable Consumption	32,804	1,253	3,537	-20,578	75,607
Assets	32,804	185,571	192,404	0	2,017,615
Age	32,804	52	19.023	15	100
Female	32,804	0.52	0.500	0	1
Education	32,804	3.17	0.999	1	6
Household size	32,804	3.04	1.244	1	12
Single woman	32,804	0.15	0.354	0	1
Single man	32,804	0.06	0.238	0	1
U.S.					
Income	99,413	34,28	23,584	1	24,1567
Wealth	99,413	105,206	226,301	-130,000	3,345,062
Non-Durable Consumption	99,413	9,487	8,726	-255	749,833
Durable Consumption	87,455	29,525	126,056	-26,946	2,453,878
Assets	99,413	62,050	129,874	-329,156	2,758,257
Age	99,413	41	17.417	15	101
Female	99,413	0.54	0.499	0	1
Education	91,148	4.11	0.860	1	6
Household size	99,413	3.07	1.563	1	14
Single woman	99,413	0.26	0.437	0	1
Single man	99,413	0.10	0.298	0	1
Race head					
White	51,141	0.58	0.494	0	1
Black	51,141	0.36	0.481	0	1
Other	51,141	0.05	0.221	0	1
Race spouse					
White	25,901	0.71	0.454	0	1
Black	25,901	0.21	0.410	0	1
Other	25,901	0.07	0.248	0	1

Note: all monetary mesures are equivalised and expressed in 2010 constant dollars.

Table 4.2: Percent contribution to volatility, by group

	U.S.			Italy		
	Income	Wealth	N	Income	Wealth	N
Status						
Single	81.6	56.0	25,358	33.6	43.7	4,850
Cohabiting	18.4	44.0	25,783	66.4	56.3	10,003
Gender						
Women	47.0	36.2	17,104	51.9	36.2	5,069
Men	53.0	63.8	34,037	48.1	63.8	9,784
Age group						
15-34	42.5	47.5	15,617	11.7	4.3	351
35-44	16.9	21.0	10,366	24.9	16.8	1,735
45-54	21.1	17.0	10,685	37.8	30.8	3,119
55-64	11.0	9.6	7,730	13.0	21.4	3,492
65+	8.4	4.8	6,743	12.6	26.6	6,156
Education (highest degree)						
None	0.8	0.5	348	1.6	4.7	679
Primary	2.8	1.4	804	19.3	29.0	4,043
Lower secondary	24.9	7.6	5,408	54.2	42.1	5,093
Upper secondary	59.3	63.4	30,842	16.4	20.1	3,646
Graduate	7.6	17.2	7,331	8.1	4.0	1,301
Post-graduate	4.7	9.9	4,694	0.3	0.0	91
Household size						
One	51.9	30.1	13,423	27.3	27.0	3,201
Two	20.4	25.4	15,287	13.7	20.0	4,754
Three	11.4	18.5	9,067	19.8	18.7	3,121
Four or more	16.3	26.0	13,364	39.2	34.2	3,777
Race						
White	33.3	55.0	29,473	-	-	-
Black	62.7	39.5	18,619	-	-	-
Other	4.1	5.5	2,635	-	-	-
Total	0.93	43.42	51,141	0.18	4.63	14,853

Notes: The numbers in columns “Income” and “Wealth” indicate percentages. Each percentage expresses the relative contribution of each group to the overall income (wealth) volatility and is derived from the sub-group decomposition of the MG1 transitory component of the variance of income (wealth). Individual characteristics (e.g. age, gender) refer to household heads. The row Total reports total income and wealth volatility in the U.S. and in Italy, as well as the size of the sample of households in each country.

Appendix 4.A: Income and Wealth Components in PSID and SHIW

4.A1 Income components

In order to account for the economies of scale deriving household members sharing economic resources, we here attribute each individual a measure of equivalised income, using a square-root equivalence scale. In SHIW, family income is the sum of each household member's net disposable income. This can be derived as the sum of net labour income, transfer income, business and self-employment income, and asset income. Labour income includes both net salaries or wages and other forms of monetary compensation, as well as in-kind benefits (e.g. company car).[5] Transfer income includes pensions and other state transfers; from 1998 this category further includes transfers to and from non-cohabiting relatives or friends). Business and self-employment income is the sum of profits, dividends, and self-employed income, net of capital depreciation. Finally, asset income includes income sources deriving from real estate assets (i.e. rents) and financial assets (i.e. interests).

Similarly, family income in PSID is measured as the sum of each household member's taxable income, transfer income, and social security income. Taxable income is any income deriving from assets (i.e. interest, dividends, trust funds, rent), earnings (e.g. wage or salary, overtime, tips, commissions), and net profit from farm or business. Transfer income encompasses all transfers received by the family members, including pensions and annuity income. Note that the latter refers to defined contribution pension plans, such as the 401(k), and not to other forms of private annuities or IRAs, which are instead included in the computation of wealth. Missing values for all income components are imputed using overall median substitution by income source and recipient (for further details on the imputation method refer to PSID, 2019). Unlike in SHIW, however, labour income in PSID is measured before taxes. In order to ensure consistency between our measures of income across the two countries, we compute a measure of net family income in PSID by estimating federal and state income taxes with the TAXSIM module (Feenberg and Coutts, 1993).

[5] Note that housing is not included as a form of in-kind benefit.

4.A2 Wealth components

Both in PSID and in SHIW wealth is only available at the family level, which makes equivalisation necessary in order to perform an individual-level analysis. In SHIW, family wealth is computed as the sum of real and financial assets, net of financial liabilities. Real assets include houses, businesses, and land and buildings, as well as valuable objects (e.g. jewels, furniture). Financial assets are the sum of the value of checking and saving accounts; bonds; stocks, funds, and other financial instruments; credits towards relatives or friends. Lastly, debts towards banks or financial corporations, commercial debts, and debts towards other families make up financial liabilities. The PSID provides two aggregate measure of family wealth, one without and the other with housing equity. The latter is the sum of the values of seven asset types (farms and businesses; checking and saving accounts; other real estate; stocks; vehicles; other assets, e.g. bonds, funds, valuable collections; private annuities or IRAs), net of debt value (credit card debt; student loans; debt deriving from medical or legal expenses; loans from relatives), plus the value of home equity (measured as the self-assessed market value of the house, net of self-reported mortgage debt). Most wealth components overlap across the two countries. The main discrepancy is given by the attribution of vehicles: as these are not included in the definition of wealth in SHIW (figuring instead as durable consumption), we here use a measure of wealth in PSID which is net of the value of vehicles.

Appendix 4.B: Volatility and the Equivalence Scale Parameter

We here look at the sensitivity of our volatility measures to the choice of the equivalisation parameter α.[6] In Figures 4.B1 and 4.B2 we plot the relationship between income and wealth volatility and the parameter for the years in which all of our volatility measures were available, namely 2004, 2006, 2008, and 2010. Depending on the measure used and on the year considered, we find either a U-shaped or a negative relationship between income volatility and α. This is consistent with the considerations on income inequality and the equivalence scale parameter by Cowell and Mercader-Prats (1999), who find a similar U-shaped relationship using Spanish data. The measure that seems to be the most sensitive to the choice of α, especially in the U.S., is the transitory component of the income variance derived with the MG1 method. We find a flatter relationship instead for other measures, especially the standard deviation.

Figure 4.B2 shows a linear relationship (with negative slope) between wealth volatility and the equivalisation parameter. This comes as no surprise, since by construction there is an approximately linear relationship between the inverse hyperbolic sine transformation we used to rescale equivalised household wealth and the equivalisation parameter. Again, the measure that appears to be less sensitive to the choice of α is the standard deviation of the individual percentage changes in wealth between two consecutive periods (SD in the Figures).

[6] The parametric equivalence scale we use divides household income (wealth) by the number of household members raised to the parameter α, $\alpha \in [0, 1]$.

Figure 4.B1: INCOME VOLATILITY AND THE EQUIVALENCE SCALE PARAMETER

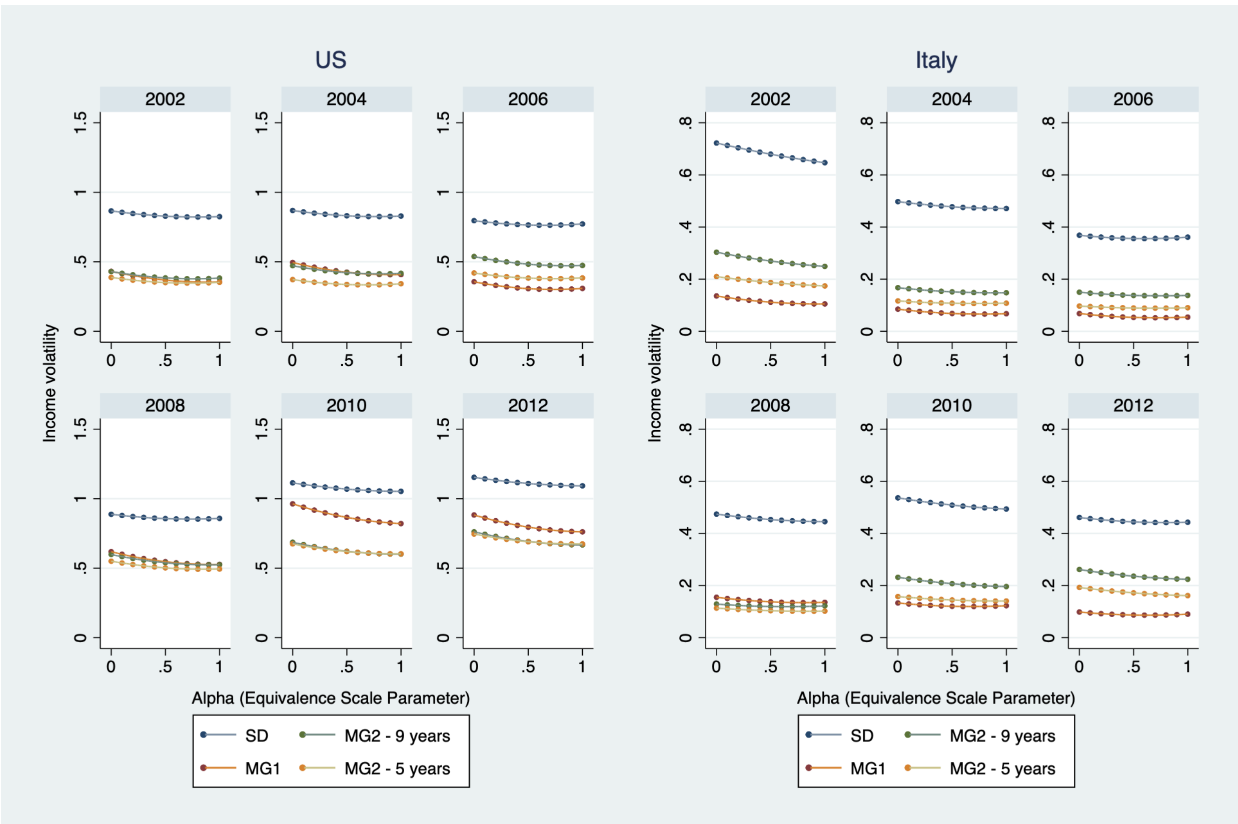
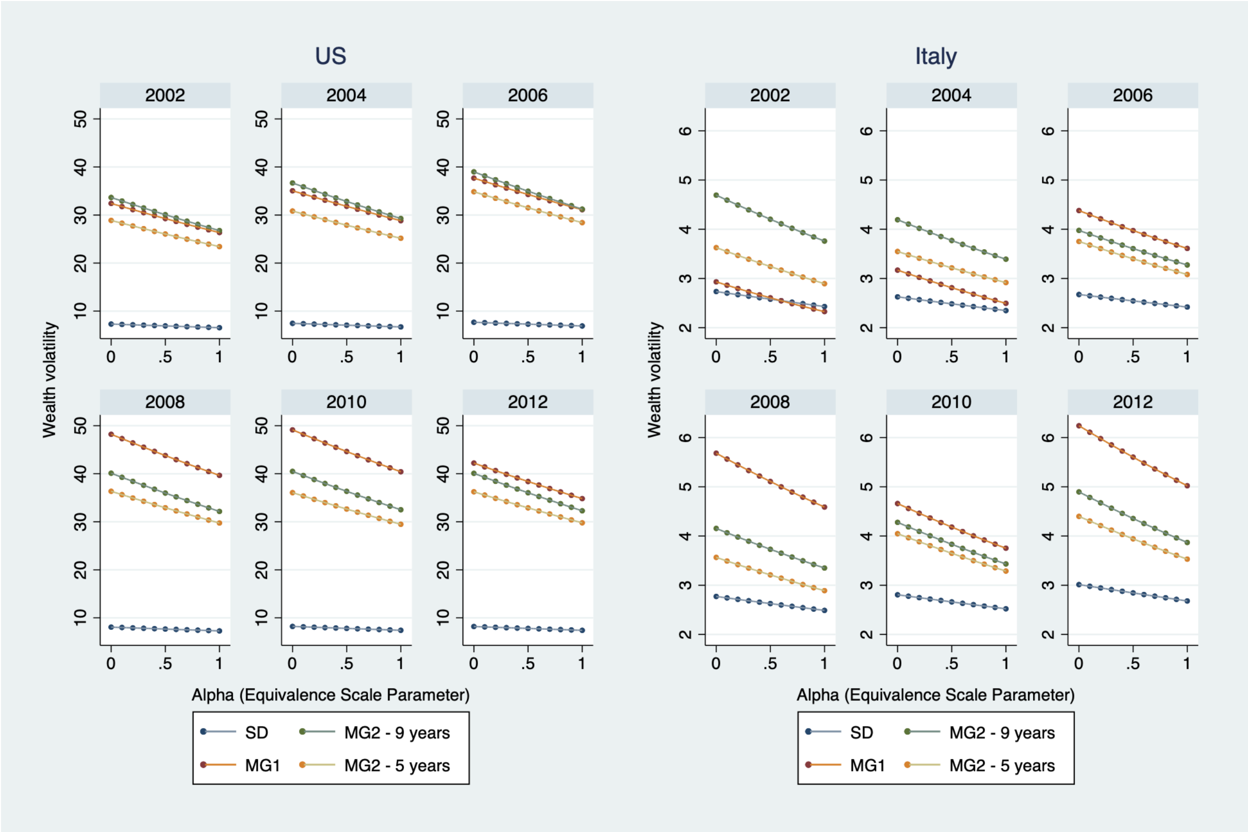


Figure 4.B2: Wealth volatility and the Equivalence Scale parameter



Appendix 4.C: Other Figures and Tables

Figure 4.C1: Income and Wealth volatility in Italy and the U.S.

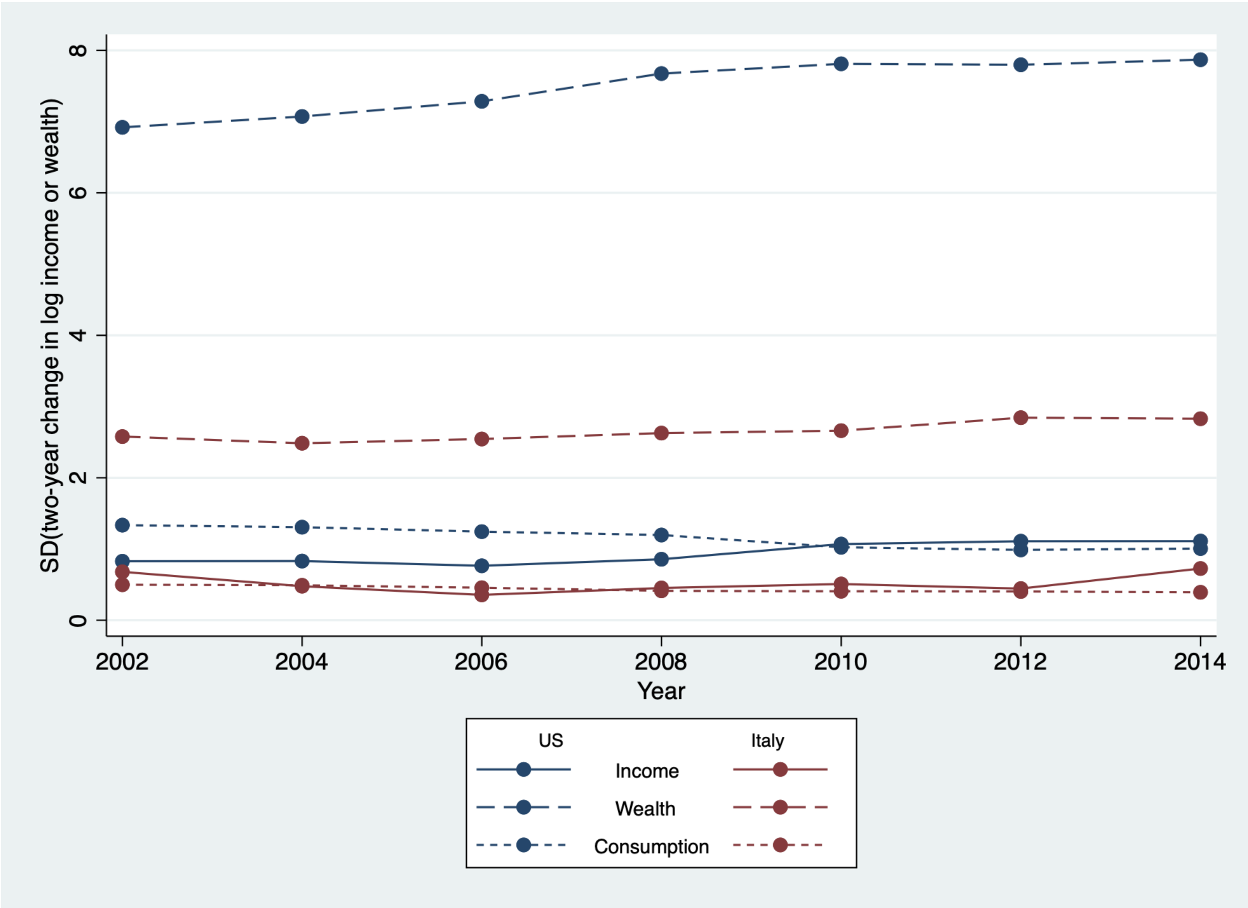


Figure 4.C2: Residual and Explained levels of Wealth in PSID, by wealth type

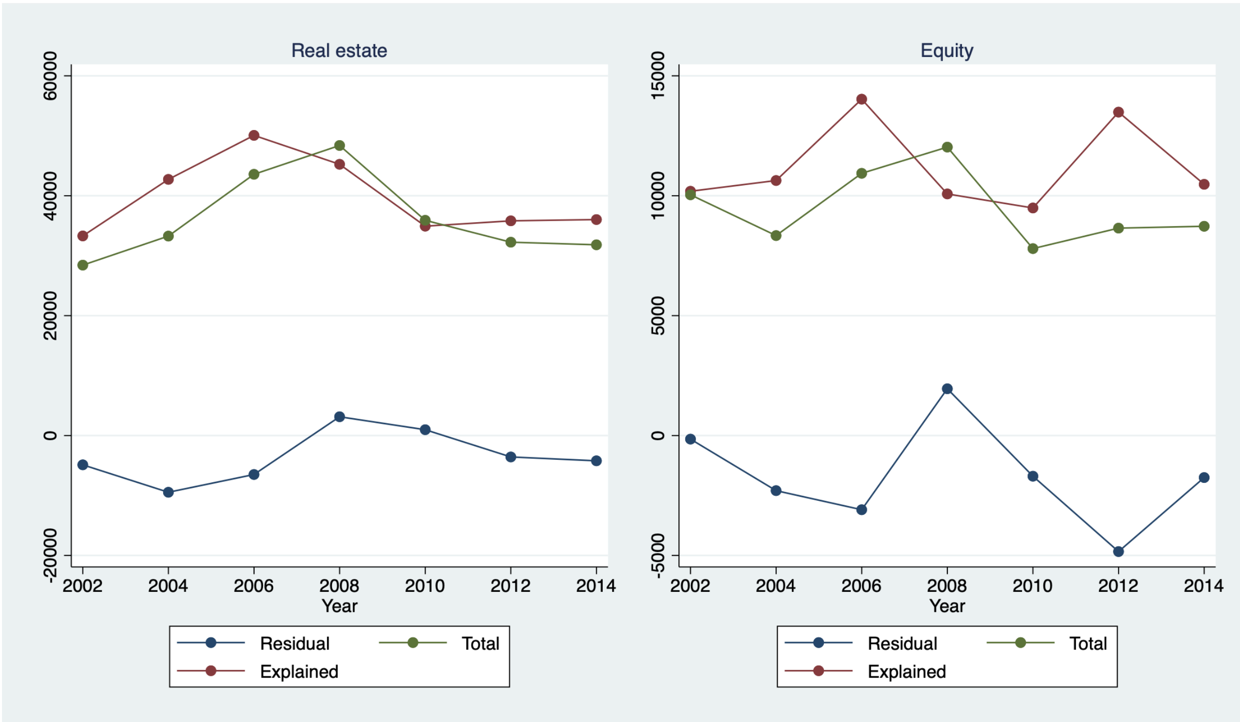


Figure 4.C3: Residual and Explained levels of Wealth in SHIW, by wealth type

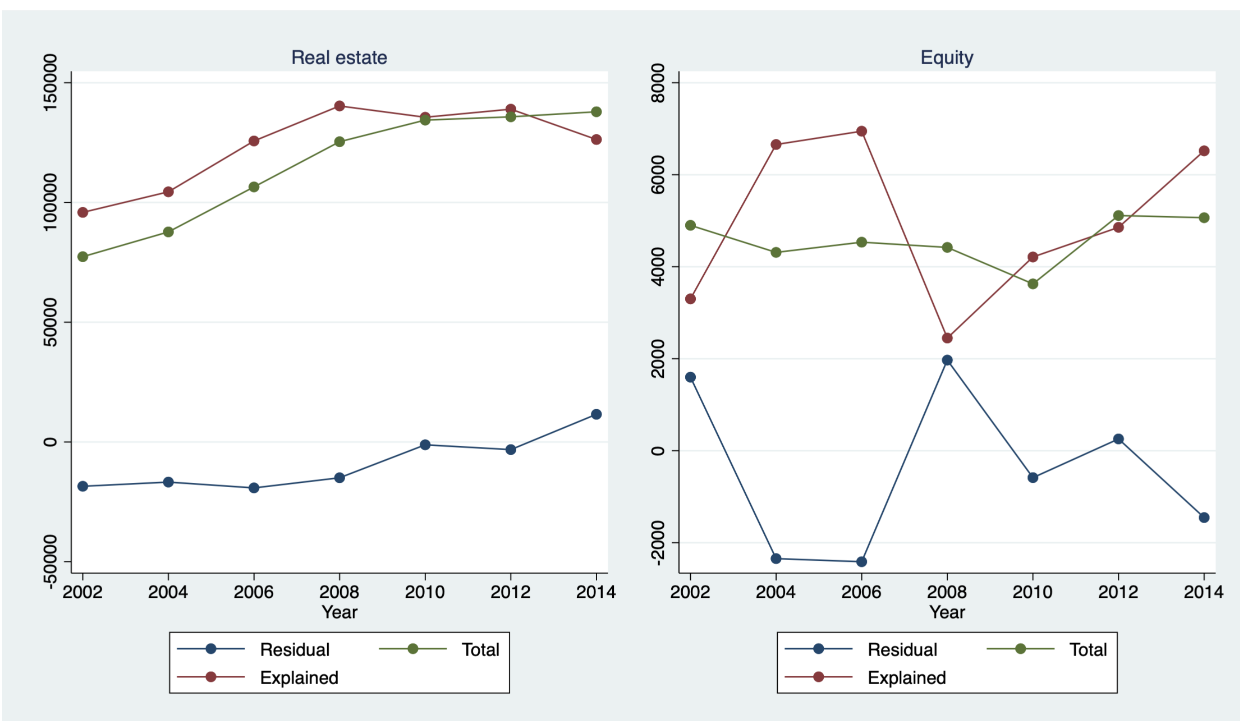


Table 4.C1: Percent changes in asset prices over time

Time span:	2002 - 2006	2006 - 2008	2008 - 2010	2010 - 2014
U.S.				
Equity	62.8	-14.8	17.9	92.3
Housing	46.9	-3.3	-6.2	29.4
Italy				
Equity	137.5	-44.8	14.8	21.5
Housing	41.2	11.8	-0.5	-6.9

Note: Return rates used to compute the percentage changes in the table are adjusted for inflation.

Future Developments and Research

Future Developments and Research

Social Science Genetics Research

Gene-Environment Interactions

While Chapter 1 used genetic data as a source of identification to assess the relationship between maternal depression and child human capital, the scope to apply genetic data into social science research is much wider, as outlined in the General Introduction of this manuscript.

One interesting avenue that I plan on exploring is that of gene-environment interactions, or $G \times E$. Drawing from the evidence described in the Introduction (Barcellos, Carvalho and Turley, 2018; Biroli and Zünd, 2020; Biroli and Zwyssig, 2021; Schmitz and Conley, 2016, 2017), I am interested in addressing the role of differences in genetic endowments in moderating the associations between exogenous changes in working and living conditions (that can be seen as changes in environmental exposures in the $G \times E$ framework) and individual behaviours and wellbeing. Using limited-access genetic data from the UK Household Longitudinal Study, I plan on exploiting policy discontinuities introduced by labour-market and pension reforms, such as the 2013 UK Enterprise and Regulatory Reform Act and the 1995 to 2011 Pensions Acts, to identify the interplay of exogenous environmental exposures and genetic predispositions (as proxied by polygenic scores) in shaping the health and behaviours of individuals. These quasi-natural experiments provide the ideal setting for identification strategies such as Difference-in-Differences or Regression Discontinuity Designs, as already shown by Cribb, Emmerson and Tetlow (2016) among others.

Using labour-market reforms to identify exogenous changes in environmental exposures is not always a panacea. They rarely offer the possibility to identify plausible control groups and their external validity is often limited. As a complement, I plan to investigate the impact of life events, such as job losses or obtaining a permanent work contract, on individual wellbeing and behavioural outcomes. Although these life events are easily observable in longitudinal survey data and occur with higher frequencies than do natural experiments, they do not come about at random so that endogeneity concerns need to be addressed. To do so, one possibility is to adopt a life-event study approach (Kleven *et al.*, 2019; Clark and Georgellis, 2013) and interact life-event variables with polygenic scores to account for genetic differences.

From the Environment to Genes: The Case of Epigenetics

While a person's DNA is fixed at conception, genetic expression is not. In a cell's nucleus, the physical disposition of DNA filaments can be altered by the surrounding architecture of proteins (called histones) and chemical bounds (typically, methyl groups). Modifications to these surrounding structures (the so-called 'epigenome') can alter gene expression, i.e. the production of RNA or proteins coded by a certain gene. Epigenetic changes such as DNA methylation and histone modification are mostly the result of environmental exposures and are, by definition, heritable (i.e., they can be transmitted from parents to children). Although knowledge about the precise biological mechanisms and the variety of circumstances inducing epigenetic changes in humans is still relatively scarce, a wide range of environmental factors has been found to be associated with changes in DNA methylation, including smoking, alcohol and diet, as well as stress and accidents (Alegría-Torres, Baccarelli and Bollati, 2011). A substantial literature has shown that DNA methylation in certain genetic regions can predict a shorter lifespan (Lin *et al.*, 2016; Lu *et al.*, 2019) and accelerated biological ageing (Horvath, 2013; Horvath and Raj, 2018; Hannum *et al.*, 2013; Levine *et al.*, 2018) – the latter being associated with the occurrence of diseases such as cancer (Horvath, 2013; Dugué *et al.*, 2018; Perna *et al.*, 2016), Alzhimer's disease (McCartney *et al.*, 2018), and ALS (Zhang *et al.*, 2020).

It is however less clear what is the effect of accelerated epigenetic ageing on more distal factors such as educational attainment, behavioural problems, or emotional health. In childhood, especially, being epigenetically older with respect to one's peers might exert a positive influence on early developmental outcomes (Simpkin *et al.*, 2017).

Thanks to the availability of epigenetic information for both mothers and children, the same dataset used in Chapter 1 (the Avon Longitudinal Study of Parents and Children, or ALSPAC) can be used to investigate the causes and consequences of early-life age acceleration at three time points (child ages 0, 7, and 15-17). DNA methylation profiles can additionally be identified based on trajectories of age acceleration and linked to later life outcomes, such as the likelihood of pursuing higher education, risky behaviours or fertility decisions.

Government Responses to COVID-19: Consequences on Income and Time-Use

From its onset in early 2020, the COVID-19 pandemic has affected the health, income, and general wellbeing of individuals all over the globe. Governments have had to come up with carefully balanced sets of measures and restrictions, with the objective of limiting the diffusion of the virus, while ensuring at least a certain degree of continuity of their countries' economic activity. The existence of cross country differences in the intensity and extent of containment and economic support policies over time is what motivates the following part of my research agenda, which aims at investigating how governments' policy responses to the COVID-19 emergency have affected individuals' incomes and the time they spend in paid and unpaid work.

In a descriptive paper using real-time longitudinal survey data for five European countries (the University of Luxembourg 'COME-HERE' dataset), I show that poverty increased by 1 percentage point on average from the beginning of the pandemic to September 2020 (Menta, Forthcoming). However, especially in the case of developed economies, poverty hardly captures the whole picture, as individuals can be severely affected by income losses without necessarily falling into standard definitions of income poverty. For this reason, I plan on focusing on the impact of policy responses to COVID-19 on household income losses, other than the likelihood of falling into poverty. Preliminary results show that more stringent confinement policies are associated with a higher risk of experiencing household income losses and transitioning into poverty. These effects are partly counterbalanced for high enough levels of economic support to individuals and households.

Chapter 2 of this thesis identifies the effect of an arguably exogenous increase in family size on the contribution to housework in childhood and, later on, in adulthood. Here, the underlying assumption is that an increase in family size generates the need for higher levels of home-production, as posited by intra-household decision models such as Blundell, Chiappori and Meghir (2005) and Cherchye, De Rock and Vermeulen (2012). The confinement measures following the onset of COVID-19 have also likely created an increase in the demand for housework, with both adults and children spending more time at home – due to job losses and remote working on one side and school closures on the other. The viability of outsourcing the production of housework to the market being compromised by confinement measures and arguably tighter budget constraints, it is reasonable to assume that a substantial share of the increase in housework has been absorbed by household members themselves. In order to investigate potential gender differences in the contribution to this higher housework load, I will use the COME-HERE dataset to describe the evolution of time spent in paid and unpaid work across gender, as well as its changes in response to more severe lockdowns. Preliminary results on a sample of partnered individuals with children suggest that, conditional on being employed and on the number of hours worked, more stringent lockdowns have translated into an equal increase in housework participation for men and women and into a disproportionately larger increase in childcare for women with respect to men.

References

Alegría-Torres, Jorge Alejandro, Andrea Baccarelli, and Valentina Bollati. 2011. "Epigenetics and lifestyle." *Epigenomics*, 3: 267–277.

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: E9765–E9772.

Biroli, Pietro, and Christian Lukas Zünd. 2020. *Genes, Pubs, and Drinks: Gene-Environment Interplay and Alcohol Licensing Policy in the UK.* Mimeo.

Biroli, Pietro, and Laura Zwyssig. 2021. *Moral Hazard Heterogeneity: Genes and Health Insurance Influence Smoking after a Health Shock.* IZA Discussion Paper, 14176.

Blundell, Richard, Pierre-André Chiappori, and Costas Meghir. 2005. "Collective labor supply with children." *Journal of Political Economy*, 113: 1277–1306.

Cherchye, Laurens, Bram De Rock, and Frederic Vermeulen. 2012. "Married with children: A collective labor supply model with detailed time use and intrahousehold expenditure information." *American Economic Review*, 102: 3377–3405.

Clark, Andrew E., and Yannis Georgellis. 2013. "Back to baseline in Britain: adaptation in the British household panel survey." *Economica*, 80: 496–512.

Cribb, Jonathan, Carl Emmerson, and Gemma Tetlow. 2016. "Signals matter? Large retirement responses to limited financial incentives." *Labour Economics*, 42: 203–212.

Dugué, Pierre-Antoine, Julie K. Bassett, JiHoon E. Joo, Chol-Hee Jung, Ee Ming Wong, Margarita Moreno-Betancur, Daniel Schmidt, Enes Makalic, Shuai Li, Gianluca Severi, *et al.* 2018. "DNA methylation-based biological aging and cancer risk and survival: Pooled analysis of seven prospective studies." *International Journal of Cancer*, 142: 1611–1619.

Hannum, Gregory, Justin Guinney, Ling Zhao, Li Zhang, Guy Hughes, SriniVas Sadda, Brandy Klotzle, Marina Bibikova, Jian-Bing Fan, Yuan Gao, *et al.* 2013. "Genome-wide methylation profiles reveal quantitative views of human aging rates." *Molecular Cell*, 49: 359–367.

Horvath, Steve. 2013. "DNA methylation age of human tissues and cell types." *Genome Biology*, 14: 1–20.

Horvath, Steve, and Kenneth Raj. 2018. "DNA methylation-based biomarkers and the epigenetic clock theory of ageing." *Nature Reviews Genetics*, 19: 371–384.

Kleven, Henrik, Camille Landais, Johanna Posch, Andreas Steinhauer, and Josef Zweimüller. 2019. "Child penalties across countries: Evidence and explanations." *AEA Papers and Proceedings*, 109: 122–126.

Levine, Morgan E., Ake T. Lu, Austin Quach, Brian H. Chen, Themistocles L. Assimes, Stefania Bandinelli, Lifang Hou, Andrea A. Baccarelli, James D. Stewart, Yun Li, *et al.* 2018. "An epigenetic biomarker of aging for lifespan and healthspan." *Aging (Albany NY)*, 10: 573–591.

Lin, Qiong, Carola I. Weidner, Ivan G. Costa, Riccardo E. Marioni, Marcelo R.P. Ferreira, Ian J. Deary, and Wolfgang Wagner. 2016. "DNA methylation levels at individual age-associated CpG sites can be indicative for life expectancy." *Aging (Albany NY)*, 8: 394–401.

Lu, Ake T., Austin Quach, James G. Wilson, Alex P. Reiner, Abraham Aviv, Kenneth Raj, Lifang Hou, Andrea A. Baccarelli, Yun Li, James D. Stewart, *et al.* 2019. "DNA methylation GrimAge strongly predicts lifespan and healthspan." *Aging (Albany NY)*, 11: 303–327.

McCartney, Daniel L., Anna J. Stevenson, Rosie M. Walker, Jude Gibson, Stewart W. Morris, Archie Campbell, Alison D. Murray, Heather C. Whalley, David J. Porteous, Andrew M. McIntosh, *et al.* 2018. "Investigating the relationship between DNA methylation age acceleration and risk factors for Alzheimer's disease." *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, 10: 429–437.

Menta, Giorgia. Forthcoming. "Poverty in the COVID-19 era: Real-time data analysis on five European countries." *Research on Economic Inequality*, 29.

Perna, Laura, Yan Zhang, Ute Mons, Bernd Holleczek, Kai-Uwe Saum, and Hermann Brenner. 2016. "Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort." *Clinical Epigenetics*, 8: 1–7.

Schmitz, Lauren L., and Dalton Conley. 2016. "The long-term consequences of Vietnam-era conscription and genotype on smoking behavior and health." *Behavior Genetics*, 46: 43–58.

Schmitz, Lauren L., and Dalton Conley. 2017. "The effect of Vietnam-era conscription and genetic potential for educational attainment on schooling outcomes." *Economics of Education Review*, 61: 85–97.

Simpkin, Andrew J., Laura D. Howe, Kate Tilling, Tom R. Gaunt, Oliver Lyttleton, Wendy L. McArdle, Susan M. Ring, Steve Horvath, George Davey Smith, and Caroline L. Relton. 2017. "The epigenetic clock and physical development during childhood and adolescence: Longitudinal analysis from a UK birth cohort." *International Journal of Epidemiology*, 46: 549–558.

Zhang, Ming, Paul M. McKeever, Zhengrui Xi, Danielle Moreno, Christine Sato, Tessa Bergsma, Philip McGoldrick, Julia Keith, Janice Robertson, Lorne Zinman, *et al.* 2020. "DNA methylation age acceleration is associated with ALS age of onset and survival." *Acta Neuropathologica*, 139: 943–946.

List of Figures

List of Tables