Essays on Social Mobility

Bertha Rohenkohl

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Sheffield

Faculty of Social Sciences

Department of Economics

December 2020

Para minha mãe, Cleonice.

This thesis is dedicated to my mother, Cleonice.

Abstract

This thesis comprises three empirical studies that explore contemporary topics within the area of social mobility. I investigate the mechanisms that contribute to the perpetuation of inequalities in the UK.

In the first empirical chapter, I examine the degree of intergenerational income mobility in the UK. Using new data from the harmonised British Household Panel Survey (BHPS) and Understanding Society (UKHLS) datasets for cohorts born between 1973 and 1992, I find evidence of substantial intergenerational persistence in the transmission of resources at the household level. I propose a novel methodology to estimate income mobility. The two-stage residual approach provides improved measures of long-run income for both generations. The intergenerational elasticities indicate that around one quarter of every additional percentage of income advantage enjoyed by parents is passed on to their children. The percentile rank coefficients, a measure of positional mobility, corroborate these findings. An increase of 10 points in the rank distribution of parents relates to an improvement of around 3 points in the ranks of their children. The strength of this association is an important indicator of how family background influences living standards in adulthood. My results are robust to changes in the specifications of the model, sample restrictions and to the use of alternative measures of income.

In the second empirical chapter, I unpack heterogeneities in intergenerational income mobility by three key dimensions, and present new facts about income mobility in the UK. First, I examine income mobility by gender, and I find that it is similar for sons and daughters. There is assortative mating by income, which reinforces the dynamics of income persistence due to family background. Second, I investigate variations in income mobility across the income distribution. Mobility is lower in the tails of the distribution of parental income. However, estimators of rank directional mobility reveal small rank movements even in these tails. Third, I explore geographical variations in absolute and relative mobility by region

of upbringing. There are stark differences across the UK, regions in the North of England display substantially lower levels of both relative and absolute income mobility than in the South. This chapter demonstrates the importance of considering these heterogeneities and of looking beyond aggregate measures that reflect average income mobility.

The final empirical chapter provides new insights on the role of expectations of facing future labour market discrimination on the educational achievement of ethnic minority pupils. Using unique data on discriminatory expectations from Next Steps for a cohort of students born in 1989-1990, I find that ethnic minority pupils who anticipate discrimination on ethnic grounds achieve better results in their GCSE exams at age 16. The evidence is consistent with the idea that discrimination can endogenously influence human capital investment decisions even prior to labour market entry. Variations in the magnitude and sign of this association indicate that there are heterogeneous effects across ethnic groups, with some groups showing greater effort in attempting to counteract the ethnic penalty that they expect to face in the labour market.

Taken together, the results presented in this thesis show that opportunities for social mobility in the UK are unevenly distributed across society. The implications drawn from this research provide valuable insights that need to be taken into consideration when planning policies to equalise opportunities.

Acknowledgements

I would like to express my sincere gratitude to my supervisors Andy Dickerson, Nicolas Van de Sijpe and Anita Ratcliffe for their invaluable comments, guidance and support throughout my PhD. They have helped me develop and advance my skills as an independent researcher.

In addition, I would like to thank my PhD examiners, Jo Blanden and Gurleen Popli, for their constructive and helpful comments, which have considerably improved my thesis. I would also like to thank all the academic and support staff within the Department of Economics for their continual support and help. I am particularly grateful to Mark Bryan and Arne Risa Hole for their comments and feedback in my confirmation review. The financial support received from the University of Sheffield is also gratefully acknowledged.

I am also extremely grateful to my fellow PhD students, who made the long days and nights in the office a lot more fun. A special thanks to Alessandro and to Emily for their friendship in good and bad times, and to Tina, Sam, Anna and Maria.

I would also like thank my wonderful friends from all around the world for their support and friendship, especially Daniel, Johanna, Olga, Hubertus, Julia, Raquel, Debora, and all my beloved Querendas. I am also very grateful to Mike for making the end of this journey so special.

Above all, I am always grateful to my family; my dear mother, Cleonice, and my stepfather Delauro have been my rock, encouraging me to start the PhD and persist in this journey. None of this would be possible without their never ending love and support.

Declaration

I, the author, confirm that this thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means. I hereby declare that except where specific reference is made to the work of others, the contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. An early version of Chapter 2 has been published online in the Sheffield Economic Research Paper Series (SERPS No. 2019017). This thesis contains fewer than 97,000 words including appendices, bibliography, footnotes, tables and equations.

Chapters 2 and 3 of this thesis use data from the British Household Panel Survey (BHPS) and Understanding Society (UK Household Longitudinal Study), produced by the Institute for Social and Economic Research (ISER) and supplied by the UK Data Service. Chapter 4 is based on data from Next Steps, which has been produced by the Department for Education (DfE) and supplied by the Secure Data Service (SDS) at the UK Data Service. The use of this data does not imply the endorsement of the data owner or the UK Data Service at the UK Data Archive in relation to the interpretation or analysis of the data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

Bertha Rohenkohl

December 2020

Table of contents

List of figures **xv**

List of tables **xvii**

1 Introduction **1**

1.1 Background and Motivation 1

1.2 Aims, Research Questions and Contributions 8

1.3 Structure and Content of this Thesis 9

1.3.1 Overview of Chapter 2 10

1.3.2 Overview of Chapter 3 11

1.3.3 Overview of Chapter 4 12

2 Intergenerational Income Mobility in the UK **13**

2.1 Introduction 13

2.2 Intergenerational Income Mobility: Background and Measurement 17

2.2.1 Measuring intergenerational mobility 21

2.2.2 Current evidence: The UK in an international perspective 27

2.3 Data and Methodology 36

2.3.1 Data 36

2.3.2 Key variables 38

2.3.3 The matched sample 40

2.3.4 Descriptive statistics 41

2.3.5 Estimation strategy 42
2.4 Results 50
2.4.1 Estimates of intergenerational income mobility in the UK 50
2.4.2 Robustness analysis 58
2.5 Discussion and Conclusion 68

Appendix A 73

A.1 Sample representativeness 73
A.2 Income variables 75
A.3 TSRA: Bootstrapped standard errors 77
A.4 Supplementary results 78

3 Unpacking Intergenerational Income Mobility in the UK 83

3.1 Introduction 83
3.2 Background and Related Literature 87
3.2.1 Intergenerational income mobility among daughters and sons 87
3.2.2 What kind of mobility are we talking about? 90
3.2.3 The UK postcode lottery 95
3.3 Data 101
3.4 Estimation and Results 103
3.4.1 Intergenerational income mobility by gender 103
3.4.2 Income mobility at different points of the income distribution . . . 112
3.4.3 Regional variations in income mobility 126
3.5 Conclusion 137

Appendix B **141**

B.1 Supplementary figures 141

B.2 Supplementary results 145

4 Anticipated Discrimination and Educational Achievement **147**

4.1 Introduction 147

4.2 Theoretical framework: Labour market discrimination 151

4.3 Background: Ethnic Minorities in the UK 161

4.3.1 Ethnic minorities in the labour market 161

4.3.2 The educational performance of ethnic minorities 164

4.4 Data and Methodology 170

4.4.1 Data: Next Steps 170

4.4.2 Descriptive statistics 178

4.4.3 Estimation strategy 184

4.5 Results 188

4.5.1 Anticipated discrimination and GCSE results 188

4.5.2 Ethnic differences 193

4.5.3 Selection on unobservables 199

4.5.4 Considering prior achievement: Value-added models 203

4.5.5 Robustness checks 209

4.6 Conclusions 211

Appendix C **217**

C.1 Variable definitions 217

C.2 Supplementary results 219

5 Conclusions 245

5.1 Summary of Results 246

5.2 Policy Implications, Limitations and Future Research 248

References 255

List of figures

2.1 Intergenerational earnings elasticity for fathers and sons - various countries 29

2.2 Observation of the key variables over the life cycle 39

2.3 IGE and rank coefficients by age groups (TSRA): household income 60

2.4 IGE and rank coefficients by age groups (TSRA): personal income 62

3.1 Percentage of local authorities (by region) within each Social Mobility Index performance quintile 98

3.2 National transition matrix 115

3.3 National transition matrix by gender 116

3.4 Directional rank mobility by gender 119

3.5 Unconditional quantile regression estimates of IGE 124

3.6 Relative and absolute mobility 129

3.7 Association between children's and parent's percentile ranks in England . . 132

3.8 Regional variation in intergenerational income mobility 134

B.1 National transition matrix - unadjusted household income 141

B.2 National transition matrix (personal income) 141

B.3 Directional rank mobility by gender (personal income) 142

B.4 Map of the UK and regions of England 143

B.5 Association between children's and parent's percentile ranks in the UK . . . 144

4.1 Pupils achieving 5+ A*-C GCSE and equivalent including English and Maths by ethnic group and gender (%), England, 2005. 167

4.2 Educational achievement by ethnicity and expectation of discrimination in the labour market 182
4.3 Educational achievement by ethnicity and expectation of discrimination in the labour market (cont.) 183
4.4 Kernel density distribution of KS2 scores by anticipation of discrimination 205
4.5 Kernel density distribution of KS2 scores by anticipated of discrimination - separately by ethnic group 206

List of tables

2.1 Summary statistics - key variables 42

2.2 Estimates of the intergenerational income elasticity (IGE) and rank coefficient in the UK 52

2.3 IGE and rank coefficient based on different measures of income 56

2.4 Cohabitation status by age groups 64

2.5 IGE and rank coefficient by co-residency status 65

2.6 Summary of treatment of outliers 67

A.1 Differences in means: in and out of sample 74

A.2 Parental characteristics: employment and education 76

A.3 OLS estimates of intergenerational income elasticities for the UK 79

A.4 First stage regressions (TSRA) 79

A.5 Second stage regressions (TSRA) 80

A.6 Estimates of the percentile rank coefficient for the UK 80

A.7 IGE and rank coefficient estimated for different age groups 81

A.8 Differences in sample means by co-residency status 81

3.1 Differences in means by gender 102

3.2 Estimates of intergenerational income elasticities and rank coefficients by gender 105

3.3 Intergenerational income mobility and assortative mating 109

3.4 Estimates of IGEs by quantiles 125

3.5 Absolute and relative mobility in the North and South of England 130

B.1 Summary statistics by gender 145

B.2 Estimates of IGEs by quantiles - conditional quantile regressions 146

B.3 Estimates of relative and absolute mobility by regions 146

4.1 The Educational System in England 165

4.2 Educational progression of the Next Steps Cohort 172

4.3 Anticipation of labour market discrimination by ethnic minority group (%) . 178

4.4 Raw gaps in GCSE performance by anticipated labour market discrimination 179

4.5 Anticipated discrimination and GCSE achievement - summarised results (OLS and LPM) 190

4.6 Anticipated discrimination and GCSE achievement, by ethnic group – Model 1 (OLS and LPM) 195

4.7 Anticipated discrimination and GCSE achievement, by ethnic group – Model 2 (OLS and LPM) 196

4.8 Bounding statements about β 202

4.9 Statements about δ (δ for $\beta = 0$) 203

4.10 Anticipated discrimination and GCSE achievement - summarised results value-added models VA-M1 and VA-M2 (OLS and LPM) 207

4.11 Anticipated discrimination and GCSE achievement - Summarised results value-added model VA-M2 (OLS and LPM); by ethnic group 208

C.1 Anticipation of labour market discrimination by ethnic minority group (%) . 219

C.2 Anticipation of labour market discrimination by ethnic minority group (%), unweighted 219

C.3 Raw gaps in GCSE performance by anticipated labour market discrimination (unweighted) . . . 219

C.4 Differences in sample means by anticipated discrimination: pooled sample . 220

C.5 Differences in sample means by anticipated discrimination: Indian 221

C.6 Differences in sample means by anticipated discrimination: Pakistani . . . 222

C.7 Differences in sample means by anticipated discrimination: Bangladeshi . . 223

C.8 Differences in sample means by anticipated discrimination: Black Caribbean 224

C.9 Differences in sample means by anticipated discrimination: Black African . 225

C.10 Differences in sample means by anticipated discrimination: Mixed/Other . 226

C.11 Anticipated discrimination and GCSE achievement - summarised results, models with and without school fixed effects (OLS and LPM) 227

C.12 Anticipated discrimination and GCSE achievement - complete results Model 1 (OLS and LPM) . . . 228

C.13 Anticipated discrimination and GCSE achievement - complete results Model 2 (OLS and LPM) . . . 229

C.14 Anticipated discrimination and GCSE achievement - complete Results Model 2 (OLS and LPM); Indian group only . . . 230

C.15 Anticipated discrimination and GCSE achievement - complete Results Model 2 (OLS and LPM); Pakistani group only . . . 231

C.16 Anticipated discrimination and GCSE achievement - complete Results Model 2 (OLS and LPM); Bangladeshi group only . . . 232

C.17 Anticipated discrimination and GCSE achievement - complete Results Model 2 (OLS and LPM); Black Caribbean group only . . . 233

C.18 Anticipated discrimination and GCSE achievement - complete Results Model 2 (OLS and LPM); Black African group only 234
C.19 Anticipated discrimination and GCSE achievement - complete Results Model 2 (OLS and LPM); Mixed/Other group only 235
C.20 Differences along the distribution of KS2 scores - summarised results Model 2 (OLS and LPM) 236
C.21 Separating 'Yes' and 'Don't Know'(DK) categories - summarised results (OLS and LPM) 237
C.22 Separating 'Yes' and 'Don't Know' categories - summarised results Model 2 (OLS and LPM); by ethnicity 238
C.23 Differences by gender - summarised results (OLS and LPM) 239
C.24 Differences by interview month (post July/05) - summarised results (OLS and LPM) 240
C.25 Anticipated discrimination and GCSE achievement - summarised weighted results (OLS and LPM) 240
C.26 Anticipated discrimination and GCSE achievement - summarised weighted results value-added models VA-M1 and VA-M2 (OLS and LPM) 241
C.27 Anticipated discrimination and GCSE achievement - summarised weighted results Model 2 (OLS and LPM) by ethnicity 242
C.28 Anticipated discrimination and GCSE achievement - summarised weighted results value-added model VA-M2 (OLS and LPM), by ethnicity 243

Chapter 1

Introduction

1.1 Background and Motivation

Increasing income and wealth inequalities at the global-scale (Keeley, 2015; OECD, 2011) and their serious implications for political, economic and social dynamics, have become a topic of concern and extensive research interest in recent years. In 2015, the OECD reported that the richest 10% of the population earned, on average, 9.6 times the income of the poorest 10% in OECD countries (OECD, 2015). This represents a considerable increase since the 1980s, where the richest 10% earned 7 times more than the 10% poorest. Such high levels of income inequality have raised concerns in the socio-political environment. Specifically, regarding the consequences of the prolonged persistence of high inequality levels over time and the exacerbation of social divides when inequality is associated to low levels of social mobility. As such, recent years have seen a growing interest in the study of 'equality of opportunities', with the objective of measuring the extent of equal opportunities, identifying the most vulnerable groups and problematic areas of opportunity hoarding and understanding the mechanisms through which social mobility occurs. The term 'equality of opportunities' usually refers to the belief that a person's chances to prosper in life - for example, to receive a high quality education or to obtain a favourable job - should be unrelated to ascribed characteristics such as gender, ethnicity or socio-economic origins (Roemer, 1998; Roemer and Trannoy, 2015). This contrasts with the notion of equality of outcomes or conditions, which refers to distributional inequalities of income, wealth and living conditions. The body of research on equality of opportunities is aimed at the design and promotion of policies that aim to 'level the playing field' between people from different backgrounds.

Social mobility is frequently considered as an indicator of the level of equality of opportunities,[1] and thus a growing body of literature has been dedicated to the study of different types of social mobility. A great emphasis has been given to studying social mobility through an intergenerational perspective. Research in this vein aims to quantify and understand the association between the socio-economic status of parents and the economic outcomes of their children as adults. Or, as the OECD describes, "[intergenerational] mobility reflects the extent to which individuals move up (or down) the social ladder compared with their parents" (OECD, 2010, p.184). Intergenerational mobility is typically measured through the degree of association between parental outcomes and those of their children as adults.[2] The study of intergenerational mobility encompasses various facets of socio-economic status, including mobility of social class, occupation, income and education. Although each of these aspects are regarded as important on their own, the primary focus of most economic research on intergenerational mobility, including this thesis, is to analyse income and earnings mobility. Since income and earnings are intrinsically related to the other dimensions of socio-economic status, this represents the most direct way of measuring economic mobility.

A strong association between the status of parents and their children in adulthood reflects low intergenerational mobility and is often identified as a violation of equal opportunity norms. Conversely, a high degree of intergenerational mobility is often considered one of the pillars of a fair and equal society, where fairness and equality mean that hard work and talent (or ability) are rewarded rather than purely inherited advantages, that is, that there is greater equality of opportunities (Narayan et al., 2018). Under the context of rising income inequality, family background is likely to play a bigger role in determining adult outcomes of individuals, with individual efforts playing a smaller role, thus increasing intergenerational

[1] Although the link between these two concepts is not straightforward. As Roemer (1998) points out, assuming that equality of opportunities reflects a complete disassociation between outcomes of parents and their children implies a narrow definition of equality of opportunities. He defends that a more flexible approach to equality of opportunity would be focusing on a scenario where material circumstances of children do not have a determining influence on their later outcomes. Roemer's ideas have served as base for another strand of the literature on inequality of opportunities, which focuses on determining the weight that various socio-economic circumstances have on the outcomes of an individual.

[2] An early review of the international literature on intergenerational mobility is presented by Solon (1999). This has been later updated by Black and Devereux (2011).

inequalities (Corak, 2013; Narayan et al., 2018; OECD, 2010, 2018*b*). If incomes are strongly associated across generations[3] - i.e. if there is high intergenerational income persistence - this indicates that children from poorer families are more likely to have limited opportunities to overcome this reality. In the context of high income inequality, a lack of social mobility can have serious implications for productivity and economic growth (Narayan et al., 2018; OECD, 2018*b*). Firstly, it is more likely that these levels of income inequality will persist, or worsen over time. Secondly, from the perspective of economic efficiency, a lack of upward mobility at the bottom could be especially harmful if the human potential of those from disadvantaged backgrounds is wasted and underdeveloped and if investment opportunities are missed out on or mis-used. As described by the OECD, rising inequality can prevent upward social mobility, by making it more difficult for hard-working people to climb the social ladder (OECD, 2010). Conversely, lack of downward mobility at the top of the distribution could mean persistent gains for the few rich and the hoarding of opportunities of access to education and health, among other important resources.

Many studies have presented evidence of this association between the lack of social mobility and high levels of income inequality. The empirical evidence points to low levels of intergenerational earnings mobility in countries with high income inequality (Corak, 2006, 2013; d'Addio, 2007; OECD, 2010, 2018*b*), such as the United States (US) and the United Kingdom (UK). Conversely, intergenerational mobility is considerably higher in the Scandinavian countries, where income is distributed more evenly. This relationship has been illustrated empirically with the Great Gatsby Curve, which shows that countries with greater income inequality, as measured by the Gini coefficient, are typically countries in which economic advantages and disadvantages are strongly linked across generations.[4] If a society

[3]Note that 'generation' is a term used loosely throughout this thesis, referring to a group of individuals from a set of birth cohorts - this is similar to what is done by other authors in this field (see for example Solon 1999). Even if the children are born in the same year, it does not mean that their parents were necessarily born in the same year, as they may have begun parenthood at different ages.

[4]The term Great Gatsby Curve was first used by Alan Krueger in the speech "The Rise and Consequences of Inequality," to the Center for American Progress in 2012. Representations of this curve have been presented by several studies (see for example Corak 2006, 2013, Andrews and Leigh 2009, Blanden 2009; Ermisch et al. 2012, Jerrim and Macmillan 2015, OECD 2018*b* and d'Addio 2007).

is highly unequal and has very low levels of intergenerational mobility then it is likely that inequalities will persist.

In most OECD and developing countries, upward intergenerational mobility has been observed in absolute terms (OECD, 2018*b*). This means that there has been social progress over time, and that general living conditions have improved in comparison to those observed for the previous generation. On average, individuals are benefiting from higher income, higher education levels, have better health outcomes and more opportunities than their parents. However, thinking about intergenerational mobility in terms of an economic ladder, some individuals (or groups) are more likely to be in a relatively better situation than their parents, reflecting upwards mobility between generations. Conversely, other individuals may be stuck in similar, or even worse positions on the ladder, reflecting downwards mobility. When considering the concept of intergenerational mobility, it is important to make this distinction between absolute and relative mobility; high levels of *relative* mobility reflect both upward opportunities and downward risks.

In the UK, concerns about the steep rise in income inequality in the 1980s and its persistence at high levels in the recent decades (Belfield, Blundell, Cribb, Hood and Joyce, 2017; Blundell and Etheridge, 2010; Brewer and Wren-Lewis, 2016) alongside the growing perception that intergenerational mobility has declined, has contributed to a resurgence of social mobility as a topic of interest in public and academic debates. In one of her first speeches as Prime Minister, Theresa May stated “I want Britain to be a place where advantage is based on merit not privilege; where it’s your talent and hard work that matter, not where you were born, who your parents are or what your accent sounds like" (May, 2016). A statement that clearly reflects the importance and relevance of the intergenerational dimension of social mobility today.

Previous research has described the UK as a country with relatively low levels of intergenerational earnings (Blanden, 2009; Blanden et al., 2004; Dearden et al., 1997; Jäntti et al., 2006) and educational mobility (Jerrim and Macmillan, 2015). In spite of these research efforts, the extensive data requirements for this type of analyses on intergenerational mobility have meant that this body of evidence is still relatively thin. As a result, there is still uncertainty about the current levels of intergenerational income mobility in the UK. In particular, the current evidence is largely based on survey data from only two birth cohorts. There is uncertainty surrounding the accuracy of their estimates, their comparability and whether they are representative of the entire UK population. The research on British income mobility is also surprisingly sparse, as most datasets only contain information on individual earnings for the second generation, which does not fully capture all sources of income and can only be observed for individuals in employment. This represents a key knowledge gap in our understanding of social mobility in the UK which I aim to address in this thesis.

The initial part of this thesis focuses on the measurement of the degree of intergenerational income mobility in the UK, and on improving our understanding of how the levels of social mobility may vary between different social groups. Distinguishing between upwards and downwards movements and identifying groups of individuals with higher or lower mobility rates are of crucial importance for policy design, especially when the goal is to equalise opportunities. In addition, the particularly high levels of regional inequalities observed within the UK invites the study of intergenerational mobility from a regional perspective i.e. is there some kind of 'postcode lottery' whereby the influence of family background for future outcomes varies with the region of upbringing?

In recent years, new developments in the social mobility literature attributed to the acquisition of new and detailed longitudinal data, have allowed researchers to obtain new measures of mobility and compare how intergenerational income and earnings mobility vary across countries, across regions and over time (Black and Devereux, 2011; Blanden et al., 2004, 2013; Chetty et al., 2014; Corak, 2013; Jäntti et al., 2006; Landersø and Heckman, 2017; OECD, 2018*b*). This surge in empirical work on social mobility has not only provided greater understanding and more precise measurement of intergenerational mobility, but has also afforded the opportunity to identify and explore the channels through which transmission of socio-economic status across generations takes place. Understanding the causal mechanisms underlying this process is crucial for the development of appropriate equalising polices.

Theoretically, and in the empirical literature, investments in human capital, for example in education and skills, have been described as being critical channels of transmission of socio-economic (dis)advantage across generations (Becker and Tomes, 1986; Blau and Duncan, 1967; Breen and Jonsson, 2005; Jerrim and Macmillan, 2015; Solon, 2004). Increasing investments in education and skills not only raises young people's hopes and aspirations, but also provides the means to achieve higher future earnings and to attain greater economic status than their parents. However, such investments are costly and, from an economic efficiency perspective, only worthwhile for the individual as long as they yield private returns in the labour market.[5] Significant socio-economic differences in the cognitive ability of children are observed at all stages of education. These differences could be a reflection of lower investments in human capital among children from disadvantaged backgrounds. The socio-economic gaps open up at young ages, even before the start of formal schooling (Becker, 2011; Cunha et al., 2006; Heckman and Mosso, 2014). They tend to persist (and in some countries, can even become wider) during school years, with factors such as school selection and quality, neighbourhood effects and parental investments also influencing the acquisition of skills (Björklund and Salvanes, 2010).[6] By the end of secondary school, several socio-economic gaps are apparent with respect to children's aspirations and other non-cognitive skills as well as achievement (OECD, 2010).

This body of research points to a strong connection between the persistence of income inequality across generations and the unequal distribution of educational attainment (Björklund and Salvanes, 2010; Black and Devereux, 2011). Thus, access to, and the quality of education could help remove obstacles to social mobility and promote equality of opportunities for all. This notion has generated debates in the academic and political spheres, mainly related to

[5]However, the broader benefits of education and training may go beyond the private returns and spillover to the economy, yielding social returns.

[6]For a detailed overview of the links between education and many indicators of family background, see Björklund and Salvanes (2010).

policies targeting improved quality of education for all and access to education, especially for children from disadvantaged backgrounds.

It is therefore imperative to study the fundamental determinants of pupil's educational outcomes, such as those associated with family background and personal characteristics, in order to better understand how to design education policies that can also help improve social mobility. Despite considerable efforts from the economics of education literature to understand various determinants of educational achievement (see for example Almlund et al., 2011; Björklund and Salvanes, 2010; Hanushek and Woessmann, 2010; Heckman et al., 2006; Sacerdote, 2010), little is known about how perceptions and expectations about the existence of (equal) opportunities can influence real decisions and behaviours related to the accumulation of human capital.

The perception of an 'unfair' society, in which opportunities are unequal between social groups due to discrimination based on certain characteristics, such as gender or ethnicity, could influence an individual's motivation and effort and affect their decisions over investment in education. Notably, expectations of facing discrimination reflect beliefs about the equality of opportunities in the labour market. It is possible that perceptions and expectations of being treated differently in the labour market, for example because of skin colour or sexual orientation, could reinforce beliefs about the 'unfairness' of society. According to various theoretical frameworks, discrimination can influence individual decisions and behaviour as a response to expected differential treatment, particularly those decisions related to human capital investment (Arcidiacono et al., 2010; Coate and Loury, 1993*b*; Lang and Manove, 2011; Lundberg and Startz, 1983). Thus, self-fulfilling discriminatory beliefs could generate persistent inequality (Piketty, 2000). Disentangling the role of discriminatory beliefs from that of other channels through which socio-economic status is transmitted across generations is a major challenge for empirical research (Piketty, 2000). As a result, the link between discrimination in the labour market and its influence on pre-market outcomes, such as decisions to invest in human capital and educational attainment, has remained elusive and represents a key knowledge gap in the economic literature. In the second part of this thesis, I delve deeper into this topic and examine the role of expectations of future labour market discrimination on the educational achievement of ethnic minority pupils.

1.2 Aims, Research Questions and Contributions

These topics form the focus of this thesis, which will include three main chapters that empirically examine different aspects related to the mechanisms underpinning social mobility in the UK.

The first empirical chapter (Chapter 2) focuses on the role of family background. It examines the degree of intergenerational income mobility in the UK at the national level. The main motivations here are to; firstly, update and improve current estimates of income mobility for the UK using a new dataset that allows me to examine the level of income mobility for individuals from a younger generation; and secondly, to use new methodological developments in this literature to obtain robust measures of income mobility.

In Chapter 3, I examine heterogeneities in income mobility by three relevant dimensions: by gender, across the income distribution and by region of upbringing. This chapter aims to provide a more complete picture of the social mobility puzzle in the UK. By unpacking intergenerational income mobility over these three dimensions, a better understanding about the nature of social mobility in the UK can be reached and provide key insights towards informing policy and promoting more equal opportunities.

Chapter 4 expands the analysis beyond the influence of socio-economic background to individual factors that may affect prospects of social mobility. Here, I examine how perceptions of unequal opportunities in the labour market might influence decisions to invest in human capital. In particular, given the links between educational investments and opportunities for upward social mobility, I seek to understand whether and how expectations of facing future discrimination in the labour market are associated with actual decisions of investing in education and developing skills through educational attainment. I focus specifically on individuals from ethnic minority origins, motivated by the observation of various ethnic gaps in labour market outcomes in the UK. I link the literature on theories of discrimination and determinants of educational achievement among ethnic minorities to empirically examine the role of discrimination in influencing human capital acquisition.

This thesis contributes to the understanding of social mobility in multiple ways. First, it provides updated estimates of the current levels of intergenerational income mobility in the UK. Secondly, it provides a more nuanced picture of the patterns of social mobility, by examining heterogeneities in income mobility over several key dimensions. Knowing which groups are presented with fewer opportunities for mobility, and especially upwards mobility, is of crucial importance for policy design that aims to level the playing field. Thirdly, it contributes to the literature on discrimination and the economics of education by empirically examining the influence of expectations of discrimination on actual decisions to invest in human capital through educational effort and achievement. Fourth, it contributes to the literature on the determinants of educational achievement by revealing factors specifically relevant for ethnic minority pupils. Furthermore, in Chapter 2, I propose a novel approach to obtain a comparable proxy of long-run income for two generations. The two-stage residual approach (TSRA) method provides an improved measure of income, adjusting for life-cycle effects by controlling flexibly for the age-income profiles of young individuals and their parents. Lastly, this thesis makes a quantitative contribution to the existing literature by making use of two very detailed panel datasets with exceptionally rich information on income and earnings across generations in the second and third chapters, and on pupil's characteristics and expectations, in the fourth chapter.

1.3 Structure and Content of this Thesis

This thesis consists of three separate empirical studies presented in Chapters 2, 3 and 4. These studies utilise individual-level data and adopt modern econometric techniques to deepen our understanding of the mechanisms that contribute to the perpetuation of inequalities in the UK.

Chapter 5 concludes this thesis. The three chapters constituting the main part of this thesis are briefly discussed below.

1.3.1 Overview of Chapter 2

Chapter 2 investigates the degree of intergenerational income mobility in the UK, focusing on measuring the association between family background and children's incomes (in adulthood) in a precise and robust manner. Specifically, this chapter uses new data from the harmonised British Household Panel Survey and Understanding Society to measure intergenerational income elasticities and rank coefficients for the UK, for cohorts born between 1973-1992. Here, for the first time, income mobility is examined for this younger generation.

In addition to the traditional OLS approach, I propose a method to the estimation of income elasticities and rank coefficients, the two-stage residual approach (TRSA). This method provides an improved measure of long-run income for both generations, in order to overcome common estimation issues related to measurement error and life-cycle bias. The first stage estimates parents' and children's long-run incomes, while the second stage uses these estimates to obtain measures of relative income mobility. Comparisons with the standard estimates obtained via OLS are also provided.

My findings suggest a strong degree of income persistence across generations in the UK. The intergenerational elasticities based on household income indicate that for every additional (reduced) 10% of parental income advantage (disadvantage) around a quarter (2.6%) will be passed on to the next generation. Estimates based on alternative income variables for the second generation and the results for rank coefficients indicate a similar degree of immobility. In the final part of this chapter, I examine the sensitivity of my results to different methodological and model specification choices, changes in sample restrictions and treatment of outliers.

1.3.2 Overview of Chapter 3

Chapter 3 builds on the findings from the previous chapter and is concerned with deepening our understanding of intergenerational income mobility in the UK. Using data from the harmonised BHPS and Understanding Society, I examine heterogeneities in income mobility over three key dimensions.

Does income mobility differ between sons and daughters? Considering mobility in terms of household income, I find similar levels of income mobility by gender. The focus on the transmission of economic resources at the household level affords the opportunity to study dynamics of household formation and assortative mating. Focusing on individuals with partners, I find evidence of assortative mating on income. Further, I find that the contribution of assortative mating to income persistence relates to the share of partner's contribution to household income.

Does income mobility vary across the income distribution? Here, I examine income mobility at different points of the income distribution of both generations. This characterisation of relative mobility provides insights on groups that experience less mobility, which can be of policy interest. Using quantile transition matrices, I find that there is more persistence at the tails of the parent's income distribution. Then, distinguishing between patterns of upward and downward mobility by using estimators of rank directional mobility, I find indication that the patterns of directional movements vary by gender. I find, however, little variation in income mobility across the children's income distribution.

Is there regional variation in income mobility in the UK? Examining intergenerational income mobility based on childhood location, I find that stark differences in absolute and relative mobility exist across regions in the UK. Particularly, I identify a clear north-south divide in England. Regions in the North of England display much lower levels of both relative and absolute mobility than in the South of England. A more nuanced picture emerges when looking at a higher level of regional disaggregation. These findings reinforce the notion that there is a ‘postcode lottery’ in the UK, whereby regional characteristics partly determine the potential for social mobility.

1.3.3 Overview of Chapter 4

Chapter 4 examines the role of expectations of facing future labour market discrimination on the educational achievement of English ethnic minority pupils at the end of secondary school. This chapter addresses a key knowledge gap in the existing literature by empirically examining the direct influence of expectations of discrimination on actual behaviour, for which very little evidence currently exists. This analysis builds on theories of discrimination and on research on determinants of ethnic inequalities, in an attempt to gain an understanding of the influence of discriminatory beliefs on educational achievement.

I employ unique data on direct expectations of facing labour market discrimination from Next Steps, for a cohort of British teenagers from ethnic minority backgrounds born in 1989/90. This chapter implements a standard least squares approach to investigate the influence of anticipating labour discrimination on a number of educational achievement outcomes reflecting performance at GCSE level.

This study finds a positive and significant relationship between anticipating labour market discrimination and crucial educational outcomes, while controlling for a series of individual, family and parental characteristics, as well as for school fixed effects. The findings of this study suggest that ethnic minorities have an incentive to over-invest in education and, as a result, achieve higher results in high stake exams when they expect future discrimination, compared to their similar counterparts who do not anticipate discrimination. This chapter further investigates heterogeneous effects of anticipated discrimination by ethnicity. My results reveal variations across ethnic groups, indicating that certain ethnic minorities might have a greater incentive to try to overcome the expected ethnic penalty in the labour market.

Chapter 2

Intergenerational Income Mobility in the UK: New evidence using the BHPS and Understanding Society

2.1 Introduction

A common goal of governments and societies is to ensure their citizens have equal opportunities for social advancement and to understand how this can be achieved in practice. This makes the study of intergenerational mobility a topic of long-standing interest to economists, sociologists and other social scientists, as well as to and policy-makers. The understanding of how economic status is linked and transmitted across generations is a crucial part of the puzzle of why inequalities persist.

Since the 1970s, a number of studies have examined the extent of intergenerational earnings mobility for a select group of developed countries for which data were available, including the US, Germany, the UK and a number of Scandinavian countries.[7] A common observation amongst these studies is the declining levels of relative intergenerational mobility in recent years, with family background playing an increasingly important role in the determination of an individual's socio-economic status in adulthood. This process is closely related to the perpetuation of inequalities in society as both socio-economic advantages and disadvantages are transmitted across generations (OECD, 2018*b*). In the UK, the increasing levels of income

[7] For detailed reviews of this literature, see Solon (2002), Black and Devereux (2011) and Jäntti and Jenkins (2015).

inequality in the 1980s and early 1990s and the persistent high levels thereafter (Belfield, Blundell, Cribb, Hood and Joyce, 2017; Blundell and Etheridge, 2010; Brewer and Wren-Lewis, 2016) have raised concerns of the government and population alike, regarding the general availability of equal opportunities. This has brought social mobility to the forefront of the political agenda. Despite the critical importance that intergenerational mobility plays in understanding the income dynamics of a society, and its subsequent relevance to inform policy, the empirical evidence surrounding the degree of income mobility in the UK remains relatively scarce.

The primary objective of this chapter is to deepen our understanding of the levels of social mobility in the UK. More specifically, in this chapter I measure the degree of intergenerational income mobility at the national level and investigate the extent to which family resources at childhood are related to an individual's living standards in adulthood. By making use of a dataset combining the British Household Panel Survey (BHPS) and Understanding Society, I provide contemporary estimates of intergenerational income elasticities and rank coefficients.

The data requirements to undertake this type of intergenerational analysis are usually extensive. The typical variables of interest (e.g. income, earnings or educational levels) need to be observed for at least two generations, and most surveys are not conducted over a sufficiently long time period to allow for this. Previous research concerned with the measurement of intergenerational earnings mobility (Blanden et al., 2004; Dearden et al., 1997), educational mobility (Jerrim and Macmillan, 2015) and, more recently, income mobility for the UK (Belfield, Crawford, Greaves, Gregg and Macmillan, 2017; Gregg et al., 2017, 2019) have mainly relied on data from two sources: (i) the National Child Development Study (NCDS), a cohort study for individuals born in one particular week in March 1958, and (ii) the British Cohort Study (BCS), which surveyed individuals born in one single week in April 1970. The results from these studies and other wider cross-country comparisons place the UK in the group of countries with relatively low levels of mobility by international standards (Blanden, 2009; Jäntti et al., 2006; OECD, 2018*b*; Raaum et al., 2008).

Most early studies typically focused on the association between father's (and sometimes parents') earnings and children's (mostly sons') earnings (Couch and Lillard, 1998; Dearden et al., 1997; Solon, 1992; Zimmerman, 1992). In this respect, they have fallen short of being able to consider the transmission of economic status and well-being in a broader sense, at the household level. The more recent international literature on intergenerational mobility has shifted towards a more family-based approach, with the aim of capturing the association between family resources available in childhood and socio-economic outcomes of individuals in adulthood, focusing on living standards and household well-being (Black and Devereux, 2011; Chadwick and Solon, 2002; Chetty et al., 2014; Jäntti and Jenkins, 2015). Studies for the UK, usually based on the NCDS and BCS data, are somewhat constrained in their ability to understand the broader context of economic mobility because they rely on labour earnings for the second generation (usually sons), rather than family income (Blanden et al., 2004; Gregg et al., 2017). This precludes the potential to uncover relevant insights into the role of pooled household resources for both generations. In addition, their estimates reflect the levels of mobility of these two specific birth cohorts of individuals only, while the situation of other cohorts remains obscure. Further, various data limitations have restricted the methodological possibilities of some of this work and raised concerns about the exposure of results to biases due to life cycle effects and measurement error. Much of this discussion has generated uncertainty around the wide-ranging estimates of intergenerational mobility present in the literature, thus limiting the opportunities for comparing the levels of intergenerational mobility over time and making international comparisons more difficult.

This chapter contributes to the literature by estimating intergenerational income mobility and considering the transmission of economic status between parents and their children at the household level. To do this, I use a new dataset comprised of the British Household Panel Survey (BHPS) and Understanding Society (UKHLS), spanning a period of 27 years, which affords the opportunity to overcome some of the inherent methodological limitations present in previous studies. In particular, this chapter investigates the extent of intergenerational income mobility in the UK for individuals born between 1973-1992. Previous studies on the UK have mostly focused on individuals from the 1958 and 1970 birth cohorts, who are now approximately 62 and 50 years old. Here, for the first time, I examine intergenerational income mobility for younger generations.

This chapter addresses a key gap in the literature by investigating the current state of intergenerational mobility in the UK focusing on these younger cohorts, which are of critical relevance for current and future policy design and implementation. The wealth of information contained in this panel dataset, which includes observations of income for two generations (parents and their offspring) in adulthood over multiple years, allows the estimation of robust parameters of income mobility. Moreover, this data allows me to test the sensitivity of the results to a series of methodological and empirical choices, previously not possible given the limitations of the NCDS and BCS data sets. To the best of my knowledge, this is the first study to use this harmonised data set to estimate intergenerational income mobility in the UK.

I start my empirical analysis by employing the standard Ordinary Least Squares (OLS) to estimate intergenerational income elasticities (IGEs) as well as recently developed rank-based measures of mobility (Chetty et al., 2014; Dahl and DeLeire, 2008), the rank coefficients. The OLS approach has been widely applied in this literature and allows for comparisons with other studies to be made. Further, this chapter proposes a novel methodology to the estimation of IGEs and rank coefficients, the two-stage residual approach (TRSA). The TSRA provides an improved measure of long-run income for both generations, adjusting for life cycle effects by controlling flexibly for the age-income profiles of young individuals and their parents. Moreover, it enables the full set of multi-year information on income available in the data, addressing an important source of attenuation bias. The first stage estimates parents' and their children's adjusted incomes, while the second stage uses these estimates to derive measures of relative mobility.

Using both OLS and TSRA, I find that the intergenerational income elasticity in the UK is 0.255-0.269 and that this value is precisely estimated. These values indicate the fraction (in %) of every additional 1% of parental income advantage (or disadvantage) that will be passed on to their children. In addition to estimating IGEs, I estimate intergenerational rank coefficients. Rank coefficients are an alternative measure of relative mobility that characterises intergenerational mobility from a positional perspective, analysing the persistence of percentile ranks of parents and children in the income distribution. The rank coefficient is estimated to be 0.291-0.301, meaning that an increase of 10 positions in the percentile rank distribution of parents relates to an improvement of around 3 positions in the ranks of their children. To complement this analysis, I examine the sensitivity of my main results to a series of methodological and model specification choices, including the use of alternative income measures, relaxing the sample's age restrictions, differences by cohabitation status of the children as adults and the implications of treating outliers in income.

The remainder of the chapter is structured as follows. The next section presents an overview of the relevant literature on intergenerational mobility, describes the standard methodology and most common estimation issues and provides a brief summary of the current evidence for the UK. A detailed description of the data sources and the estimation strategies employed in this chapter are presented in Section 2.3. Section 2.4 presents the results from the main analysis alongside robustness checks and discussion. The chapter ends with a conclusion in Section 2.5.

2.2 Intergenerational Income Mobility: Background and Measurement

Intergenerational mobility researchers are usually interested in how socio-economic status is associated and transmitted across generations. Since this transmission of advantages and disadvantages is an important contributor to the perpetuation of inequalities in society, measuring the degree of persistence and understanding the mechanisms behind intergenerational transmissions are essential aspects of the study of social mobility. Socio-economic status encompasses a series of factors, including income, earnings, consumption power, health, education and occupation, among others. As a result, there are various ways of measuring intergenerational mobility depending on the main focus of the researcher.

While intergenerational mobility refers to the transmission of status between generations, another common way to look at social mobility is from the intragenerational perspective (Jäntti and Jenkins, 2015). Intragenerational mobility refers to how an individual's socio-economic status changes between years during their lifetime. The recent work by the Social Mobility Commission (SMC), particularly the Social Mobility Index, is a good example of research focusing on intragenerational mobility in Britain (Social Mobility Commission, 2016*a*, 2017*a*,*b*).[8] The SMC studies intragenerational inequalities at various ages and stages of life and emphasises the importance of policy actions towards young children and schools to improve equality of opportunities. The fundamental conceptual differences between intra- and intergenerational mobility have created two separate, although complementary, strands in the literature of social mobility. This chapter focuses on intergenerational mobility. Throughout this thesis, I will refer to different types of social mobility from the intergenerational perspective. It is important to start by distinguishing between two types: absolute and relative mobility. Absolute mobility refers to changes in the absolute value of an outcome of interest (income, education, health) across generations and indicates how much living standards have improved/worsened over time. On the other hand, relative mobility reflects the extent to which individuals can reach a better or worse *position* in the distribution of income, education, or any given outcome of interest than that of their peers, depending on their parents' position in the social ladder. For example, in the case of income mobility, relative mobility looks at whether individuals who rank high or low in terms of income also had parents who ranked high or low. The main focus of this chapter is on relative income mobility.

The theoretical framework for the empirical analysis of intergenerational income mobility was developed by Becker and Tomes (1979, 1986). The set-up of the model by Becker and Tomes begins with a utility function that each family maximises across several generations.

[8]The Social Mobility Commission is an independent advisory non-departmental public body sponsored by the Department of Education. It was created in 2016 with the aim of monitoring and improving the state of social mobility in the UK.

Parents can (and want to) influence the future earnings capacity of their children directly, through investing time and money to provide their children with additional human capital and non-human capital. The degree of parental investment will, of course, depend on parents' own preferences and on their budget and credit constraints. In addition to these investments, the future income of children will also be (indirectly) influenced by other family-related 'endowments', such as genetic disposition, ability, culture, values, family connections and other skills, knowledge and goals provided by the family environment (Becker and Tomes, 1979). This mixture of human and non-human capital and other family-related endowments will be generically called 'family resources'.

This framework was further developed by Solon (2004) with a model that highlights important channels of intergenerational persistence. In the model by Solon (2004), the steady-state intergenerational income elasticity is defined as a function of four key factors: mechanical transmission of income generating traits (e.g. genetics), the efficacy of the investment in children's human capital, the earnings return to the accumulation of human capital and the public investment in education. He argues that differences in mobility over time and place could be a consequence of differences in any of these factors. Ultimately, the influence of 'family resources' on the children's future earnings or income will also depend on the broader context, such as the social institutional environment, or even luck - these factors are responsible for differences in the return to human capital. Commonly, this literature on intergenerational income/earnings mobility uses regression analysis to estimate the extent of mobility. In this framework, the generational association between parents' and children's income is measured by the intergenerational elasticity (IGE).

Since the early studies of intergenerational mobility from the 1970s, a growing body of literature has focused on estimating to what extent socio-economic status is transmitted between generations and identifying the mechanisms behind this association. While economists tend to focus on income, earnings or education as representative measures of permanent socio-economic status (Black and Devereux, 2011; Jäntti and Jenkins, 2015; Solon, 2002), sociologists have often focused on social class and occupational status (Breen and Jonsson, 2005; Erikson and Goldthorpe, 1992, 2002). The first papers on intergenerational mobility by economists focused on intergenerational earnings and wage mobility between pairs of fathers and sons. Later, other studies started focusing on a number of other outcomes that are closely related to wages, such as educational and occupational mobility. Comprehensive reviews of these studies on earnings and income mobility are provided by Solon (1999), Black and Devereux (2011) and Jäntti and Jenkins (2015).

Based on the idea that it may both influence and reflect the other factors directly (health, occupation, education, neighbourhood, etc), income has been considered as a broader, more representative indicator of long-run economic status, and the focus of more recent studies has therefore shifted to measuring intergenerational income mobility. The literature on income mobility suggests that parental income is a good predictor of children's economic situation in adulthood (Chadwick and Solon, 2002; Chetty et al., 2014; Lee and Solon, 2009). Income measures (rather than earnings or occupation) are not only a broader representation of economic status, but also more useful for the study of mobility among women since, unlike wages or earnings, it avoids the complex issues of selection into labour force participation and part time work and therefore enables the inclusion of mothers and daughters in the analyses of social mobility.[9] Therefore, the main motivation behind estimating intergenerational income mobility relates to the idea that parents invest time and money into their children and that these investments, alongside the availability of family resources during childhood, are an important determinant of children's socio-economic status in adulthood. If the income of individuals in adulthood is closely associated with their parental or family income during childhood years (i.e. low relative income mobility), this implies that having less or more access to resources and opportunities (at least relatively) during childhood could have several implications for their future economic success in the long term. In the next part of this chapter, I explain how intergenerational persistence (or mobility) of socio-economic status is measured in practice, and outline the main estimation issues commonly faced by researchers in this field.

[9] Women are increasingly important in the labour force, but were frequently excluded from the earlier studies on intergenerational mobility due to a lack of information on their wages or earnings.

2.2.1 Measuring intergenerational mobility

Intergenerational elasticities (IGEs)

Based on the theoretical framework by Becker and Tomes (1979, 1986), the standard empirical strategy used in the literature to estimate intergenerational earnings or income mobility consists of relating, across generations, a proxy of individual, family or household 'permanent' income or earnings. Typically, the studies on intergenerational mobility use a regression-based approach to estimate the generational association between parents' and children's income (Black and Devereux, 2011; Jäntti and Jenkins, 2015). The typical equation used to estimate the intergenerational elasticity is:

$$log(Y_i^{child}) = \alpha + \beta \; log(Y_i^{parent}) + \varepsilon_i \tag{2.1}$$

where Y_i^{child} is a measure of long-run economic status (here, income) of children in adulthood and Y_i^{parent} is the same variable for the parents, captured during childhood years.

The main coefficient of interest, the intergenerational elasticity (β) is a measure of intergenerational persistence and represents the extent to which parental income is transmitted to the next generation. A larger coefficient indicates greater persistence in incomes, or less mobility. On the other hand, $1-\beta$ is a measure of mobility. When β is small, $1-\beta$ will be large, indicating less persistence in incomes, that is, more mobility across generations. Referring back to the theoretical framework, the intergenerational elasticity captures a mixture of factors: the direct investments in human and non-human capital, and also the broader inherited family 'endowments' (Becker and Tomes, 1979, 1986). Empirically, estimates of intergenerational elasticity tend to lie between 0 and 1. However, other values are theoretically possible. In summary:

- If $0 < \beta < 1$: there is some intergenerational persistence of income, which means that higher parental income is linked to higher child income in adulthood;

- If $\beta = 0$: there is complete intergenerational mobility, which means that children's income in adulthood is not associated with parental income;[10]
- If $\beta = 1$: there is complete persistence, and this would mean full determination of children's income in adulthood by parental income;
- If $\beta < 0$: there is a negative relationship between parental and child income, which means that higher parental income is associated with lower child income in adulthood. This case is a theoretical possibility but has not been observed empirically.

Intergenerational elasticities are, therefore, a measure of relative immobility, or relative persistence (Chetty et al., 2014; OECD, 2018*b*).[11] They compare the socio-economic outcomes (in this chapter, income) of children (as adults) from families located at different points of the social ladder, i.e. of children from low- and high-income families relatively.

Common estimation issues

Two common issues related to the estimation of IGEs have been repeatedly described in the intergenerational mobility literature: (i) measurement error due to transitory variation in observed income measures and (ii) life cycle bias. The presence of these problems can produce biases in the estimates of intergenerational elasticities.

The first issue, measurement error can arise when using short-run proxies as measures of long-run economic status. Lacking direct information on long-run income or earnings, these must be derived from the observed variables present in the data sets - usually measures of annual, monthly or weekly income or earnings. For example, the very first studies estimating earnings elasticities relied on single year measures of income or earnings due to data availability issues (Solon, 1992). This source of bias has been widely discussed in the literature on intergenerational income mobility (Grawe, 2006; Solon, 1989, 1992; Zimmerman,

[10] This result is very unlikely to be found empirically, given that the broader inherited family endowments also captured by the coefficient, such as ability, values and genetic predisposition.

[11] This differs from the concept of *absolute* mobility, which indicates how much living standards (health, education, income) have improved or deteriorated across generations (i.e. between parents and their children).

1992). The use of a short-run proxy for long-run status can imply that 'permanent' income will be measured with error due to a transitory fluctuation around long-run status that would be captured in the data.[12] This results in a classical errors-in-variables attenuation bias, leading to a reduction of the estimated intergenerational income or earnings elasticities (Grawe, 2006; Solon, 1989, 1992). In other words, β will be biased downwards towards zero, giving the impression that intergenerational mobility is higher than it actually is. Based on the classical model of measurement error, although error in the measurement of the dependent variable (children's income) would not bias the estimate of β, it could lead to a loss of precision.[13] Conversely, measurement error in the explanatory variable, parental income, could have serious implications for the consistency of β.

Typical corrections to this classical measurement error problem in the literature include using a multi-year average of parental income observations in order to reduce the transitory variation (Chadwick and Solon, 2002; Chetty et al., 2014; Gregg et al., 2017; Mazumder, 2005; Solon, 1992), or utilising an IV approach to predict long-run parental economic status based on other parental characteristics, such as education and occupation (Dearden et al., 1997; Nicoletti and Ermisch, 2008; Solon, 1992).[14] [15] In this chapter, I use a rich longitudinal data set that allows me to calculate multi-year averages of parental income during childhood years and also of child income in adulthood, in order to minimise the potential bias from measurement error.

The second important issue to consider when estimating intergenerational elasticities relates to the life cycle bias, and is also a consequence of using snapshots of income or

[12] This transitory fluctuation is both due to transitory shocks and random measurement error (Solon, 1999).

[13] So long as the measurement error is uncorrelated across generations and the error and permanent income are uncorrelated.

[14] Examples of instruments used in the literature are indices of father's socioeconomic status (social class, occupation) and father's education. However, it would be important that the instruments do not explain children's earnings (i.e the exclusion restriction is valid), otherwise the IV results would be biased upwards (Dearden et al., 1997; Solon, 2002), generating an amplification bias.

[15] Mazumder (2005) and Haider and Solon (2006) point out that even estimates based on five-year averages of the earnings variable for fathers could still be subject to some attenuation bias. Under the classical measurement error model, this bias is minimised the more periods of data are used to obtain the averages. Mazumder (2005) shows with a simulation exercise that approximately 20 to 25 years of income data are needed to obtain a very good proxy for the permanent component of earnings with high reliability rate.

earnings data as a proxy for long-run economic status (Grawe, 2006; Haider and Solon, 2006; Jenkins, 1987; Nybom and Stuhler, 2016). This becomes a problem when these snapshots are not a good proxy for long-run outcomes because of the time of observation. Previous studies suggest that the relationship between current (observed) and long-run income varies greatly over the life cycle (Haider and Solon, 2006; Jenkins, 1987; Nybom and Stuhler, 2016). Thus, current income might not be a good proxy for long-run income, contingent on the age at which it is observed. Based on the model developed by Haider and Solon (2006), the income of children can be written as:

$$log\ {Y_{ia}}^{child} = \lambda_a\ log\ Y_i^{child} + u_{ia} \tag{2.2}$$

and for parents:

$$log\ {Y_{ia}}^{parent} = \mu_a\ log\ Y_i^{parent} + v_{ia} \tag{2.3}$$

where $log\ Y_i$ is the log of log-run income, which is not observed in practice and hence proxied by current income $log\ Y_{ia}$. The subscript a represents the age at which current income is observed.

In this model, λ_a and μ_a represent the source of the bias that arises when children's income (for λ_a) and parental income (for μ_a) at age a are not representative of their long-run economic status. These equations imply that the parameters λ_a and μ_a may vary with age. Indeed, Haider and Solon (2006) describe that the slopes λ_a and μ_a generally start at a value much lower than one in the early adult years and that they increase monotonically with age.[16]

Individual annual incomes tend to grow rapidly between the ages of 20 and 30, reach a maximum and flatten between the ages of 40 and 50 and decline thereafter (Corak, 2004). This pattern of growth is heterogeneous across individuals due to differences in income profiles, deviations from the average profile are usually correlated with both individual and

[16]For the US, Haider and Solon (2006) estimate that λ is as low as 0.2 for men before the age of 25.

family characteristics (Grawe, 2006; Haider and Solon, 2006; Nybom and Stuhler, 2016). For example, this could be related to investments in human capital. The classical example is of highly educated young individuals who experience a rapid growth of earnings after they complete education, but receive lower earnings in the beginning of their careers, at this point similar to the earnings observed for lesser educated individuals. If earnings/income are observed in early-career years, before the differences in the growth rate manifest, it is likely that the gap between low- and high-earners would be underestimated relative to what would be observed mid-working life. Thus, the degree of intergenerational persistence would also be underestimated. This is an important issue in practice, because it is likely that current income will be observed relatively later for parents and relatively earlier for children (Black and Devereux, 2011) in most data sets.

Since the relationship between the short-run proxies and long-run measures of economic status evolves over the life cycle and is age dependent, estimates of intergenerational elasticities will be sensitive to the age at which both children's and parents' incomes are observed in the data.[17] Haider and Solon (2006) demonstrate that the life cycle bias can explain the commonly observed pattern that estimates of IGEs tend to increase with the age of sons in the analytical sample.[18] Fortunately, the evidence suggests that the life cycle bias varies predictably across age (Grawe, 2006). In particular, using a long series of Swedish data containing nearly complete income histories of fathers and sons, Nybom and Stuhler (2016) derive a benchmark estimate of lifetime income and explore the extent of the life cycle bias from approximating lifetime incomes by annual incomes at different ages. The results from Haider and Solon (2006) and Nybom and Stuhler (2016) suggest that it is possible to minimise the extent of measurement error from life cycle effects when children's and parents' earnings

[17]Grawe (2006) highlights the importance of father's age when income is observed. He shows that intergenerational earnings persistence is negatively associated with the age at which father's earnings is observed. Assuming that sons are observed at some point in mid-life, we would observe a lower persistence (lower β) if parental income is observed at much older ages. Grawe argues that 20% of the variance in IGE estimates among studies using similar methodologies and data can be attributed to differences in fathers ages when income is observed. Here, to avoid this issue of observing the income of very old parents, I restrict the maximum age of parents to 65.

[18]Böhlmark and Lindquist (2006) apply the Haider and Solon model to Swedish data and find notably similar patterns.

and income are measured around the middle of the life cycle (between ages 30-40) and at a similar point for both generations. In this chapter, I use multiple income observations for parents measured around mid-life and for children's income measured no earlier than the age of 25. I also employ a two-stage residual approach to obtain improved measures of long-run income adjusted for time and age effects. These strategies should help alleviate the biases from both measurement error of observed income and life cycle effects.

Rank coefficients

Although intergenerational elasticities are linked to an established theoretical framework, these estimates suffer from further limitations. Recent studies have argued IGEs could be limited estimates of mobility because: (i) the relationship between parental and child income could be non-linear (or non-log linear); (ii) they can be sensitive to the treatment of children with zero income (Bratsberg et al., 2007; Chetty et al., 2014); (iii) to the use of different measures of income (Landersø and Heckman, 2017); and (iv) because they do not differentiate between upward and downward movements (Corak et al., 2014). Recent developments in this literature have proposed an alternative way to measure intergenerational mobility while taking into account some of these issues - the use of rank-based measures. This approach, first proposed by Dahl and DeLeire (2008) has become well-known after the influential study by Chetty et al. (2014) that uses rank-based estimates to compare income mobility across geographical areas in the US.

Rank-based measures focus on the analysis of the correlation between parents' and children's rank position in the distribution of income, instead of looking at the values of income variables directly. The ranks of parents and children are constructed separately, based on the respective income distributions of each generation. Then, the rank coefficient is estimated using the following equation:

$$Rank\ Y_i^{child} = \alpha + \gamma\, Rank\ Y_i^{parent} + u_i \tag{2.4}$$

The rank-rank slope, or rank coefficient (γ) represents the association between a child's rank in the income distribution and their parents' income rank. Similarly to IGEs, the rank coefficient is also a measure of relative mobility, however it captures solely the extent of re-ranking across generations without taking into account the spread of the income distribution. To better understand this property, the joint distribution of parents' and children's incomes can be separated into two components: (i) the joint distribution of their ranks, and (ii) the marginal distributions of parent's and children's incomes, reflecting the degree of inequality within each generation (Chetty et al., 2014). While IGEs combine both marginal and joint distributions capturing the extent of re-ranking across generations and the spread of the income distributions, rank-based measures focus solely on the re-ranking. As explained by Gregg et al. (2017), if the income distribution is represented by a ladder, re-ranking describes people switching rungs on the ladder and inequality describes how far apart the rungs of the ladder are.

The rank approach is a way of estimating mobility without having to assume a (log) linear relationship between parental and children's incomes (Dahl and DeLeire, 2008). In addition, evidence from other recent studies reveals that the rank-rank relationship is almost perfectly linear for multiple countries, that it allows for changes in inequality across generations (Bratberg et al., 2017) and that it is much less sensitive to different model specifications and to measurement error and life cycle bias, as it is scale invariant (Chetty et al., 2014; Gregg et al., 2017; Mazumder, 2015; Nybom and Stuhler, 2017). Following these new methodological developments, in this chapter I also estimate rank coefficients to complement the analysis of intergenerational elasticities. I then subject these estimates to the same robustness checks as the IGEs.

2.2.2 Current evidence: The UK in an international perspective

Estimates of intergenerational income and earnings mobility have been obtained for a set of countries for which longitudinal data is available, such as the US, Denmark, Sweden and

Germany and the UK. This body of evidence suggests that intergenerational mobility varies greatly across countries, and a number of studies have ranked the UK as a country with relatively low intergenerational mobility (Blanden, 2009; Corak, 2006; Jäntti et al., 2006; OECD, 2018*b*; Raaum et al., 2008).[19]

While cross-national comparisons can greatly contribute to the understanding of the mechanisms underlying the transmission of socio-economic outcomes between generations, comparing the estimates and the extent of intergenerational mobility across independent studies and countries often constitutes a challenging task. Several methodological and data choices are involved in such comparisons, making it important to question whether the observed cross-country differences arise from genuine mobility differences or are a product of different estimation methods or data incompatibilities (Jäntti et al., 2006). Some studies have taken up on the challenge to provide reliable comparisons of intergenerational mobility for different countries and most have focused on the variation and ranking of intergenerational elasticity estimates (IGEs).[20] Recently, the OECD (2018*b*) has provided an updated ranking of intergenerational earnings elasticities by country based on a number of individual country studies, illustrated in Figure 2.1.

Other studies estimate internationally comparable intergenerational elasticities directly for different countries using similar sources of data. This approach allows researchers to make similar methodological decisions and apply the same sample restrictions across data sets, making cross-country comparisons more credible. The papers by Jäntti et al. (2006) and Raaum et al. (2008) are examples of such studies. These authors compare the IGEs for the US, UK, Denmark, Sweden, Finland and Norway.[21] They find significant differences across countries, with the US being less mobile than the UK, which in its turn is less mobile than

[19]On the other hand, comparisons focusing on social class or occupational mobility classify the UK as a country with medium levels of mobility (Erikson and Goldthorpe, 1992).

[20]See, for example, Solon (2002) for a survey of independent studies for a sample of countries and their estimates on the earnings elasticities between pairs of fathers and sons; Corak (2004, 2006) for a survey of the existing literature on earnings mobility combined with a meta-analysis for the US; or Blanden (2009) for a broad review of the main elasticity estimates not only of earnings but also mobility of social class, status and educational outcomes across a sample of 65 countries.

[21]Raaum et al. (2008) is also one of the first studies to focus on measuring mobility considering women.

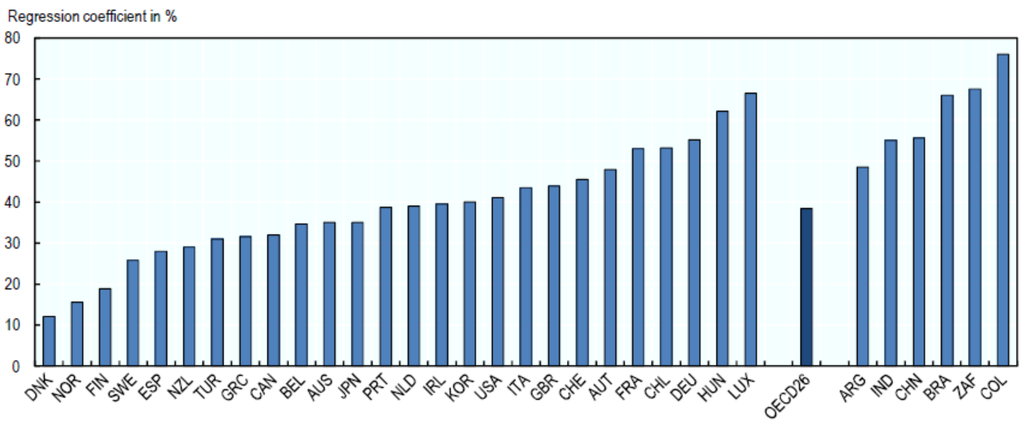


Figure 2.1 Intergenerational earnings elasticity for fathers and sons - various countries

This figure presents intergenerational earnings elasticity estimates for a sample of countries (late 2000s). Each bar represents the point estimate of the intergenerational earnings elasticity. A higher estimate means a higher persistence of earnings across generations, and a lower intergenerational mobility. Source: Figure taken from OECD (2018*b*)

the Nordic countries. Even in this more direct comparison, cross-country differences in the collection of data and definition of the key variables are unavoidable and remain a challenge.

Overall, the results from cross-national comparisons and the rank of countries illustrated in Figure 2.1 suggest that the level of intergenerational mobility in the UK is relatively low, especially in comparison to Scandinavian countries. However, the estimates for the UK reported in these cross-country studies vary substantially: Corak (2006) reports an estimate of intergenerational elasticity for son's earnings in the interval 0.43-0.55, Raaum et al. (2008) obtain an estimate of 0.41, Blanden (2009) estimates it at 0.37 and Jäntti et al. (2006) at 0.30. It is noteworthy that the current empirical evidence on intergenerational mobility for the UK is still relatively scarce and that “there is a lot of uncertainty for the UK" (Björklund and Jäntti, 2011, p.507). While the country's position in the cross-national rank is often based on a handful of selected studies that use cohort surveys to calculate mobility, in reality, the results produced by the literature are more diverse and often conflicting. The estimates of intergenerational persistence tend to vary with the dataset that has been used, with the cohorts studied, with the variable definitions and with the methods applied to the data.

The first study of intergenerational mobility in Britain using nationally representative survey data was undertaken by Dearden et al. (1997) using the National Child Development

Study (NCDS).[22] The NCDS is a longitudinal data set that follows a cohort of individuals born in Britain in one particular week in March 1958. Dearden, Machin and Reed examine the extent of intergenerational mobility between pairs of fathers and sons and fathers and daughters, in terms of labour earnings (weekly wages) and education (years of schooling) measured for the younger generation at the age of 33. For the older generation, only a single measure of father's earnings is available when children were aged 16. Other potential problems with the measure of parental earnings in the NCDS include that it is obtained by retrospective questioning (i.e. subject to recall error) and reported in bands (i.e lack of precision). In addition, the lack of income data over multiple periods means that the estimates from this study likely suffer from attenuation bias related to measurement error from various sources. To address this, the authors also use an instrumental variable (IV) approach with different variables as instruments, such as father's social class and father's education. However, they acknowledge that the estimates using IV are likely to be upwardly biased. The true value of the parameter likely lies somewhere between the estimates from the OLS and the IV approaches - they report IGEs in the wide range of 0.216-0.594 for sons and 0.351-0.695 for daughters.

Following this study, Blanden et al. (2004) estimate the intergenerational elasticity of earnings for the NCDS and for the British Cohort Survey (BCS), born in 1958 and 1970, respectively, with the aim of understanding how intergenerational mobility has changed over time and between these two cohorts. They observe a decline in intergenerational mobility for the younger cohort born in 1970 (i.e. the estimated IGE is higher). For sons, the estimated intergenerational elasticity of earnings is around 0.17 for NCDS and 0.26 for BCS. For daughters, they obtain an IGE of 0.17 for NCDS and 0.23 for the BCS.[23] They hypothesise that an important part of the fall in intergenerational mobility between the 1958 and 1970 cohorts is related to the unequal increase in educational attainment over this period, which mostly benefited children from richer parents.

[22]The very first evidence on intergenerational income mobility in Britain was presented by Atkinson using an original data set created from a household survey in York (Atkinson, 1980).

[23]Results presented in Blanden et al. (2004), earnings regressions.

The finding of a decline in intergenerational mobility between these two cohorts has created some controversy in the literature. Other research examining social class mobility has observed no evidence of changes between these cohorts using the same data (Erikson and Goldthorpe, 2010; Goldthorpe, 2013). The different results obtained when looking at earnings versus social class have been partly attributed to problems of comparability of the NCDS and BCS data and due to the data available on earnings. The variables used as a proxy for long-run economic status are not the same for the two cohorts. For the NCDS, separate (net) measures of father's earnings and mother's earnings when the child is aged 16 are used, and children's earnings were observed at one point, at age 33.[24] Conversely, the BCS only has data on parental earnings combined, and measured at two points in time, when children are aged 10 and 16. Children's earnings are observed once in adulthood, at age 30. In addition, in both data sets the sample has to be restricted to individuals at work when earnings are observed, i.e. working during the specific week when the interview was conducted.

Blanden et al. (2013) further investigate the factors behind these differences in the trends observed when examining social class or earnings mobility. They develop a framework that examines the relationship between father's social class, 'permanent' income and current income. This is applied to BHPS data to analyse how much of the variance in father's 'permanent' and current income is explained by their social class, by other common income predictors and by a residual.[25] This study finds that a considerable share of 'permanent' income is not explained by social class, which could explain the divergent results on trends from Blanden et al. (2004) and Goldthorpe (2013). They also emphasise that for the younger BCS cohort, the within-class income of fathers during childhood has become increasingly important to explain children's outcomes in adulthood and that the link between father's social class and family income has changed between the two cohorts, possibly because of an increasing role of mothers contributing financially to the household.

[24]In the NCDS, there is a single measure of father's and mother's earnings. In addition, these earnings are measured by net weekly earnings (wages) and only reported in bands, with no exact value being observed, which could increase the bias from measurement error.

[25]Other income predictors include parents' education, age and employment, housing tenure, financial difficulties and region (Blanden et al., 2013).

Apart from these methodological considerations, the existence of this declining trend in mobility has been questioned based on the argument that only two estimates might not provide enough evidence of a trend (Jäntti and Jenkins, 2015). As Goldthorpe (2013) points out, a ‘consensus view’ that social mobility has been in decline over recent decades has emerged. However, this is based on the data from these two cohorts, and is therefore open to question. In spite of these criticisms, the results from these studies have been widely used in the UK public policy debates about social mobility. However, not always with the appropriate interpretation - politicians often interpret these estimates as referring to absolute mobility, rather than relative mobility.

Other studies have used the British Household Panel Survey (BHPS) data to estimate intergenerational mobility. Ermisch and Francesconi (2004) use a father-child matched sample from the BHPS and are able to measure intergenerational mobility for individuals of different cohorts and backgrounds.[26] This was the first study to use the BHPS for intergenerational analysis, using the initial waves of data, from 1991-1999. Firstly, they estimate occupational mobility by measuring the association between the Hope-Goldthorpe score of occupational prestige of fathers and sons for a sample of individuals who provide retrospective information on their parents’ occupation when they were aged 14.[27] Then, using a matched sample of fathers and children who were all respondents in the survey, they turn to the study of mobility of earnings and income. Employing the traditional OLS method, their estimates of IGE for monthly earnings and annual income for sons are around 0.05. These estimates clearly suffer from the short period of data available (only the first 8 waves of the BHPS) and extremely young ages at which children’s earnings are observed, from the age of 16. Thus, it is very likely that these estimates of earnings mobility are affected by the life cycle bias. They also use an IV approach, with four different sets of instruments parent’s education, parent’s HG-Index, parent’s childhood family structure and parent’s local unemployment

[26] Ermisch and Francesconi (2004) focus on pairs of fathers and sons born between 1970 and 1983.

[27] The Hope-Goldthorpe index is based on a ranking of occupations obtained from a random sample of individuals interviewed in England and Wales in 1972. Ermisch and Francesconi (2004) argue that the HG-index is likely to be a good measure of permanent socio-economic status, as it is highly correlated with earnings and thought to be relatively stable over the working life.

rate.[28] The IV regressions yield IGEs of around 0.10 for the first two instruments and 0.20 for the third and fourth instruments. Even though mobility is estimated for a group of individuals from multiple birth cohorts, the wide-ranging results presented by Ermisch and Francesconi (2004) do not allow us to draw strong conclusions about the current levels of intergenerational mobility in the UK for cohorts born after 1970.

Also using the BHPS, Nicoletti and Ermisch (2008) extend this analysis of intergenerational earnings mobility using data from 1991 to 2003. Since there is no data on both fathers' and sons' earnings for all these cohorts in Britain, they attempt to overcome this lack of information by combining two samples from the BHPS and estimating the elasticities by two-sample two-stage least squares (TS2SLS).[29] Their main sample is comprised of sons born between 1950-1972, who are employed and have at least one observation of earnings between the ages of 31 and 45 and whose fathers were born between 1918-1949 and were aged between 31 and 55 when their sons were 14 years old. This first sample contains information on sons' earnings and a set of educational and occupational characteristics of their fathers collected through retrospective questions. The second sample, the supplemental sample, is given by all men born between 1923-1946, who the authors argue should be a representative sample for the potential fathers, with observations on earnings, education, age and occupational characteristics. They combine the two samples using the TS2SLS method. Firstly, they use the second (supplemental) sample to estimate a log earnings equation for the potential fathers, using their age, education and occupational information as explanatory variables. Then, to estimate the IGE, the best linear predictor of fathers' earnings is plugged into the traditional OLS equation together with the information on sons' earnings from the first sample. The reported IGEs by cohort for single year earnings are estimated to be in the range 0.20-0.30

[28] Some of these instruments, such as parental education and occupation are very likely to be correlated with sons' earnings, which probably creates an upward bias on these estimates, providing an upper bound of IGE.

[29] The TS2SLS approach is often used when there is no information on parental earnings/income in the data set (Jäntti and Jenkins, 2015; OECD, 2018*b*). First used in the context of intergenerational mobility by Björklund and Jäntti (1997), the method consists of using a second sample to predict earnings/income for the parental generation. Thus, it is based on a sample of children including information on their income distribution and key predictors of parental income and another sample, of parents, which contains information on the unconditional distribution of income in that generation.

(Nicoletti and Ermisch, 2008). Their main results suggest that there are no significant changes in intergenerational mobility in Britain between 1950 and 1960 and a small decline in mobility between 1961-1972. Due to the nature of the income predictors used in this analysis, it is likely that they are positively correlated with children's earnings indirectly, through fathers' earnings, but also directly. If this is the case, the TS2SLS estimator would be biased upwards, and mobility is underestimated.

In another recent study, Gregg et al. (2017) present revised IGE estimates for the NCDS and BCS cohorts. Gregg, Macmillan and Vittori attempt to minimise both life cycle and attenuation bias by using the two measures of parental income available in the BCS study (at ages 10 and 16) and estimating intergenerational mobility at various points along the life cycle. They show that the use of two observations of parental income rather than one reduces the attenuation bias for the BCS estimates.[30] Examining mobility at different ages, they also show that for both cohorts the IGE estimates start very low during the early twenties and rise constantly until the individuals (children) are in their mid-forties.[31] Moreover, Gregg et al. (2017) were the first to calculate rank coefficients for the UK as a complement to the more traditional IGE measures. They note that the rank coefficients follow a similar pattern to the IGEs across the life cycle. However, their results suggest that rank coefficients seem less attenuated than IGEs at lower ages and also less affected by attenuation bias from measurement error or transitory shocks in the parental income variable. Based on this, the authors suggest that rank estimates might be more appropriate when income is observed at early ages. Overall, their best revised IGE estimates for the UK are 0.25 for the NCDS cohort and 0.43 for the BCS cohort.[32]

All the aforementioned studies focused on measuring intergenerational mobility considering labour market earnings for either the second generation, or for both. The shift in the literature to study mobility with respect to family income in the two generations is relatively

[30] Due to lack of multiple observations of parental income in the NCDS they could not present comparable estimates for this cohort.

[31] For the NCDS cohort, IGEs go from 0.042 at age 23 to 0.259 at age 46. For the BCS, they go from 0.203 at age 26 to 0.397 at age 42.

[32] The correspondent rank coefficients are 0.195 for the NCDS and 0.298 for the BCS.

recent. The emphasis on capturing all available childhood resources makes sense, if the goal is to capture the association between the living standards of parents and their children. When moving from studying the association in earnings to analysing the links between family or household incomes, it is important to acknowledge the role of the process of partnering and assortative mating, which will influence the living standards of families.

Belfield, Crawford, Greaves, Gregg and Macmillan (2017) measure the intergenerational transmission of living standards, by estimating the association in net family income between parents and children. Using data from the NCDS and BCS, the authors show how their estimates change when using different definitions of sons' income and earnings - sons' individual gross earnings, gross private income and net family income - while holding the variable for parental income constant as net family income.[33] [34] For the NCDS cohort, using sons' earnings, the authors obtain an IGE of 0.22. Using net family income, the IGE is estimated as 0.17. Using gross private income, the IGE is estimated as 0.20. For the BCS cohort, the authors obtain an IGE of around 0.36 when using sons' earnings. For gross private income, they obtain an IGE of 0.37 and for net family income, they obtain an IGE of 0.31. Using these different measures of income for the second generation, this study emphasises the importance of understanding what exactly is the nature of association being measured and demonstrates that IGE estimates might be sensitive to the choice of income variable. In the international literature, this sensitivity has also been discussed by Landersø and Heckman (2017).

It becomes clearer after this more detailed examination of the empirical evidence on intergenerational mobility for the UK that comparing estimates across studies can be a challenge due to the variety of methods, datasets and variables being used. This literature has certainly developed in the recent years, following new methods and estimation strategies, but

[33] In this model, they use a one-point parental income observation when children were age 16 and sons' income or earnings is captured when they were 42 years old.

[34] Net family income is the closest measure to household income, although not exactly the same - in this paper, Belfield and colleagues create continuous measures from imputing incomes for each band observed in the data, for 'usual family income' reported by the parents in the BCS and a sum of father's earnings, mother's earnings and other household income in the NCDS.

many parts of it remain unexploited. As Jäntti and Jenkins put it, “a key conclusion that we draw about the UK debate [...] is that much richer data than those provided by the NCDS and BCS cohort studies are needed to draw firm conclusions about the level and trend in UK income mobility." (Jäntti and Jenkins, 2015, p. 911).

2.3 Data and Methodology

The availability of new data with the BHPS for 1991-2008 and Understanding Society for 2009-2017 opens up new research avenues for the examination of the extent of intergenerational income mobility, and enables the estimation of income mobility for younger cohorts of individuals. To my knowledge, no other studies have measured intergenerational income mobility using this harmonised dataset. In this section, I discuss the main characteristics of the data and the estimation strategy.

2.3.1 Data

As noted before, studies on intergenerational mobility have very strict data requirements, a factor that has contributed to the existing empirical evidence for the UK being rather limited. Earlier studies that have used the NCDS and BCS cohort studies have faced a series of limitations related to their focus in one single cohort of children and only having one or two observations of income or earnings for parents and children. Within this chapter, I make use of two household longitudinal surveys that collect income data across generations: the British Household Panel Survey (BHPS) and the UK Household Longitudinal Survey (UKHLS), also referred to as Understanding Society.

The BHPS is a panel survey of households in the UK. Having started in 1991, the BHPS interviewed a nationally representative sample of 5,500 households and 10,300 individual respondents. Every year, the same individuals have been re-interviewed. Even though there is some sample depletion through attrition, responses rates are high and the sample is frequently replenished with the addition of new households and individuals that join participant households. If individuals leave the original household to form new households, all adult members (aged 16+) in this household are added to the survey. In addition, children of the original households are also interviewed once they reach the age of 16. The BHPS is comprised of 18 waves of data, spanning from 1991 to 2008.

Complementing the BHPS, I also use data from Understanding Society, the UK Household Longitudinal Study (UKHLS). The two studies have been recently adapted and harmonised, and as such provide an opportunity to increase the main sample as well as to extend the analysis to more recent years. The UKLHS data collection started in 2008 and this data provides a larger and more wide-ranging continuation of the BHPS. From its second wave, the Understanding Society main study additionally includes information collected for continuing participants of the BHPS. Of around 8,000 BHPS participants invited to join, almost 6,700 accepted the offer and are being interviewed in Understanding Society every year since 2010. The identification of BHPS participants in the UKHLS is possible through their unique person identifier.

The long harmonised panel BHPS + UKHLS is particularly suited to the estimation of intergenerational mobility. Firstly, differently to the NCDS and BCS data, children and parents come from multiple heterogeneous cohorts. Secondly, the longitudinal nature of this dataset makes it possible to link children to their parents and obtain the relevant variables directly from these individuals, as opposed to relying on variables originated from retrospective questions that could be affected by recall error. Another second advantage of this data is that the information on income for parents and children comes from multiple years and is not restricted to a single time-point measure of income. Considering all 27 available waves of data (years), the harmonised data set is sufficiently long to calculate an approximation of parental permanent income during childhood for a sub-sample of young individuals who can be linked to at least one parent within the survey. Thirdly, the harmonised data allows us to observe directly the income of sons and daughters as they become adults. This is a very important point due to estimation concerns related to the life cycle bias, as described before. Earlier studies (e.g. Ermisch and Francesconi, 2004) that have used the BHPS were restricted to observing the income of children from the age of 16, arguably too young to obtain a sensible proxy of permanent income for these individuals.

The use of the harmonised BHPS and UKHLS will allow me complement and expand the analysis of the intergenerational mobility of income in the UK started by researchers who previously used either or both of the two British cohort studies. Using this rich data, I can observe individuals from multiple cohorts and analyse the sensitivity of my parameters of interest (IGEs and rank coefficients) to a series of changes in the model specifications, sample restrictions and income measures. This also represents an addition of at least 10 years of data since the closest related studies that have used the BHPS, by Ermisch and Francesconi (2004) and Nicoletti and Ermisch (2008).

2.3.2 Key variables

Following recent studies in the intergenerational literature (Chetty et al., 2014; Lee and Solon, 2009; Mayer and Lopoo, 2004), the main variable of interest throughout this chapter is gross (pre-tax) household income in the month before the interview.[35]

Parental income is the gross household income captured for parents during childhood years (when children are aged 0-18). When a child lives with just one biological parent, parental household income equals that of that parent, regardless of their marriage status. A restriction is imposed so that to be in the matched sample, children must have lived in the same household with at least one of their biological parents, for at least one of the childhood years. This restriction means that children must be no older than 18 years old in 1991, that is, they have to be born on/after 1973. Moreover, in order to avoid potential confounding effects due to retirement, parental income is only captured between the ages of 25-65.

[35] Here, as done Chetty et al. (2014), Heidrich (2017) and others, I focus on pre-tax income and thus the role of redistribution by taxes is not considered. However, if taxes do not generate rank reversals, using post-tax instead of pre-tax would not change the rank-based results in a significant way.

Child income is the gross household income of the child when they become adults, measured at age 25 or above. The choice of 25 as the cut-off minimum age for the income observation relates to this being the age at which most young individuals will have already left education and entered the labour market. This classification is also used by the OECD labour markets statistics, which considers individuals to be 'young' until the age of 24 (OECD, 2018*c*). In order to have at least one observation for income when children are 25 and older, another restriction is imposed on the sample: only individuals born on/before 1992 are included. As a robustness check, further in the chapter I examine the implications of varying these age restrictions in the sample.

Figure 2.2 illustrates how the main income variables are captured over the lifetime of individuals (the children) in my sample.

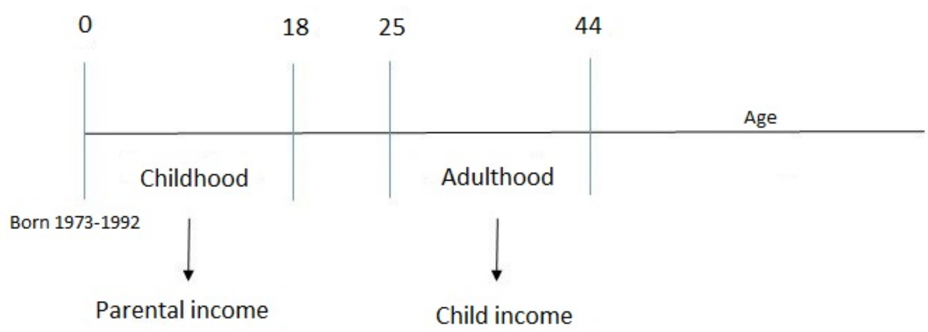


Figure 2.2 Observation of the key variables over the life cycle

This diagram shows how the main variables for this analysis are collected over the life cycle. Source: Own elaboration.

The shift to using family or household income has been widely discussed in the literature. Household income is considered a good indicator of broad economic welfare, living standards, and overall socio-economic status (Chadwick and Solon, 2002; Chetty et al., 2014; Lee and Solon, 2009). Hence, the use of parents' household income as opposed to parental individual income or earnings is supported by the attempt to measure the influence of all resources available during childhood to the outcomes of young adults. Using household income for the second generation also highlights the relationship between childhood resources for the overall living standards of this generation. This is especially important to account for the role of pooling of resources within the household, and how its members may benefit from economies of scale and common goods. As a robustness check, I also examine the implications of choosing a different measure of children's income. More detailed information about the exact definition of the income variables can be found in Appendix A.2.

The main income variables used in this analysis have been defined in the same way in the harmonised BHPS and UKHLS, and are collected every year from individual respondents using the same question. All income variables used in this analysis reflect pre-tax income, are measured in GBP and deflated using the Consumer Prices Index (CPI) of each respective interview year, with 2015 as the base year.

Finally, it is important to note that all age variables used as controls in the main model specifications throughout the analysis reflect the age at the time of interview and when the relevant income variable is observed. Regarding parental age, when the child lives with both biological parents, parental age is equal to father's age, as fathers are heads of household in over 94% of these households. If the child lives with only one biological parent, the age of this parent is considered.

2.3.3 The matched sample

In the BHPS+UKHLS dataset, it is possible to match individuals (children) to their parents using cross-wave personal identifiers. As explained before, this allows me to obtain direct information on individual characteristics and incomes of children as adults and of their parents as these children were growing up (childhood years). Taking into account the age requirements necessary to obtain the income data, individuals need to meet the following conditions (C) in order to be included in the final matched sample:

C1) need to be born between 1973 and 1992;

C2) can be matched to at least one of their biological parents within the BHPS;

C3) income is observed at least once in adulthood (after turning 25 years of age); and

C4) parental income is observed at least once during childhood years (before age 18).

After imposing conditions C1 and C2, 5505 individuals are matched to at least one biological parent and meet the birth cohort requirements for the analysis, of which 89% are original BHPS sample members. To obtain the income data for the two generations, I restrict the sample to those children with at least one observation of income in adulthood and with some parental income observed during childhood years. Applying these restrictions leads to a drop in the effective sample, since not all children have been followed into adulthood.[36] The final sample after imposing restrictions C1-C4 is comprised of 2126 individuals matched to their parents, which corresponds to 39% of individuals from the eligible birth cohorts matched to their biological parents and 80% of these who have been interviewed at ages 25 and over. For these matched pairs, I rely on a total of 15631 observations of income for the parents and 13479 for the children in adulthood. This is the sample used throughout the remainder of this chapter. The main source of attrition after imposing conditions C3 and C4 is losing the children as they get older and potentially move out of the original (parents') household. Following individuals over a long period of time is the major challenge here, which is common to most longitudinal studies, including the previous UK studies using the NCDS and BCS. This comes with concerns that the data may not be fully representative of the target population. I examine the representativeness of the analytical matched sample in Appendix A.1 - this investigation reveals that the final sample remained broadly representative.

2.3.4 Descriptive statistics

Table 2.1 presents the summary statistics for the main variables used in the intergenerational income analysis of the BHPS + UKHLS. Panel A refers to parental characteristics and Panel

[36]The large drop in sample size occurs after imposing condition C3, which restricts the sample to 2636 matched pairs.

B refers to the characteristics of sons and daughters in the matched sample. The summary statistics reveal that, on average, children have higher real household income in adulthood than their parents, which reflects the improvement of living standards across generations in absolute terms. According to recent reports, upward mobility in absolute terms has been occurring in most OECD and emerging countries (OECD, 2018*b*). This table also reveals differences in the ages at which parental and child income are observed - parents are on average 12 years older when their income is observed. In an attempt to account for these differences in the age of income observation across generations, I include flexible controls for age in my regression models, as described in the next section. Finally, I also present the summary statistics for alternative measures of child income that will be used later in this chapter. Average personal income for the children is, by definition, higher than (or at least equal to) labour income because it also encompasses income from non-labour sources. More information on the definition of the different income variables can be found in Appendix A.2.

Table 2.1 Summary statistics - key variables

Panel A: Parents (N=2126)	Obs	Mean	SD	Min	Max
Parental household income	15631	3616.6	2717.5	15.3	102426.2
Parental age when income observed	15631	42.0	7.0	25	65
Panel B: Children (N=2126)	Obs	Mean	SD	Min	Max
Adult household income	13479	4147.0	2427.2	1.7	41794.4
Adult personal income	13420	1972.8	1381.9	0	18337.4
Adult labour earnings	13412	1733.1	1439.8	0	18337.4
Age when household income observed	13479	30.1	4.4	25	44
Child age when parental income observed	15631	13.0	4.2	0	18

Notes: Unweighted summary statistics based on the main matched sample of 2126 individuals for whom all of the variables are defined. Children's personal income and labour income refer to N=2107, which corresponds to a smaller number of observations. All income variables are pre-tax income measured in real pounds per month and constructed as described in the text in section 2.3.2. Source: BHPS and Understanding Society.

2.3.5 Estimation strategy

It may be useful at this point to reiterate that the main aim of this chapter is to assess the degree of intergenerational income mobility in the UK by estimating IGEs and rank coefficients for a cohort of young individuals born between 1973 and 1992, using the harmonised BHPS and UKHLS data. Furthermore, this study is concerned with whether the estimated coefficients are robust to changes in the model specifications, variable definitions and sample restrictions.

In this section, I present and discuss the methodology. Initially, I employ the traditional Ordinary Least Squares (OLS) approach to estimate the intergenerational income elasticity and to offer comparability with previous studies that have applied this method. Then, a novel two-step residual approach (TSRA) is proposed to estimate the intergenerational elasticities while better controlling for the presence of life cycle effects. Thirdly, I estimate rank coefficients, a complementary measure of intergenerational mobility that characterises mobility from a positional perspective, using both OLS and TSRA.

Traditional OLS approach

Based on the traditional approach used in the literature on intergenerational mobility (Black and Devereux, 2011; Jäntti and Jenkins, 2015; Solon, 2002), I estimate the intergenerational income elasticity using the following equation:

$$log\ (\overline{Y_i^{child}}) = \alpha + \beta\ log(\overline{Y_i^{parent}}) + X_i'\theta + \varepsilon_i \tag{2.5}$$

where $\overline{Y_i^{child}}$ is the multi-year average of all the observations of real income for individual i in adulthood, and $\overline{Y_i^{parent}}$ is the multi-year average of all the available income observations for parental real income during individual i's childhood years. The coefficient β from this regression is the intergenerational income elasticity. X is a vector of control variables at the individual level, including the average age of children and of the main parent when income is observed, their average age squared, as well as birth cohort dummies for both generations, which capture cohort effects. All regressions use robust standard errors, clustered at the parental level, to allow for arbitrary correlation within families, since there are a few individuals matched to the same parents in the sample. The idea behind this clustering is that if parents have more than one child, the error term for these individuals (children) could be cross-sectionally dependent.

The elasticity estimates β based on this model provide a comparative benchmark to the findings of previous studies that use a similar methodology. In order to address the measurement error problem from using short-run proxies for long-run income, parental income is taken as the multi-year average of observed household incomes during childhood, that is, average household income when children are aged 0-18. On average, parental income is observed in 6.9 childhood years in my matched sample. Therefore, the measure of approximate long-run parental income is a product of multiple observations of income over childhood years rather than being a single point in time measure. This approach reduces the attenuation bias from transitory shocks to 'permanent' income and is an important advantage of using the BHPS data as compared to the cohort studies. For example, previous studies using the NCDS rely on a single measure of parental income during childhood,which is measured in income bands. Similarly, studies that use the BCS only have two available observations of parental income during childhood, at ages 10 and 16.

Further, child income in adulthood is also taken as a multi-year average, this time of all observations of income after the age of 25. Although under the classical measurement error model assumptions, the main source of attenuation bias arises from measuring the independent variables with error, using a multi-year average of child income is also useful because it reduces the year-to-year variability of income and the influence of episodes of very low income, as well as contributing to attenuate the life cycle bias (Nybom and Stuhler, 2016). The potential life cycle bias from observing incomes of children and parents at different and non-comparable ages is addressed by controlling for a function of age and restricting the minimum age at which income is observed for both generations.[37] Arguably, the cut-off age of 25 can still be considered relatively young in terms of obtaining a good proxy for long-run income in adulthood and could potentially remain a cause of underestimation of the

[37]For parents, I also restrict the maximum age to 65.

intergenerational income elasticity.[38] Later in this chapter, I examine the sensitivity of my results to changes in the children's age restrictions in the sample.

Two-stage residual approach (TSRA)

In addition to the traditional OLS approach, I propose the use of a two-stage residual approach (TSRA) to obtain more accurate estimates of long-run income for both generations using all the available data.

Due to the temporal generational structure of this long panel data, I observe parental income and child income across a varying number of years and at different time periods. The income observations are temporally staggered and need to be collapsed to allow the use of traditional regression-based analyses. In the traditional OLS regression-to-the mean model, this issue of collapsing the observations for each individual is addressed by averaging the incomes (for parents and children) and also averaging the age of parents and children at which income is observed. This methodology has been adopted by most studies in the intergenerational literature and is also what I utilise to obtain my first estimates of IGE. However, this way of collapsing or aggregating the multiple observations on income and age for each individual means that specific information on the age-income profiles of individuals could get lost in the process.

The relationship between the current observed income and 'permanent' income changes over the life cycle (Haider and Solon, 2006). In addition, there is evidence of heterogeneity in the individuals' age-income profile (Nybom and Stuhler, 2016), which reflects the development of income at different ages. Because of the existence of such age-income profiles, the accuracy and the meaningfulness of the averaged current income to represent long-run income will depend on the ages at which current income is observed. For example, if current income is observed and averaged during the period of lower levels and fast growth, such as in early-career years, it is probable that the obtained average is understating lifetime income. In

[38] Note that for both generations the adequate income measures comprise of the log of multi-year averages, not the multi-year average of log incomes.

my data, income in adulthood is first observed when individuals are 25 years old, and so this might be a relevant issue.

The new methodology proposed in this chapter, the two-stage residual approach or TSRA, aims to address this issue by using all the available data on income and *when* it is observed. It provides a more flexible way of collapsing the multiple observations of current income, considering the exact age and time at which it is observed. In practice, in the first stage of TSRA, I apply least squares (separately for each generation) to a regression of current (observed) income on a function of age and year dummies, as shown in equations 2.7 and 2.8. The remainder is a residual capturing any variance in income that is not explained by age and time. The residuals from these auxiliary regressions are used to generate a measure of income adjusted by age and year effects, which could serve as a better proxy of long-run income because age and time-related differences between current income and long-run income will be conditioned out.

TSRA First Stage: Auxiliary Regressions

The general form of a first stage equation is:

$$log\ Y_{it} = f(age_{it}, time_t) + u_{it} \tag{2.6}$$

For parents, the estimated equation is:

$$log\ Y_{it}^{parent} = \alpha + \zeta_1 P_{it} + \zeta_2 P_{it}^2 + \zeta_3 P_{it}^3 + \zeta_4 P_{it}^4 + \delta' D_t + w_{it} \tag{2.7}$$

where the vector D_t contains year dummy variables for each $t = 1991, 1992..., 2008$. The regression also includes control for a quartic in parental age P_{it} at the time parental income is observed during childhood.

For children, the first stage equation is:

$$log\ Y_{it}^{child} = \alpha + \theta_1 A_{it} + \theta_2 A_{it}^2 + \theta_3 A_{it}^3 + \theta_4 A_{it}^4 + \delta' D_t + v_{it} \tag{2.8}$$

where the vector D_t contains year dummy variables for each $t = 1998, 1999..., 2017$. I also control for a quartic in age A_{it} at the time children's income is observed in adulthood.

In the second stage of TSRA, the residuals are averaged for each individual. I then use this adjusted income in the subsequent analysis to derive the measures of intergenerational mobility, as shown in equation 2.9. The underlying idea behind this alternative methodological approach is to examine the intergenerational association between the age- and time-adjusted incomes of parents and children. This allows me to employ all the information available in the data, reducing the bias from observing incomes of parents and children at different stages of life.

TSRA Second Stage: Estimating Income Mobility

The second stage equation is:

$$\overline{\hat{Y}}_{res_i}^{child} = \alpha + \beta\ \overline{\hat{Y}}_{res_i}^{parent} + e_i \tag{2.9}$$

where $\overline{\hat{Y}}_{res_i}^{child} = \overline{v_i}$ and $\overline{\hat{Y}}_{res_i}^{parent} = \overline{w_i}$. β is the main coefficient of interest, the intergenerational income elasticity.

Age and cohort effects cannot be separately defined without imposing strong functional form restrictions (Lee and Solon, 2009). For example, at age 25 comparisons between individuals born in 1973 and 1974 are not necessarily the same as comparisons between the years 1998 and 1999. In this chapter, I opt to include year dummies in the first stage auxiliary regressions, recognising that while controlling for age, these effects are similar to cohort effects.[39] In any case, δ should be interpreted as reflecting a combination of time and cohort effects.

[39] A similar approach is used by Lee and Solon (2009).

It is important to note that when employing the TSRA method, the main variables used in the second stage are residuals (age- and time-adjusted income), which are generated in the first stage. When using generated regressors, it is important to adjust the standard errors to account for the sampling variance of generated variables constructed in the first stage (Pagan, 1984). Therefore, as well as reporting the usual robust standard errors clustered at the parental level, I also report the adjusted standard errors calculated using a bootstrap procedure adapted to TSRA. First, bootstrap samples of parents and children are drawn to obtain adjusted incomes in the first-stage regressions. Secondly, a bootstrap sample of adjusted incomes is drawn and used for running the second-stage regression to obtain estimates of mobility. This process is repeated 1999 times and the bootstrapped standard error is estimated by the deviation of the distribution of the bootstrap estimates. Further information on this bootstrap adjustment can be found in Appendix A.3.

A similar residual approach is used in the literature that studies *intragenerational* mobility trends, more specifically, the research on the instability of individual (especially men's) earnings. These studies first run regressions of earnings controlling for differences in age, education and work experience and then use the earnings residuals in a second stage (Jäntti and Jenkins, 2015). For example, Shin and Solon (2011) use a residualised measure of the change in log earnings by regressing, in the first stage, log earnings on a function of age. In another study, Moffitt and Gottschalk (2012) use a similar method by regressing log earnings on education, age polynomials and interactions between them. In the second stage, they compare the earnings residuals for the same individuals at different points in time. Furthermore, some earlier studies measuring earnings mobility refer to the use of a residual approach to obtain a measure of 'permanent' status adjusted for age effects (Couch and Lillard, 1998; Dearden et al., 1997). However, the use of this nomenclature was employed to denote the inclusion of age controls as independent variables in the traditional OLS regression. Nonetheless, to the best of my knowledge, this is the first time that this two-stage approach has been used to obtain estimates of intergenerational income mobility for the UK.

Rank coefficients

Finally, moving beyond the estimation of intergenerational elasticities, I also estimate rank coefficients of income mobility. As discussed previously. the IGE combines both marginal and joint distributions of parents' and children's incomes, capturing both the extent of re-ranking across generations and the spread of the distributions. Meanwhile, when estimating rank coefficients, the focus is on the re-ranking across generations, and the spread of the distribution is standardised.

The original ranks are obtained by ranking parents with respect to the income distribution of parents in the sample and children with respect to the income distribution of children.[40] In addition, it is useful to standardise and transform these ranks from absolute to a relative rank – the most common approach in this literature is to transform them into percentile ranks. The rank-rank coefficient (γ) is then obtained from:

$$Rank\ \overline{Y_i^{child}} = \alpha + \gamma\ Rank\ \overline{Y_i^{parent}} + X_i'\theta + \varepsilon_i \tag{2.10}$$

where the percentile rank for each individual i on the left-hand side is obtained from ranking all children based on the averaged child income ($\overline{Y_i^{child}}$) over multiple periods. The same is done for the rank of parents, which are based on the averaged parent income ($\overline{Y_i^{parents}}$). Similar to the equation used to obtain intergenerational elasticities (equation 2.5), X is a vector that contains individual controls for the average age of parents and children and birth year dummies.

Analogously, I use the TSRA method to obtain the rank coefficient from the relationship between children's and parent's ranks in terms of the age- and time-adjusted incomes. In this case, the second stage regression of TSRA used to obtain the rank coefficient can be written as:

[40] The results were unaffected by ranking parents among all parents matched to their children with income information available during childhood years and also when ranking children among all individuals matched to their parents with some income available in adulthood.

$$Rank\ \overline{\hat{Y}}^{child}_{res_i} = \alpha + \gamma\ Rank\ \overline{\hat{Y}}^{parent}_{res_i} + n_i \tag{2.11}$$

where the percentile rank of individual i is obtained by ranking all children based on their average adjusted income ($\overline{\hat{Y}}^{child}_{res_i}$). Similarly, parents are ranked with respect to their average adjusted income ($\overline{\hat{Y}}^{parent}_{res_i}$).

2.4 Results

2.4.1 Estimates of intergenerational income mobility in the UK

I start by examining the degree of income mobility in the UK by documenting the traditional intergenerational income elasticities (IGE). Panel A of Table 2.2 presents the IGE estimates obtained using both OLS and TSRA methods. Focusing on the first row, Model 1 shows the OLS results with no additional controls, while Model 2 includes controls for age and birth cohort dummies for both generations.[41] Again, the dependent variable, child income, is the log of averaged household income for individuals measured at/after the age of 25. The main independent variable, parental income, is the log of averaged parental household income observed during childhood years. The OLS results using Model 2 indicate that the intergenerational income elasticity for the UK at the national level is around 0.269 and it is precisely estimated. This can be interpreted as: an additional 10% in parental household income advantage would give children a 2.69% income advantage as adults.

Panel A also presents the IGE estimates obtained using TSRA. As discussed previously in section 2.3.5, this method is proposed as an alternative to the traditional OLS specification and consists of a two-step estimation procedure. In the first stage, the income variables are age- and time-adjusted using separate auxiliary regressions for each generation. The first

[41] Here, I present the results including controls for the average age and average age squared of parents and children. However, identical results were obtained when controlling for a quartic function of the average age of parents and children.

stage auxiliary regressions for the model with full controls are presented in Table A.4 in the Appendix. In the second stage, these residuals are then averaged for each individual and parent and used as the key variables in the main regression. Parents' averaged residuals will be called adjusted parental income and, similarly, children's averaged residuals will be called adjusted child income. The new IGEs are obtained by regressing the age- and time- adjusted child income on the adjusted parental income. While the OLS approach only allows me to control for the average age at which income is observed and for birth cohort dummies, the TSRA allows me to control more flexibly for the exact age and year at which incomes are observed in the data and use of the full information available. As a result, this likely presents an improved measure of long-run income. The second row of Table 2.2 shows that using the TSRA the IGE is estimated to be 0.255, a slight decrease from the coefficient estimated using OLS. Table 2.2 also reports the bootstrapped standard errors computed following the bootstrap procedure described in Appendix A.3 for the TSRA method. The bootstrapped standard errors are slightly smaller than the robust standard errors. This is because both stages of TSRA are bootstrapped simultaneously, therefore the bootstrap samples are drawn from all available observations of income.[42]

Combining the two sets of results obtained by OLS and TSRA, the intergenerational income elasticity for the UK is estimated to be 0.255 to 0.269. These values reflect my preferred estimate of IGE for individuals born between 1973-1992. How these results compare to previous estimates for the UK is discussed in later in section 2.5.

Panel B of Table 2.2 presents the estimates of rank coefficients obtained by ranking children and parents in terms of their position in the income distribution of each generation and then regressing the percentile rank of the children on the percentile rank of the parents. Parents and children are ranked based on their household income among the final sample of 2126 matched parent-child pairs. When using OLS, parents and children are ranked within each generation based on their average household income over the period, while for TSRA

[42]For estimates based on household income, this corresponds to 15631 observations of parental income and 13479 observations of child income.

Table 2.2 Estimates of the intergenerational income elasticity (IGE) and rank coefficient in the UK

	Method	Model (1)	Model (2)
Panel A: IGE			
β	OLS	0.292***	0.269***
		(0.027)	(0.028)
β	TSRA	-	0.255***
			(0.027)
			[0.021]
Panel B: Rank Coefficient			
γ	OLS	0.323***	0.301***
		(0.023)	(0.024)
γ	TSRA	-	0.291***
			(0.023)
			[0.018]
N		2126	2126

Notes: Robust clustered standard errors in parentheses. Bootstrapped standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Child income and parental income are measures of household income, constructed as described in the text in section 2.3.2. Model (1) estimated by OLS does not include any additional controls. Model (2) estimated by OLS controls for the average age of parents and children when their income is observed, their average age squared and birth cohort dummies for both generations. Model (2) estimated by TSRA controls for a quartic in age and year dummies in the first stage regressions for each generation. The TSRA standard errors have been corrected using bootstrapping techniques, as described in Appendix A.3. Source: BHPS and Understanding Society.

they are ranked based on the age- and time-adjusted household income. As displayed in this table, I estimate the rank coefficient as 0.291-0.301. This means that, on average, a 10 percentile point increase in the rank of the parents would translate in a 3 percentile point increase in the rank of income for children.

For both rank coefficients and income elasticities, it is notable that accounting for the age at the time income is observed slightly decreases the point estimates. As discussed previously, the relationship between observed and long-run income varies over the life cycle (Grawe, 2006; Haider and Solon, 2006; Nybom and Stuhler, 2016). The literature describes that predicting the direction of this bias can be a challenge (Haider and Solon, 2006; Jenkins, 1987), thus reinforcing the importance of adequately accounting for possible age differences at the time income is observed. The great similarity between the results from OLS and TSRA suggests that they perform similarly in this sample at addressing the life cycle bias. The main advantage of the TSRA is using all the information available in the data. The slightly smaller estimates from TSRA reflect this feature, as this model controls for year effects (at which income is observed in the data) in addition to age and cohort effects, while OLS only accounts for the latter.

Overall, these results indicate a similar (and considerable) degree of intergenerational income persistence when looking at elasticities and rank coefficients. The slightly higher values of rank coefficients compared to the IGEs, are likely a consequence of the ratio between the standard deviations of the household income distributions of both generations. In contrast to the intergenerational elasticity, the rank coefficient is not affected by changes in the variance of the distribution between generations.

It is important to acknowledge that despite all efforts to obtain the most accurate estimates of income mobility, it is possible that the estimates obtained are still subject to some residual attenuation from measurement error. In addition, even with the flexible age controls included in the models, children in the data are, on average, relatively younger than their parents when their income is observed. This means that the observations of income could provide a poorer proxy for the long-run income for children than it does for parents. Therefore, the estimates of mobility could be also subject to a residual life cycle bias (Haider and Solon, 2006; Nybom and Stuhler, 2016). However, it is likely that both these issues could only be completely solved if the complete income histories of the two generations were observed in the data. While not perfect substitutes for the elasticities, rank coefficients are known to be less affected by these estimation issues (Chetty et al., 2014; Nybom and Stuhler, 2017). Using career-long income histories of parents and children in Sweden, Nybom and Stuhler (2017) show that rank-based measures are the least attenuated by measurement error and the most stable over age (smallest life cycle bias), as they are scale invariant.

Using alternative income definitions as the dependent variable

In this section, I estimate income mobility again using alternative measures of income for the second generation. The choice of the income variable that is used as a proxy for long-run economic status is a common area of debate in the empirical body of work concerned with the measurement of intergenerational mobility (Belfield, Crawford, Greaves, Gregg and Macmillan, 2017; Björklund and Jäntti, 2011; Björklund and Salvanes, 2010; Landersø and Heckman, 2017). In practice, this choice is frequently determined by the availability of data and is limited by the data-intensive nature of intergenerational analyses. Ideally, the definition of income chosen should reflect the main purpose of the study, the concept of social mobility being examined and the research question being investigated (Björklund and Jäntti, 2011).

Various measures of income are available in the BHPS and Understanding Society data, which allows me to examine the implications of using alternative definitions of income.[43] Here, I estimate income mobility using three different measures of children's income in adulthood, namely labour earnings, total personal income and the sum of couple's combined personal incomes. The respective results are presented in Table 2.3. The first row (1) displays the results obtained earlier when considering household income for parents and children.

Row (2) presents the estimates based on labour earnings for the second generation. This variable represents the usual pre-tax monthly labour earnings from the main job, self-employed profit and second and/or occasional jobs. Labour earnings are often used to examine earnings mobility in the literature. This variable is present in most data sets and it captures people's earnings power in the market since labour earnings are the dominant source of income for most individuals. The estimates of IGE based on labour earnings are around 22% larger; β is estimated to be 0.329 using OLS and 0.315 using TSRA.

In addition, I also consider another measure of individual income, personal income. Personal income, sometimes also called total factor income, is a broader measure of income-generating power (income potential) than earnings alone (Björklund and Jäntti, 2011) and includes inherited capital income and other non-labour income sources (benefits, pensions, transfers), as well as labour earnings. Row (3) shows that the estimated IGEs drops by around 10% to 0.293 (OLS) and 0.270 (TSRA) when public transfers and inherited capital are

[43] A detailed description of the income variables in the data is provided in the Appendix A.2.

included in the measure of income. This decrease suggests a role of redistribution via public transfers may be increasing mobility in comparison to just earnings.

Row (4) displays the estimates for pooling personal income for partners in the same household, which reflects the income generating power of the couple.[44] Considering the couple's personal income, even though the IGEs are slightly higher than those obtained for household income, they are lower than those obtained using the other measures of individual income, pointing to the importance of partners.

Examining the rank coefficients, the estimates based on labour earnings and personal income are around 10% smaller than those obtained for household income. These estimates are strikingly similar for the two individual measures of income (personal income and labour earnings), especially so when estimated by TSRA. This is to be expected, as no major re-ranking of individuals across the distribution would be likely observed when moving between these two very similar measures of income. Considering the couple's combined income, however, the rank coefficients drop even further. This is likely related to the increased variation in the ranks added by the inclusion of partner's income. This evidence suggests that it could be important to consider how the intergenerational association of family resources may be affected by dynamics of partnership formation and assortative mating. This is an interesting topic in itself, which deserves being investigated further. The following chapter of this thesis addresses some of these issues.

The comparison between estimates of mobility based on household income and those based on individual measures of income (personal income, labour earnings) is not straightforward. These measures typically reflect different concepts of social mobility and will have different interpretations and policy implications (Björklund and Jäntti, 2011). When examining mobility at the individual level, the focus is on the influence of family background and parental resources for the children's earnings or income generating potential in adulthood. This concept is more closely linked to the ideal of providing equal opportunities for all members

[44] Here, if the individual is single, partner's income is set to zero.

Table 2.3 IGE and rank coefficient based on different measures of income

	IGE (β)			Rank coefficient (γ)		
Child income measure	N	OLS	TSRA	N	OLS	TSRA
Household income (1)	2126	0.269***	0.255***	2126	0.301***	0.291***
		(0.028)	(0.027)		(0.024)	(0.023)
			[0.021]			[0.018]
Labour earnings (2)	1924	0.329***	0.315***	2114	0.274***	0.262***
		(0.040)	(0.030)		(0.023)	(0.022)
			[0.030]			[0.017]
Personal income (3)	2099	0.293***	0.270***	2114	0.262***	0.261***
		(0.031)	(0.028)		(0.022)	(0.022)
			[0.022]			[0.017]
Couple's personal income (4)	2103	0.289***	0.268***	2114	0.228***	0.241***
		(0.034)	(0.031)		(0.022)	(0.023)
			[0.024]			[0.018]

Notes: Robust clustered standard errors in parentheses. Bootstrapped standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The measure of child income varies between the rows of this table, while parental income is household income in all models, constructed as described in the text in section 2.3.2. OLS models include controls for the average age of parents and children, their average age squared and birth cohort dummies for both generations. TSRA models include controls for a quartic in age and year dummies in the first stage regressions for each generation. Sample sizes are smaller (and varying) for the estimation of IGE than for rank coefficients because of the incidence of zeros. Sample sizes for ranks in rows (2), (3) and (4) are smaller than in (1) because this information is missing for 12 individuals who did proxy interviews. Source: BHPS and Understanding Society.

of society. The differences between earnings and personal income highlight the role of redistributive policies, transfers and benefits. On the other hand, when moving to household income for the second generation, the focus shifts to the observation of the association between living standards, or household welfare across generations. It is important to keep in mind that the estimates based on household income incorporate the role of intergenerational wealth transfers and are affected by the processes of household formation and assortative mating.

For the main analysis in this chapter, I have opted to use household income for both generations. As discussed before, this is motivated by the focus on the association of household well-being across generations. The use of family or household income measures reflects the economic resources of the household in a broader sense. As a measure of parental income, it captures the resources of the parents during childhood. When used as a measure of child income in adulthood, it represents the current resources of their households in adulthood, highlighting the contribution of all members to the well-being of the whole household. An important advantage of using this measure is that it allows for the inclusion of daughters in the analysis without having to work around the complexities of patterns of female labour force participation. While my analysis cannot pinpoint specific factors, the (small) differences observed between the estimates based on household income and individual earnings or income could occur for several reasons.

Focusing on the rank coefficients, which are not affected by the extent of inequality across generations and are less affected by the common estimation issues, the estimate based on household income is slightly higher than those obtained for the measures of individual income. This suggests that parental income during childhood appears to be slightly more strongly associated with household income than with individual outcomes. In turn, this would point to family resources being important for the transmission of advantages through endowments, such as returns to investments.

Alternatively, this difference could be a consequence of sample selection, since labour earnings are only observed for individuals who are in employment at the time of interview and hence report their earnings. Gregg et al. (2017) and Belfield, Crawford, Greaves, Gregg and Macmillan (2017) note that the exclusion of out-of-work sons usually leads to an understatement of the level of intergenerational persistence due to sample selection. This is typically the main criticism related to the use of labour earnings to examine intergenerational mobility (Gregg et al., 2017).[45] Household income, on the other hand, is also observed for out-of work individuals. In addition, for the individual measures of income, it is unclear what is the effect of the presence of daughters in my sample. Using labour earnings to measure mobility is especially complicated and controversial when women are included in the analytical sample, as in this chapter, because of issues related to the patterns of women's participation in the labour force. (Chadwick and Solon, 2002; Raaum et al., 2008). The differences in mobility by gender will be further examined in the next chapter of this thesis.

[45] An additional issue relates to the reporting of self-employment earnings, which are often missing or measured less accurately in survey data (Björklund and Jäntti, 2011).

Changing the measure of income used to estimate income mobility corresponds to a change in the concept of social mobility analysed. While these differences can be informative of underlying mechanisms of intergenerational transmissions, it is important to keep this in mind for the interpretation of comparisons with any other results. Overall, the results indicate that there is a considerable degree of intergenerational persistence in the UK, also when different income variables are used for the second generation.

2.4.2 Robustness analysis

So far, the results from my main analysis suggest a considerable persistence of socio-economic status across generations. In this subsection, I perform a series of checks to further assess the robustness of my findings. Firstly, I investigate how changes in the age restrictions imposed to the sample affect my estimates. Secondly, I examine how children's co-residency with their parents in adulthood might affect the estimates of income mobility. Finally, I examine the sensitivity of the estimates to different treatments of outliers.

Relaxing age restrictions in the sample

As discussed previously in section 2.2.1, a common issue in the estimation of intergenerational mobility relates to the life cycle bias. This problem is also a consequence of using snapshots of income to proxy long-run income, and more particularly, of issues around the ages at which incomes are observed (Grawe, 2006; Haider and Solon, 2006; Jenkins, 1987; Nybom and Stuhler, 2016). The inability to capture income in both generations at representative and comparable periods may affect mobility estimates in many ways, making it difficult to establish the direction of the bias (Jenkins, 1987). The age-dependency of elasticity estimates has been widely discussed in the literature - estimates tend to increase with the age of sampled sons (Böhlmark and Lindquist, 2006; Solon, 1999) and with the dispersion in transitory income varying over the life cycle (Björklund, 1993). This literature suggests that the life cycle bias may be smallest when incomes are observed in mid-life, around the age of 40 (Haider and Solon, 2006; Nybom and Stuhler, 2016).

For the main analysis, I have considered the incomes of both generations at ages no younger than 25.[46] [47] In addition, all OLS models include age and cohort controls for both generations, and the TSRA models control for a quartic in age and year dummies in the first stage regressions to help mitigate the potential bias from life cycle effects. The average age at which parental income is observed is 42, while the income for their children is observed at a slightly younger age, at 30. I now examine the sensitivity of my results to changes in the age restrictions imposed to the second generation, that is, varying the ages at which children's incomes are observed in adulthood.

The estimates of income mobility for different age groups using the household income variable are illustrated in Figure 2.3. The complete results are presented in Table A.7 in Appendix A. These results reveal that the estimated IGEs and rank coefficients are slightly larger at younger ages and decrease gradually as the minimum age in the sample increases from 20 to 25, becoming more stable afterwards.[48] A small increase is observed again at ages 30+, however, it is unclear whether this is a sign of higher persistence at that age, or a result of less precise estimates due to a smaller sample size.

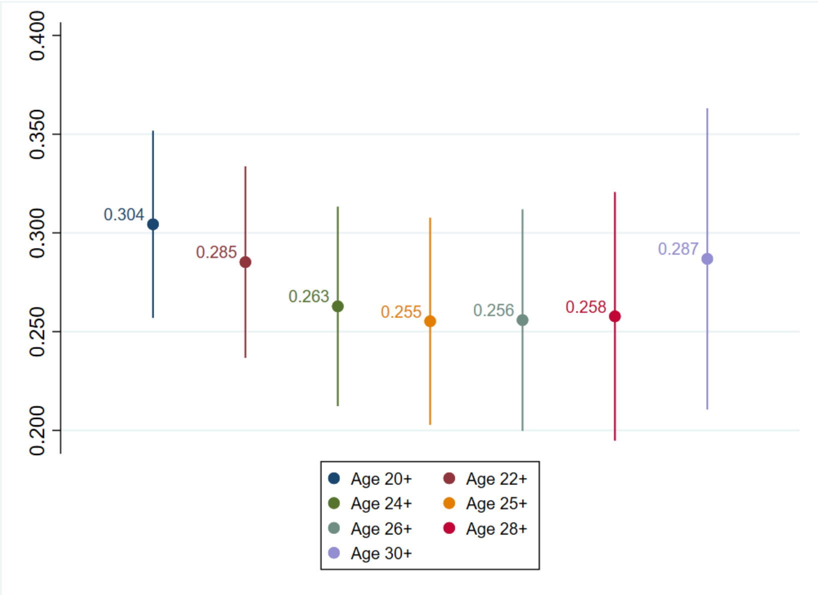
When examining the results by age groups based on household income, it is not entirely possible to disentangle the changes in income persistence related to age and those related to co-residency effects. The larger estimates at younger ages could be the result of an age effect but it is also possible that these results may be affected by changes in household composition. When household income is chosen as the measure of socio-economic status in adulthood, one extra confounding factor manifests itself. Individuals at younger ages are more likely to live with their parents (i.e. be co-residents) and for these cases the observed household income for the second generation will also include their parent's income. Therefore, if a large share

[46]For the first generation, incomes are also not considered after the age of 65.

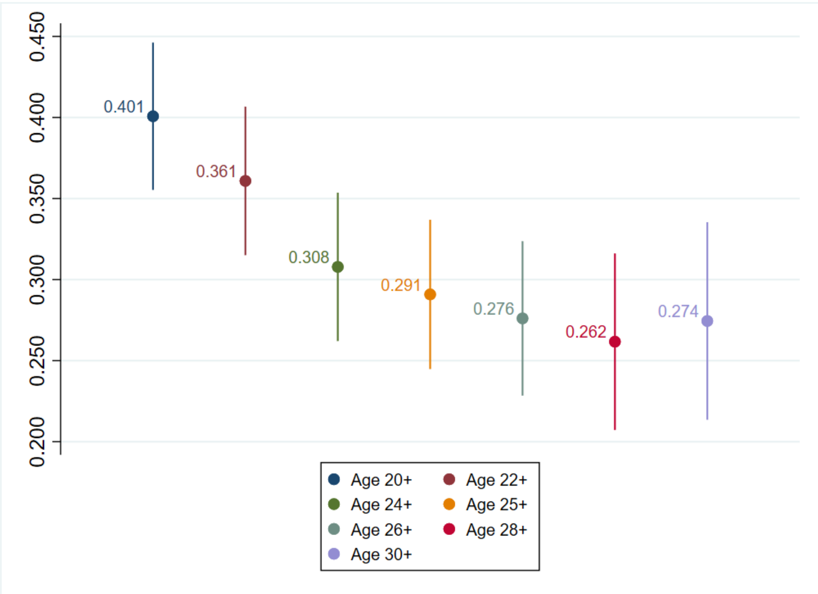
[47]Using 25 as the cut-off age is also done by other studies, see for example Lee and Solon (2009).

[48]However, these differences should be interpreted with caution because of the large, and sometimes overlapping confidence intervals.

of individuals in the sample still live with their parents at younger ages, it is likely that this would over-estimate the degree of household income persistence at these ages. To shed light on this issue, I now examine what happens to estimates of mobility based on personal income, which are not affected by the co-residency issue. Then, complementing this analysis, the results splitting the sample by co-residency status will be analysed in the following section.



(a) IGEs



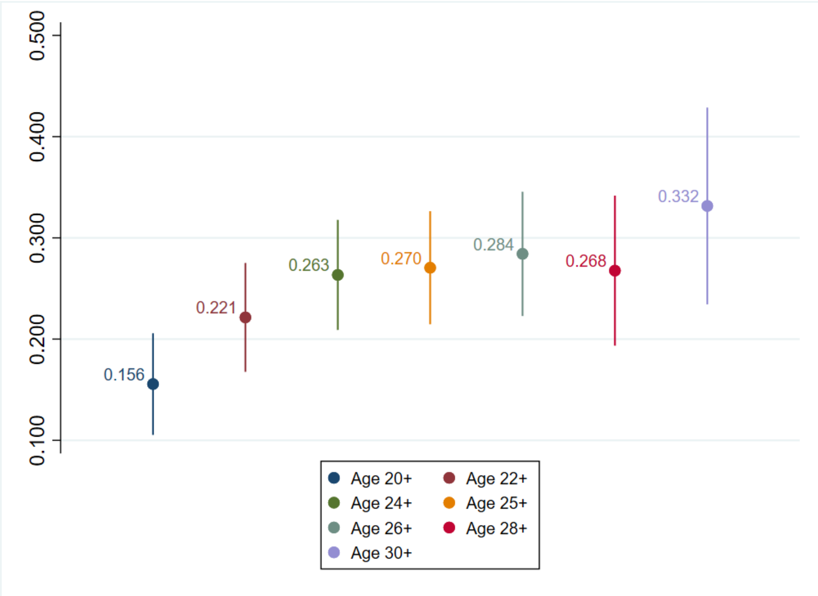
(b) Rank coefficients

Figure 2.3 IGE and rank coefficients by age groups (TSRA): household income

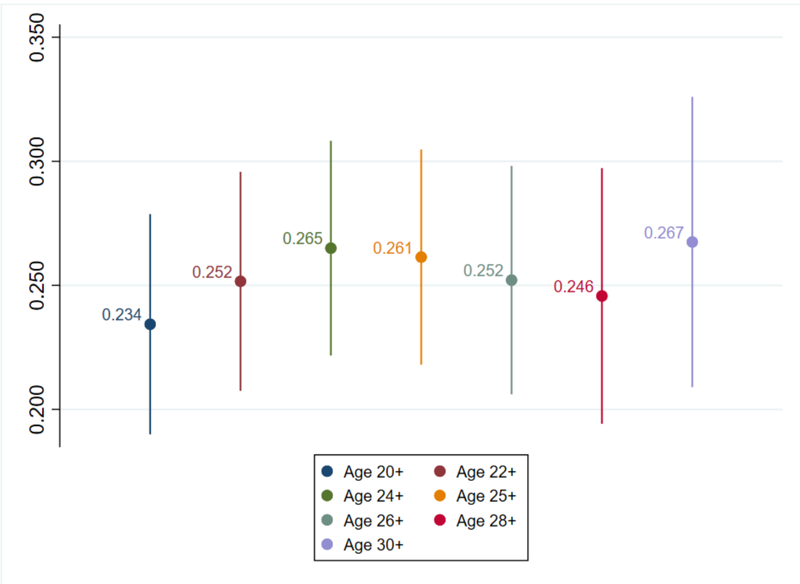
Figure 2.4 shows the estimates of income mobility across age groups considering personal income as the dependent variable. Income persistence is considerably lower at 20 years old, particularly as measured by IGEs (upper panel), and it increases slowly after this age becoming more stable after the age of 24. These results are consistent with the hypothesis that the life cycle bias attenuates the intergenerational elasticity of earnings (or personal income) when individuals in the second generation are observed at very young ages (Haider and Solon, 2006; Jenkins, 1987; Nybom and Stuhler, 2016) because these individuals have most likely not reached their earnings potential yet. A good example of estimates can be affected by this problem are the extremely low (close to zero) earnings elasticities obtained by Ermisch and Francesconi (2004) when considering children's incomes from the age of 16. Differently to the estimates obtained using household income, these results are not affected by co-residency issues. This reinforces the findings that the degree of mobility estimated for the main sample where income is observed from the age of 25 is not seriously affected by the life cycle bias.

In practice, the age at which income for the second generation is observed is often dictated by data availability. When the income of the second generation is observed at younger ages, these individuals may have not yet reached their full earnings potential, leading to an attenuation of the estimated intergenerational coefficients and to an overestimation of intergenerational mobility. For the UK, Gregg et al. (2017) examine how IGEs and rank coefficients change for the NCDS and BCS cohorts when sons' eanings are observed at different ages. They show that the estimates of persistence are much lower when the sons are in their early twenties and that they peak when sons are around mid-forties. These results are similar to what has been observed for other countries and usually attributed to systematic heterogeneity in the rates of income growth over the life cycle (Grawe, 2006; Haider and Solon, 2006; Nybom and Stuhler, 2016).

Overall, my results from examining how the estimates of income mobility change across age groups suggest that these estimates are relatively stable after the age of 25, particularly when estimated by TSRA. This is reassuring - it suggests that my estimates are not seriously affected by life cycle bias. Due to a possibility of heterogeneity of income profile across individuals, it is possible that income observed at older ages would still be a better representation of long-run income. Conversely, observing income at younger ages is not recommended as this would probably introduce large biases to the mobility estimates.



(a) IGEs



(b) Rank coefficients

Figure 2.4 IGE and rank coefficients by age groups (TSRA): personal income

Examining income mobility by co-residency status

When using household income to capture the relationship between economic status across generations, one potential confounding factor presents itself: the proportion of individuals who still live with their parents. When children live with their parents when income is observed in adulthood, this implies that their household income will also include the parents' income. This would under-estimate mobility if the share of people who co-reside with their parents is higher than the average for the population because they are being observed at younger ages.

In order to investigate whether and how my estimates of income mobility are affected by the co-residency issue, i.e. individuals still living with their parents at young ages, I split my main sample in two groups according to co-residency status in adulthood. By looking at household identifiers, I identify individuals who live with at least one of their biological parents in adulthood (co-residents) and those who live in a different household (non co-residents). To begin with, Table 2.4 presents the proportion of individuals in each group by co-residency status at the ages of 20, 25 and 30. As it would be expected, the fraction of individuals who live with their parents decreases with age. In late adolescence, at ages 16-18, over 99% of the children in the sample still live with their parents. A few years later, at the ages of 20-22, around 23.4% of individuals in the sample live in a separate household but this number almost triples at ages 24-26 (66.9%). By the age of 30, the vast majority of individuals (86.2%) live in a different household to their parents.

The change of status from co-resident to non co-resident as children get older would explain why child age displays (unexpectedly) a negative relationship with child household income (Table A.3 in Appendix A). It is likely that upon leaving the household, the 'children' initially experience a decrease in household income which, previously, at younger ages, likely also included the incomes of their parents. An analysis of sample means of the key variables by co-residency status (Table A.8 in Appendix A) supports this argument, as it reveals that co-residents at ages 24-26 have, on average, higher household income and lower personal income in adulthood than non co-residents.

Table 2.4 Cohabitation status by age groups

Status	Ages 16-18	Ages 20-22	Ages 24-26	Ages 30-32
Co-residents (%)	99.7	76.6	33.1	13.8
Non co-residents (%)	0.3	23.4	66.9	86.2
Total(%)	100	100	100	100
N	2126	2063	2066	1112

Notes: This table shows the proportion of individuals in each group when splitting the main sample by cohabitation status for different age groups. Sample sizes vary across columns because some individuals were not interviewed at these exact ages. Source: BHPS and Understanding Society.

In order to check whether the IGEs and rank coefficients vary according to children's co-residency status when their income is observed, I re-estimate the coefficients for co-residents and non co-residents separately. The results obtained using the household income measure for both generations are shown in Table 2.5. Panel A shows the results splitting the sample based on the cohabitation status at ages 24-26. As expected, I find that the IGEs for non co-residents are much lower than those estimated for individuals who live with their parents, and that the difference between the coefficients of the two groups is highly statistically significant. The rank coefficients follow the same patterns. Panel B displays the estimates when restricting the sample to individuals at ages 30+ and splitting the sample based on the cohabitation status at ages 30-32. Here too I observe that the estimates for non co-residents are much lower than those obtained for co-residents. The most important finding from this table is that the estimates for non co-residents are strikingly similar to those obtained before for the whole sample. This is reassuring that the main estimates based on household income are not only being driven by a co-residency effect.

So why not focus on the sample of non co-residents to avoid this issue? The common practice of restricting the sample to individuals who have moved out of their parents' home to start a new household, or to consider only the income of head of households also comes with limitations. These would probably be selected samples because co-residency is related to not only age but employment status, marriage and other characteristics that influence the income generating power of the household. In the population, the two groups of people are present, and applying such a restriction would likely under-represent co-residents.

To summarise, when splitting the sample by co-residency status, I have observed that the estimates of mobility - IGEs and rank coefficients - based on household income for the group of non co-residents are strikingly similar to those obtained earlier for the whole sample. This is reassuring that the intergenerational associations observed at the household level are not predominantly driven by a sample selection on co-residents.

Table 2.5 IGE and rank coefficient by co-residency status

		IGE (β)		Rank coefficient (γ)	
	N	OLS	TSRA	OLS	TSRA
Panel A: Ages 24-26					
All	2066	0.270***	0.251***	0.324***	0.291***
		(0.029)	(0.032)	(0.023)	(0.024)
Co-residents	684	0.404***	0.366***	0.425***	0.401***
		(0.042)	(0.045)	(0.038)	(0.037)
Non co-residents	1382	0.243***	0.234***	0.281***	0.279***
		(0.035)	(0.032)	(0.028)	(0.027)
Co-resident = Non co-resident (p-value)		0.003***	0.025**	0.002***	0.007***
Panel B: Ages 30-32					
All †	1112	0.279***	0.242***	0.352***	0.293***
		(0.030)	(0.034)	(0.030)	(0.029)
Co-residents	153	0.474***	0.446***	0.428***	0.425***
		(0.077)	(0.073)	(0.082)	(0.078)
Non co-residents	959	0.261***	0.243***	0.320***	0.290***
		(0.032)	(0.034)	(0.032)	(0.031)
Co-resident = Non co-resident (p-value)		0.006***	0.014**	0.136	0.100

Notes: Robust clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Estimates of mobility splitting the sample in individuals who co-reside with parents in adulthood (co-residents) and those who do not (non co-residents). Child income and parental income are measures of household income, and constructed as described in the text in section 2.3.2. OLS models include controls for the average age of parents and children, their average age squared and birth cohort dummies for both generations. TSRA models include controls for a quartic in age and year dummies in the first stage regressions for each generation. † The co-residency status at ages 30-32 is only available for the subsample of individuals born between 1973-1987. Source: BHPS and Understanding Society.

Treating outliers in the income data

It has also been emphasised in the literature that mobility estimates, especially intergenerational elasticities, might be sensitive to the treatment of extreme values and zeros, and that censoring may also produce biased results (Dahl and DeLeire, 2008; Landersø and Heckman, 2017; Nybom and Stuhler, 2016). Here, I test the robustness of my results to the treatment of outliers at the bottom and top of the income distribution using different methods of treatment of outliers. The results treating the outliers of household income for parents and children are presented in Table 2.6.[49]

Firstly, I truncate the sample to only include the income observations between the 5th and 95th percentiles of children's and parental income distributions. This has very little effect on the results using household income apart from a slight decrease in the rank coefficient, possibly an indicative of a high rank persistence at the very top and at the very bottom of the distribution. Then, I Winsorise the top and bottom incomes between 1-99% and 5-95%. Winsorising is done by limiting and re-coding extreme values in the data. In practice, I substitute the lowest income values for the values below 1% and 5% and the highest income values for the values above 99% and 95%, respectively. Panels B and C in Table 2.6 reveal that Winsorising (at 1% and 5%) very low incomes has an effect of increasing the estimated elasticities and rank coefficients slightly, possibly due to inclusion of zero and very low incomes as higher values. A similar effect is observed when top-coding incomes at 99% and 95%.[50] Conversely, rank coefficients remain remarkably stable to the treatment of outliers, possibly because this measure is scaled and unaffected by the variation of the distribution. Even though all the observed changes are only small, it is noticeable that the estimates from TSRA models have been slightly less sensitive to the treatment of outliers than those obtained by OLS.

[49]The estimates obtained after treating outliers of personal income (not reported here) follow the same pattern as those observed for household income.

[50]The very small changes observed after top-coding very high incomes could be partly due to the fact that all income variables are already top-coded in this data.

Table 2.6 Summary of treatment of outliers

	IGE (β)			Rank coefficient (γ)		
Treatment	N	OLS	TSRA	N	OLS	TSRA
Original estimates (untreated)	2126	0.269***	0.255***	2126	0.301***	0.291***
		(0.028)	(0.027)		(0.024)	(0.023)
			[0.021]			[0.018]
Panel A: Truncating sample between 5-95%						
	2027	0.267***	0.254***	2027	0.269***	0.265***
		(0.024)	(0.024)		(0.024)	(0.023)
			[0.018]			[0.018]
Panel B: Winsorising 1-99%						
Bottom-coding	2126	0.276***	0.263***	2126	0.302***	0.292***
		(0.027)	(0.027)		(0.024)	(0.023)
			[0.021]			[0.018]
Top-coding	2126	0.283***	0.263***	2126	0.300***	0.291***
		(0.026)	(0.026)		(0.024)	(0.023)
			[0.020]			[0.023]
Panel C: Winsorising 5-95%						
Bottom-coding	2126	0.284***	0.276***	2126	0.304***	0.293***
		(0.026)	(0.026)		(0.024)	(0.023)
			[0.020]			[0.018]
Top-coding	2126	0.285***	0.264***	2126	0.298***	0.293***
		(0.026)	(0.026)		(0.024)	(0.023)
			[0.020]			[0.018]

Notes: Robust clustered standard errors in parentheses. Bootstrapped standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Estimates of income mobility in this table reflect the effect of treatment of outliers on parent's and children's income variables. Untreated estimates from the main analysis are shown in the first row for comparison. Child income and parental income are measures of household income, and constructed as described in the text in section 2.3.2. The sample size in Panel A is smaller because of the exclusion of individuals with very low and very high incomes (below 5% and above 95% of the distribution). The sample in Panels B and C corresponds to my main sample. Panels B and C present separate results from bottom-coding and top-coding incomes. Percentile ranks are based on the final matched sample after outliers have been treated. OLS models include controls for the average age of parents and children, their average age squared and birth cohort dummies for both generations. TSRA models include controls for a quartic in age and year dummies in the first stage regressions for each generation. Source: BHPS and Understanding Society.

2.5 Discussion and Conclusion

This chapter provides updated estimates of intergenerational income mobility in the UK. Specifically, using the harmonised longitudinal dataset comprised of the BHPS and Understanding Society (UKHLS), I estimate the intergenerational income elasticities and rank coefficients for a younger group of individuals than those analysed in the previous literature, who were born between 1973 and 1992. In addition to the updated estimates of income mobility for the UK, I propose a novel methodology to estimate intergenerational income elasticities and rank coefficients. The two-stage residual approach (TSRA) is a two-step procedure that makes use of all available data and controls more flexibly for age and time effects to provide an improved measure of income, comparable for both generations. The estimates obtained using TSRA also appear to be less sensitive to changes in variable definitions, sample restrictions and model specifications.

The BHPS+UKHLS dataset contains a large number of pairs of individuals matched to their parents, which makes it possible to directly obtain a proxy of parental income during an individual's childhood years and then follow these children as they become adults to obtain information on their income in adulthood. Combining 27 years of data, I observe income for both generations in multiple years of the surveys and at various age points, and use all this available data to calculate a more accurate proxy of long-run income. Measuring the income of parents and children across multiple periods is particularly important to minimise potential biases from transitory shocks to income. In addition, controlling for the ages and time when incomes are observed is essential to reduce the potential life cycle bias. This estimation represents an important advancement from previous UK studies that were based on the NCDS and BCS cohorts and relied on more limited data. To the best of my knowledge, this is the first study to use the harmonised BHPS+UKHLS to estimate intergenerational income mobility in the UK.

The main analysis of this chapter consists of estimating the intergenerational income elasticities (IGEs) using the standard OLS approach from the literature and the two-stage residual approach (TSRA). Complementing this analysis, I also provide estimates of rank coefficients using both estimation strategies. Overall, my main findings indicate that the intergenerational income elasticity in the UK is around 0.255-0.269 and statistically significant from zero. This corresponds to the intergenerational association, or persistence, of economic status and is measured by the average effect that a relative change in parental income will have on their children's income as adults, while keeping other factors constant. These results suggest that for every additional (reduced) 10% of parental income advantage (or disadvantage) around a quarter (2.6%) will be passed on to the next generation. The rank coefficients suggest similar levels of mobility. Using household income measures, I obtain a rank coefficient of around 0.291-0.301. This means that an increase of 10 percentile points in the rank of the parents would mean an increase of around 3.0 percentile points in the rank of children. The slightly higher rank coefficients reflect differences in the standard deviation of the household income distribution across generations.

Comparing to similar studies for other countries, these results reinforce the findings from previous studies, depicting the UK as a country with relatively low levels of income mobility also when income is measured at the household level. Although there is no such thing as the 'optimal level' of intergenerational mobility, the UK estimates reported here are slightly smaller than the most comparable obtained for the US by Chetty et al. (2014) considering family income. There are considerable institutional, economic and cultural differences across countries, but previous research suggests that cross-country differences in the levels of inequality and returns to education strongly influence measures of income mobility (Blanden, 2009).

Furthermore, the results presented in this chapter lie within the (wide) range provided by previous studies on the UK. To my knowledge, no other studies have examined mobility for the UK considering household income for both generations. The closest to this measure is that of net family income used by Belfield, Crawford, Greaves, Gregg and Macmillan (2017)

- they estimate an IGE of 0.17 for sons from the NCDS cohort and 0.31 for sons from the BCS cohort. It is important to acknowledge that any direct comparison with other studies must be done with caution due to differences in the data (Belfield, Crawford, Greaves, Gregg and Macmillan, 2017; Blanden, 2005*a*; Blanden et al., 2004; Gregg et al., 2017), or in the very distinct methodologies used, as it is the case of studies that employ the initial waves of the BHPS (Ermisch and Francesconi, 2004; Nicoletti and Ermisch, 2008). However, my results for the intergenerational earnings elasticity are approximately coinciding with the previous results by Blanden et al. (2004) for the BCS cohort using children's labour earnings and parents combined income. Their results were around 0.30 for men and 0.40 for women, and were also similar to those obtained by Belfield, Crawford, Greaves, Gregg and Macmillan (2017) for the BCS. Although the estimates reported by Gregg et al. (2017) for sons of the BCS cohort are slightly higher (0.43) after considering multi-year earnings observations for children and adjusting for workless spells. Contrasting with another study that also used a matched sample from the BHPS, Ermisch and Francesconi (2004) estimate an IGE of around 0.05-0.10 for individuals born between 1970-1983. However, their very low estimates are heavily attenuated by life cycle effects, as sons' income is observed at extremely young ages (from 16). Finally, my rank coefficients are strikingly similar to those reported in the two other studies that estimate this measure of mobility for the UK using data on sons from BCS cohort: Gregg et al. (2017) estimate the rank coefficient to be around 0.30, and Belfield, Crawford, Greaves, Gregg and Macmillan (2017) estimate it to be 0.31. This similarity is indicative that there has been little improvement in mobility in the more recent decades.

To complement my analysis, I estimate mobility using alternative definitions of income and examine the sensitivity of my results to changes in the sample restrictions, model specifications and treatment of outliers. While the main results use household income for parents and children with the objective of capturing the general living standards and economic resources at the household level, I find that the estimates of intergenerational mobility using individual measures of income (i.e. labour earnings and personal income) are only slightly different; the IGEs obtained are slightly higher when using these alternative measures of income, while rank coefficients are slightly lower. Especially considering the estimates obtained by TSRA, they are notably robust to the use of alternative income variables. Moreover, examining the estimates for different age groups, I show that the choice of age at which income is observed in the data is of crucial importance, as highlighted by previous work in this literature (Grawe, 2006; Haider and Solon, 2006; Nybom and Stuhler, 2016). Mobility estimates obtained using income data from age 20 are likely biased for both household and personal income measures, for different reasons. On the other hand, estimates obtained using income observed after the age of 25 are more stable and similar to those obtained when restricting the sample to individuals at older ages. Another test splitting the sample by co-residency status reveals that my estimates of elasticities and rank coefficients obtained using household income are not primarily driven by co-residency with parents. Finally, I show that the main estimates of mobility are also robust to various treatments of outliers.

Although my results should be compared to the previous findings with caution due to different data sets, variables and model specifications, the fact that my estimates are robust to a number of sensitivity checks is reassuring. This indicates that some of these choices might only have limited influence for the calculation of overall measures of income mobility, as long as the common estimation issues of transitory shocks in income and life cycle bias are being addressed properly. In this context, one possible explanation for some of the differences observed in relation to estimates for the NCDS and BCS cohorts reported in other studies is that they could be driven by different cohorts or that they may relate to changes in the institutional and policy environments that occurred over time. Notably, there has been a big increase in overall income inequality between the NCDS cohort, born in 1958, and the cohorts analysed here, born between 1973-1992, which likely explains part of the reduction in mobility over this period. In addition, the rapid (but concentrated) expansion of education supply that explains differences between the NCDS and BCS cohorts (Blanden et al., 2004) is also likely to explain some of the differences between my estimates and those reported for the NCDS cohort. On the other hand, the similarity with the degree of mobility estimated for the BCS cohort, born in 1970, is indicative that there has been little improvement in the levels of mobility in the past decades. This is in spite of important changes in education policy that have occurred in this period, for example the 1988 Education Reform Act or the increase in the costs of higher education in the early 2000s, or in changes in the economic environment, such as the Great Recession. Investigating these hypotheses as well as potential differences in the levels of mobility across the birth cohorts studied here remains a challenge for future research.

This chapter represents an important advancement in the literature of intergenerational income mobility in the UK, providing updated estimates for the 21st century. Yet, many questions remain unanswered. The updated measures of intergenerational income elasticities and rank coefficients can act as a platform for further investigation of how intergenerational income mobility varies between different groups in society, and across a number of dimensions found to be important in the international literature. I will address some of these questions in the next chapter of this thesis.

Appendix A

A.1 Sample representativeness

As described in section 2.3.3, various conditions (C) are imposed to the sample to obtain the main matched sample used throughout this paper. Firstly, individuals must be born between 1973-1992 in order to be eligible in terms of age. Secondly, they need to be matched to biological parents within the BHPS. Thirdly, individuals have to be interviewed on/after the age of 25 and have at least one observation of income. Finally, there needs to be at least one observation of their parental income during childhood years. Unfortunately, the imposition of these necessary conditions means that I drop some cohort members for which this data is invalid or missing. This is a result of both item non-response and attrition patterns within the BHPS sample. If these attrition and non-response patterns are non-random, it is possible that the estimates of intergenerational mobility obtained for this sample could be different than those that would be obtained for a fully representative sample. Solon (1992) discusses the implications of using unrepresentative homogeneous samples. If the explanatory variable has a lower variance in the observed sample than in the population, for example, this could aggravate the problem of measurement error and attenuate the estimates of persistence. Therefore, after imposing conditions C1-C4 it is important to examine the representativeness of my final matched sample.

It is clear that attrition and non-response within the BHPS have meant a considerable loss in the sample. It is also important to examine whether and how these issues affect sample composition. I first examine the differences in sample means of my key variables between individuals (children) in final sample and those out of the sample (but from the same birth cohorts and matched to their biological parents). Here, I am mostly interested in understanding patterns of attrition of young adults, related to whether the characteristics of the individuals (children) who have been interviewed at/after the age of 25 are different from those who dropped out of the survey before this age. Only 2636 individuals were interviewed at ages 25+ (i.e this is the sample with restrictions C1,C2 and C3). Among these, 2126 individuals have also simultaneous information on parental income.[51] Table A.1 presents the sample means of parent's income and age comparing individuals who are in (1) and out (2) of the sample. In terms of parental characteristics, this comparison indicates the children who remain in the sample had on average slightly richer parents (+ £73.8 per month) during childhood years. Their parents were also slightly older (1.4 years) when income during childhood was observed. This is not a surprising result - if the BHPS has a slight bias to those with higher socio-economic status, it is likely that restricting the sample to follow individuals from childhood into adulthood will reinforce this pattern.

Table A.1 Differences in means: in and out of sample

	In sample (1) Mean	Out of sample (2) Mean	Diff (1)-(2)	P-value Diff
Panel A: Parents				
Parental household income	3616.6	3542.8	73.8	0.008
	(2717.5)	(2482.8)		
Parental age when income observed	42.0	40.6	1.4	0.000
	(7.0)	(6.9)		
Child age when parental income observed	13.0	12.2	.8	0.000
	(4.1)	(4.3)		
Panel B: Children				
Female (%)	54.3	46.8	7.5	0.000
Obs.	15,631	19,837		
N	2126	2542		

Notes: Standard deviation in parenthesis. The combination of samples (1) and (2) adds up to the number of individuals born between 1973-1992 whose parents have at least one income observation available during childhood (N=4668). Source: BHPS and Understanding Society.

I then examine parent's characteristics with respect to employment and education. Here, I am interested in understanding the patterns of missing values of child income in adulthood based on parental characteristics. Therefore, I focus on parent's education (which is likely very stable over time) and retrospective information on parents' employment during childhood. Table A.2 presents some background data on employment work and education for mothers and fathers of individuals in and out of the sample. It shows that individuals in the final sample

[51] The majority of cases without information on parental income have been included as part of BHPS sample boosts added in later years (1999-2000).

have slightly more educated parents; a larger share of fathers and mothers has a degree, and a smaller proportion has no qualifications. Moreover, for the individuals in the sample, the fraction of fathers and mothers who were working when their child was 14 is higher, and the share of parents who did not work at that time is smaller.

It is unfortunate that attrition and non-response lead to a considerable decrease in sample sizes used in this analysis, although this is a common problem of most if not all studies on intergenerational mobility using survey data. Overall, this analysis suggests that the pattern of attrition in the sample slightly affects more individuals from disadvantaged socio-economic backgrounds (i.e with lower parental income in childhood and less qualified parents), who have not been included in the final matched sample. Most of the difficulty in keeping the initial sample was due to the loss of observations of children in early adult years. All in all, differences in means between in and out samples were significant, however the magnitude of these differences were extremely small. Statistical significance here, could be a reflection of the large number of observations (e.g. 15631 observations for the parents in the sample) and I would not expect these small differences in means to substantially alter the interpretation of the main findings.

A.2 Income variables

Household income “sums the values of total income in the month before interview for individuals in the household" (Taylor et al., 2010, App2-5). This variable is derived in the dataset and it includes the sum of non-labour income and labour income for all individuals in the household. It is comprised of the following components:

- Household gross labour earnings: This measures the usual monthly wage or salary payment before tax and other deductions in current main job for employees, or the monthly self-employed profit variable for self-employed respondents or employees’ last

Table A.2 Parental characteristics: employment and education

	(1) In sample N=2126	(2) Out of sample N=2542	(3) Out of sample N=3379
Panel A: Fathers	[1560]	[1642]	[2306]
Highest Education	[1483]	[1519]	[2023]
Degree (%)	14.2	14.1	12.9
No qualifications (%)	18.2	24.0	26.9
Retrospective information	[1560]	[1561]	[2306]
Father working when child aged 14 (%)	75.7	67.4	69.5
Father not working when child aged 14 (%)	10.5	12.4	10.1
Missing info at age 14 (%)	13.8	20.2	20.4
Panel B: Mothers	[2061]	[2377]	[3255]
Highest Education	[2038]	[2344]	[3118]
Degree (%)	12.1	10.0	9.1
No qualifications (%)	19.4	24.0	25.1
Retrospective information	[2061]	[2268]	[3255]
Mother working when child aged 14 (%)	64.9	58.9	56.4
Mother not working when child aged 14 (%)	25.7	27.3	27.5
Missing info at age 14 (%)	9.4	13.8	16.1

Notes: Sample sizes in brackets. The combination of the samples in (1) and (2) adds up to the number of individuals whose parents have at least one income observation available during childhood (N=4658). The combination of individuals in (1) and (3) adds up to all individuals born between 1973-1992 matched to their biological parents (N=5505). Source: BHPS and Understanding Society.

wage or salary payment before tax and deductions in current main job. Income from second and occasional jobs is also added if non-missing.

- And non-labour income:
  - Household investment income : This variable totals the estimated income from savings and investments, and receipts from rented property, received in the month before interview.
  - Household benefit income: This variable totals all receipts from state benefits, received in the month before the interview.
  - Household pension income: This variable totals all receipts from non-state pension sources, received in the month before the interview.

- Household transfer income: This variable totals all receipts from other transfers, (including education grants, sickness insurance, maintenance, foster allowance and payments from TU/Friendly societies, from absent family members), received in the month before the interview.

Labour income is the pre-tax labour income in the month before the interview. It is a derived variable comprised of all labour income sources (main job, second job, self-employed).

Personal income is the sum of all the above mentioned non-labour and labour income sources at the individual level.

All these income variables include imputed data (e.g. for cases where interviews are done by proxy). The BHPS guide recommends always using imputed data in order to reduce the potential bias that would be caused by the elimination of observations with missing data (Taylor et al., 2010, A5-22). Throughout this chapter, I use the provided income variables including the imputed data.

A.3 TSRA: Bootstrapped standard errors

When using a generated variable to estimate an econometric model, that is, a variable that has been constructed (e.g. predictor or residual) from an estimated equation, it is important to consider the implications of this procedure for inference (Pagan, 1984). Particularly, it is important to adjust the standard errors to take into account the sampling variance of generated outcomes from the first stage. The usual non-adjusted standard errors would likely underestimate the actual sampling variance in the second stage regressions.

To do address this issue, I re-calculate the standard errors using a bootstrap technique in both stages of TSRA. In practice, I have written a programme in Stata to bootstrap the first and second stages simultaneously. First, I draw bootstrap first-stage samples of observations of income of parents and children (in adulthood) from which I obtain - for each generation -the residuals used to generate adjusted incomes. Second, a bootstrap sample of adjusted incomes is drawn and used for running the second-stage regression to generate estimates of mobility. Note that in this random drawing of the samples, in both stages, it is possible that some of the original observations will appear once, some multiple times and others not at all.[52] After repeating this process 1999 times, the bootstrap standard error ($\hat{se}$) is calculated by the standard deviation of the distribution of the bootstrap estimates:

$$\hat{se} = \{\frac{1}{k-1}\sum_{i=1}^{k}(\hat{\theta}_i - \bar{\theta})^2\}^{1/2} \tag{A.1}$$

and

$$\bar{\theta} = \frac{1}{k}\sum_{i=1}^{k}\hat{\theta}_i \tag{A.2}$$

where $i = 1,2,...,k$ denote the bootstrap samples and $\hat{\theta}_i$ is the value of the statistic (the intergenerational elasticity, for example) from the ith bootstrap sample and $\bar{\theta}$ is their mean.

All bootstrapped standard errors in this chapter are based on 1999 replications, following the '99-rule' recommended by Davidson and MacKinnon (2000) and the number of seeds has been chosen using a random number generator.

A.4 Supplementary results

[52] Since observations may not be independent within the same cluster, the bootstrap samples are drawn in the unit of cluster - the parental identifier.

Table A.3 OLS estimates of intergenerational income elasticities for the UK

	IGE (β)	
Log child income	(1)	(2)
Log parental income	0.292***	0.269***
	(0.027)	(0.028)
Child age	-	-0.249***
		(0.090)
Child age^2	-	0.0044***
		(0.0015)
Parental age	-	0.022
		(0.028)
Parental Age^2	-	-0.000184
		(0.00031)
Cohort dummies	No	Yes
Constant	5.84***	8.94***
	(0.22)	(1.49)
N	2126	2126

Notes: Robust clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Child income and parental income are measures of household income. Model (1) is estimated without any additional controls. Model (2), controls for the mean age of parents and children when their income is observed, their mean age squared, as well as cohort dummies. Controlling for a quartic function of mean age of children and of parents yields the exact same results. Source: BHPS and Understanding Society.

Table A.4 First stage regressions (TSRA)

	Log parental income	Log child income
Parental age	-0.88***	
	(0.32)	
Parental age^2	0.033***	
	(0.011)	
Parental age^3	-0.00050***	
	(0.00017)	
Parental age^4	0.0000026***	
	(0.00000098)	
Child age		-1.10
		(1.54)
Child age ^{2}		0.044
		(0.071)
Child age ^{3}		-0.00075
		(0.0015)
Child age ^{4}		0.0000047
		(0.000011)
Year dummies	Yes	Yes
N	15631	13479

Notes: Robust clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the first stage auxiliary regressions of the TSRA approach. Parental and child income are household income and the first stage regressions control for a quadratic function of age and year dummies. The first stage regressions also include a constant. Source: BHPS and Understanding Society.

Table A.5 Second stage regressions (TSRA)

	IGE (β)	
Adjusted child income	(1)	(2)
Adjusted parental income	0.259^{***}	0.255^{***}
	(0.026)	(0.027)
	[0.020]	[0.020]
First stage controls		
Child age	No	Yes
Parental age	No	Yes
Child age quartic	No	Yes
Parental age quartic	No	Yes
Year dummies	Yes	Yes
Constant	0.012	0.022^{**}
	(0.013)	(0.013)
N	2126	2126

Notes: Robust clustered standard errors in parentheses. Bootstrapped standard errors in brackets * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the second stage regressions of the TSRA approach. Parental and child income are household income. Model (1) controls for cohort dummies for parents and children in the first stage regressions. Model (2) controls for a quartic function of age for parents and children and time dummies in the first stage regressions. Controlling for cohort dummies instead of year dummies in the auxiliary regressions yields very similar results. Source: BHPS and Understanding Society.

Table A.6 Estimates of the percentile rank coefficient for the UK

	Rank coefficient (γ)		
	OLS		TSRA
Child Rank	(1)	(2)	(3)
Parental Rank	0.323^{***}	0.301^{***}	$.291^{***}$
	(0.023)	(0.024)	(.023)
Average Child Age	-	-0.15***	
		(0.050)	
Average Child Age ^{2}	-	0.0027***	
		(0.00085)	
Average Parental Age	-	0.018	
		(0.015)	
Average Parental Age^2	-	-0.00018	
		(0.00017)	
Child birth year	No	Yes	
Parent birth cohort	No	Yes	
Constant	0.34^{***}	2.09^{**}	0.35^{***}
	(0.013)	(0.83)	(0.013)
N	2126	2126	2126

Notes: Robust clustered standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table shows the percentile rank coefficient of income estimated by OLS and TSRA. Child income and parental income are measures of household income. Model(1) is estimated without any additional controls. Model(2) controls for the average age of parents and children and their average age squared, as well as birth cohort dummies. Model(3) controls for a quartic in age for parents and children and year dummies in the first stage regressions. In all models, ranks are based on the final matched sample. Source: BHPS and Understanding Society.

Table A.7 IGE and rank coefficient estimated for different age groups

		IGE (β)		Rank coefficient (γ)	
Age Range	N	OLS	TSRA	OLS	TSRA
20-44	2126	0.351***	0.304***	0.422***	0.401***
		(0.024)	(0.024)	(0.024)	(0.023)
			[0.019]		[0.018]
22-44	2126	0.320***	0.285***	0.378***	0.361***
		(0.024)	(0.025)	(0.024)	(0.023)
			[0.020]		[0.018]
24-44	2126	0.284***	0.263***	0.319***	0.308***
		(0.026)	(0.026)	(0.024)	(0.023)
			[0.020]		[0.018]
25-44	2126	0.269***	0.255***	0.301***	0.291***
		(0.028)	(0.027)	(0.024)	(0.023)
			[0.021]		[0.018]
26-44	1901	0.263***	0.256***	0.284***	0.276***
		(0.029)	(0.029)	(0.025)	(0.024)
			[0.022]		[0.019]
28-44	1476	0.253***	0.259***	0.291***	0.262***
		(0.030)	(0.032)	(0.027)	(0.028)
			[0.024]		[0.021]
30-44	1117	0.283***	0.287***	0.287***	0.274***
		(0.035)	(0.039)	(0.032)	(0.031)
			[0.030]		[0.024]

Notes: Robust clustered standard errors in parentheses. Bootstrapped standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Estimates of mobility for various age groups, obtained by changing the age restrictions in the sample. The estimates from the main analysis are shown in bold. Child income and parental income are measures of household income, and constructed as described in the text in section 2.3.2. OLS models include controls for the average age of parents and children, their average age squared and birth cohort dummies for both generations. TSRA models include controls for a quartic in age and year dummies in the first stage regressions for each generation. Source: BHPS and Understanding Society.

Table A.8 Differences in sample means by co-residency status

Variable	Co-residents	Non Co-residents	Diff.
Parental household income	3420.02	3693.87	-273.85***
Parental age when income observed	42.62	41.75	0.87***
Adult household income	4542.16	4002.11	540.05***
Adult personal income	1689.96	2064.08	-374.12***
Age when household income observed	29.40	30.30	-0.90***

Notes: This table presents the difference in sample means for the main variables splitting the sample by co-residency status. Significance levels are shown as *p<0.1, **p<0.05, ***p<0.01. No corrections for multiple comparisons applied. Source: BHPS and Understanding Society.

Chapter 3

Unpacking Intergenerational Income Mobility in the UK

3.1 Introduction

The overarching theme of this thesis relates to social mobility, of which two important components are the links between- and the transmission of socio-economic status across generations. So far, I have approached this topic by examining the degree of intergenerational income mobility in the UK at the national level. Whilst this has offered important insights into the strong intergenerational persistence at the household level, estimating the average mobility for the whole population can conceal important distinctions between subgroups of society. Investigating specificities across subgroups may prove valuable for targeted policy design that aims to improve social mobility. This chapter aims to deepen our understanding of intergenerational mobility and to obtain a more complete picture of social mobility in the UK by looking at differences in mobility between different groups of individuals as identified by a range of individual characteristics.

Using data from the harmonised BHPS and Understanding Society, this chapter expands upon the previous chapter on intergenerational income mobility by examining heterogeneities across three crucial dimensions highlighted by the international literature. Specifically, I investigate differences in intergenerational income mobility: (i) between sons and daughters; (ii) at different points of the income distribution; and (iii) by region of residence during childhood.

I begin this chapter by investigating whether the importance of family background as a determinant of socio-economic outcomes in adulthood differs for sons and daughters. Earlier work on intergenerational mobility has frequently downplayed the importance of studying income mobility for women, partly due to difficulties in obtaining data on women's income or earnings. However, shifting the focus to broader economic welfare, as measured by household income, rather than earnings, allows me to overcome this issue and examine income mobility for daughters. For example, we would expect income mobility to vary by gender if parents invest more in human capital of their sons than daughters, if there are differential returns to education by gender, or if assortative mating has differential consequences by gender.

This chapter advances this discussion by estimating the degree of intergenerational income mobility for sons and daughters in the UK. In doing so, in addition to the traditional income elasticities, I also compute the first rank coefficient estimates for daughters in the UK. While previous studies in the international literature find that daughters are more mobile than sons with respect to earnings (Altonji and Dunn, 2000; Bratberg et al., 2005; Hirvonen, 2008), more recent studies using family-based measures of income for two generations have obtained different results (Chadwick and Solon, 2002; Hirvonen, 2008). My findings corroborate these recent studies: household income is a similarly strong predictor of living standards in adulthood for both sons and daughters. Another aspect typically emphasised in the literature is the key contribution of assortative mating reinforcing the dynamics of intergenerational persistence (Chadwick and Solon, 2002; Hirvonen, 2008; Holmlund, 2020; Raaum et al., 2008). Here, I add to this literature by examining the role of assortative mating on income to intergenerational mobility of sons and daughters with partners. My findings provide evidence of assortative mating on income, especially for daughters. Further, I find that the contribution of assortative mating to income mobility depends on which partner contributes the bigger share in household income.

I then consider variations in intergenerational mobility based on another important aspect, namely at different points across the income distribution. While most studies focus on estimating the average persistence across generations, comparisons across the distribution of parent and child income are lacking and may provide key insights. For the purposes of improving equality of opportunity, for example, it is of interest to promote opportunities for upward mobility for individuals from disadvantaged backgrounds. As such, it is of particular interest for policy makers to be able to assess the degree of mobility at different points of the income distribution and to distinguish between upward and downward movements. Here, I examine differences in the level of income mobility at different points of the income distribution of both generations i.e. of parents and also of their children.

First, I address intergenerational transmissions across the distribution of parental outcomes. Examining mobility at the bottom of the parental income distribution - a lack of mobility here could be linked to the presence of credit constraints and potentially exacerbate income inequalities. Examining estimates of transition matrices, I find that there is more persistence at the tails of the parent's income distribution. Then, employing the estimators of rank directional mobility (Bhattacharya and Mazumder, 2011), I distinguish between patterns of upward and downward mobility for sons and daughters. Rather than only observing movements between quantiles as with transition matrices, these measures allow me to examine smaller rank movements that are obscured in the transition matrices. My findings indicate that there is a high degree mobility within quintiles of the parental distribution and provide modest evidence that daughters are more likely to be downwardly mobile and less upwardly mobile than sons across the whole parent distribution. Extending this analysis to differences in income mobility across the distribution of children's outcomes, I find that family background is an equally strong predictor of economic success across the children's income distribution.

To further complement this analysis, I examine regional differences in the degree of income mobility. Here, I address the question of whether the strength of the relationship between income across generations depends on the region where individuals grow up. For other countries, large regional differences in intergenerational income mobility have been observed (Acciari et al., 2019; Bratberg et al., 2017; Chetty et al., 2014; Eriksen and Munk, 2020; Heidrich, 2017). For example, Chetty et al. (2014) describes the US as being a 'collection of societies', rather than 'a land of opportunities'. The UK exhibits one of the highest levels of regional inequalities among developed countries (OECD, 2018*a*). However, very little is known about how intergenerational mobility vary across regions within the country. The presence of deep-rooted, persistent regional inequalities point to the existence of a stark ‘postcode lottery’(Social Mobility Commission, 2016*b*, 2017*a*), whereby regional characteristics determine the potential for social mobility. Here, I provide the first evidence on geographical disparities in intergenerational income mobility across regions in the UK.

I employ the concepts of relative and absolute mobility to characterise regions in terms of income mobility. First, I describe the levels of relative mobility estimating the traditional intergenerational elasticities and also the rank coefficients, revealing the size of the gap between the average incomes for children from high- and low-income families in each region. Then, building on the work by Chetty et al. (2014), I examine the levels of absolute mobility in each region using a rank-based measure that reflects the average income rank that a child attains in adulthood, if their parents are at the 25th percentile of the parental income distribution. In line with the literature from the US and Europe, I find that there is regional variation in both types of mobility. My results reveal a clear north/south divide. Relative and absolute income mobility are strikingly lower in the North of England than in the South of England. For absolute mobility, I also find that the South offers greater opportunities for upward mobility for children from relatively disadvantaged families. The mean income rank of a child whose parent was at the 25th percentile of their income distribution is 51.1 in the South and 38.2 in the North. Furthermore, an even more nuanced picture emerges when looking at a higher level of regional disaggregation. I observe variation in the degree of both relative and absolute mobility across regions in England and that the regions displaying high relative mobility are also those with high absolute mobility.

The remainder of the chapter is organised as follows. The following section reviews the literature on intergenerational mobility considering the specific aspects that will be analysed in this chapter. Section 3.3 presents the data and Section 3.4 presents the estimation strategy and a discussion of results separately for each of the dimensions analysed. The chapter ends with brief conclusions in Section 3.5.

3.2 Background and Related Literature

3.2.1 Intergenerational income mobility among daughters and sons

Most of the extant literature on intergenerational income and earnings mobility focuses only on men, using data on fathers and sons.[53] As discussed before, analyses of income and earnings mobility between generations are data intensive because information on income is required for at least two generations and over a sufficient period of time to derive adequate measures of long-term economic status. In general, the study of intergenerational mobility of daughters has been less prominent due to the paucity of appropriate data on women, together with the complexities that arise due to women's participation in the labour market. Lower (and non-random) labour force participation rate of women raises the incidence of non-observed earnings in adulthood and of part time work, and complicates the analyses linking parents and daughters.

This limitation in the literature has become more evident in recent decades, with the increase of women's labour market participation in most developed and developing countries. Until very recently, little was known about the extent of the intergenerational association for daughters, and whether it was of a similar nature to that for sons. Meanwhile, the availability of new sources of longitudinal data including measures of family and household income has accelerated this direction of research. Family and household income capture the broader living standards of families and as such are less affected by the difficulties in observing the economic status of women due to their pattern of participation in the labour market (Chadwick and Solon, 2002). These developments in data collection have enabled the inclusion of mothers and daughters which are not necessarily participating in the labour force in the analyses of intergenerational mobility for the first time. Moreover, instead of only considering adult children in isolation, this provides an opportunity to expand our understanding of how the

[53] A detailed review of the first studies on earnings mobility for sons is provided by Solon (1999) and a more recent review is done by Black and Devereux (2011).

dynamics of family and partnership formation affect the intergenerational transmission of advantages.

The use of family or household income distinguishes between individual and household resources, and reflects the pooling of resources between members of the household, bringing light to processes that can greatly influence intergenerational associations, such as assortative mating, marriage and intra-household labour supply decisions. The examination of income mobility using family-based measures emphasises the role of assortative mating as a contributing factor to intergenerational persistence. Assortative mating can be defined as "any non-randomness in the process of who mates with whom" (Chadwick and Solon, 2002, p.336).[54] The issue of whether individuals match on socio-economic characteristics is central for the study of social mobility dynamics, since the pooling of socio-economic advantages or disadvantages among couples can be an important source of social immobility.

Chadwick and Solon (2002) are pioneers in the analysis of intergenerational income mobility for daughters using measures of family income and with a focus on the contribution of assortative mating for mobility. In this framework, assortative mating (i.e. the systematic mate selection) is modelled as a correlation between the earnings of wives and husbands. Another similar model of assortative mating is developed by Ermisch et al. (2006). Unlike Chadwick and Solon, they consider the matching on the basis of human capital rather than income. Notably, for married (and also partnered) couples, the extent of the intergenerational association will depend on the contribution of partner's earnings to the household and on how closely partner's earnings are linked to one's family background (Chadwick and Solon, 2002; Holmlund, 2020). When couples match closely on traits related to their socio-economic status, stronger links between parents and their children's partners will be observed - as an indirect consequence of assortative mating, partners contribute to intergenerational persistence.

[54]In the context of marriage, assortative mating is also known as marital sorting. Sociologists and economists have long discussed how individuals match on socio-economic traits, usually associated with ethnic origin, religion, and family background characteristics.

Chadwick and Solon (2002) use the Panel Study of Income Dynamics (PSID) data on daughters to examine the issue for the US. Their results reveal smaller income elasticities for daughters, although these are not statistically different from those estimated for sons. Moreover, they show that assortative mating plays an important role in the intergenerational transmission of socio-economic status, with husband's earnings being similarly correlated with daughter's parents' income as the daughter's own earnings.

The evidence from this literature suggests that conclusions about the comparison by gender often depend on which the income measure being used reflects individual or family resources. Other studies that focus on family or household income also report similar levels of intergenerational income mobility for sons and daughters (Ermisch et al., 2006; Mayer and Lopoo, 2004). However, when focusing on intergenerational earnings elasticities, daughters have been found to be as mobile as sons with respect to fathers' earnings in the US (Altonji and Dunn, 2000), Norway (Bratberg et al., 2005), Canada (Blanden, 2005*b*) and Japan (Lefranc et al., 2014). Other studies have reported daughters to be more mobile than sons with respect to fathers' earnings in Australia (Fairbrother and Mahadevan, 2016) and Sweden (Hirvonen, 2008), and with respect to parents' earnings in Denmark, Finland, Norway, the UK and the US (Raaum et al., 2008).

Research examining intergenerational mobility for daughters and comparing levels of mobility by gender is relatively scarce for the UK. To my knowledge, the existing work focuses on either earnings or occupational mobility, partly due to the data limitations of the income data available in the National Child Development Study (NCDS) and British Cohort Study (BCS) cohorts. Using data from the NCDS, Dearden et al. (1997) observe similar intergenerational earnings elasticities for sons and daughters with respect to fathers' earnings, for the cohort of individuals born in 1958. Later, using data from the NCDS and also from the BCS, Blanden et al. (2004) find that the earnings elasticities with respect to combined parental income are only slightly different for sons and daughters, with some of the small observed differences falling short of statistical significance. However, Ermisch et al. (2006)

find that occupation mobility is higher for daughters than for sons using data from the first nine waves of the BHPS.[55]

The UK evidence on assortative mating is even more scarce. The early work by Atkinson (1983) suggested that it is very strong in the UK. The findings from Ermisch et al. (2006) based on occupational mobility suggest that marital sorting may be an important factor explaining the degree of intergenerational association in the UK. Blanden (2005*a*) provides further evidence on the degree of assortative mating based on education and income for the NCDS and BCS cohorts. She reports a strong association between husband's earnings and parental income in both cohorts. For sons, however, this association was only strong for the individuals born in 1970, indicating a change in the role of wives' contributions to family income over time.

To the best of my knowledge, no research has explicitly examined differences in *income* mobility among sons and daughters in the UK, making this chapter the first account on income mobility estimates by gender with specific focus on the transmission of resources at the household level.[56] It is also the first study to present estimates of rank coefficients for daughters in the UK. Moreover, using the framework developed by (Chadwick and Solon, 2002), I consider the role of assortative mating as one of the mechanisms through which the intergenerational transmission of economic status takes place. This chapter adds to this literature, being the first to provide evidence on assortative mating on income using the harmonised BHPS and Understanding Society data.

3.2.2 What kind of mobility are we talking about?

After examining differences in intergenerational mobility by gender, the second part of this chapter is dedicated to further characterising relative mobility. On studying the broader topic of social mobility, intergenerational elasticities and rank coefficients can prove very useful

[55] Using occupational prestige scores of fathers and their offspring as a predictor of permanent income.

[56] The study by Blanden (2005*a*) using the NCDS and BCS data is the most closely related to mine; however it considers combined earnings for the second generation, which is slightly different.

in summarising intergenerational mobility, but not without some important caveats. First, these aggregate measures do not allow us to distinguish between directional movements, that is, between upward and downward mobility. Second, because these estimates express average associations, they are not well suited to examine differences in mobility across the income distribution. These issues are of key importance for academics, government and policy makers because much of the interest about intergenerational mobility is motivated by concerns surrounding the potential for upward mobility, and related to providing equal opportunities to the more socio-economically vulnerable and disadvantaged members of society.

Aiming to overcome these issues, a specific strand of the intergenerational literature has focused on examining differences in mobility across the income distribution and analysing directional movements. Among the studies that examine how intergenerational mobility varies across the parental income distribution a variety of methods have been used, including transition matrices (Blanden et al., 2004; Corak and Heisz, 1999; Couch and Lillard, 2004; Dearden et al., 1997), non-linear regressions (Bratsberg et al., 2007), non-parametric techniques (Bratberg et al., 2017; Grawe, 2004) and, estimators of directional rank mobility (Bhattacharya and Mazumder, 2011; Corak et al., 2014).

Conventional wisdom and previous literature suggest that mobility might be lower (i.e. persistence is higher) at the extreme tails of the parental income distribution. The well known theoretical model by Becker and Tomes (1979, 1986) predicts that the association between parents' and children's earnings is stronger for poorer families because they face stricter borrowing constraints than richer families. The hypothesis of concavity in this relationship is referred to as the Becker-Tomes conjecture in the literature (Bratsberg et al., 2007; Grawe, 2004). In a system that requires some private financing of education, if poorer parents are more financially constrained this will impact negatively on their ability to invest in the education of their children. As a result, children from poorer families will have lower earnings in adulthood compared to those of similar ability with non- or less constrained parents. This hypothesis could explain a stronger intergenerational association at the bottom of the distribution.

The first studies in this area have focused on testing the Becker-Tomes conjecture and determining the extent to which borrowing constraints affect intergenerational earnings correlations. The results from this small literature suggest that the patterns of intergenerational association differ between countries, and have often provided contrasting conclusions about the validity of the Becker-Tomes conjecture. Mazumder (2005) provides evidence that supports this hypothesis for the US. Splitting the sample by net worth, he estimates a significantly higher earnings elasticity for families of low net worth. However, other studies have reached different conclusions. For example, Corak and Heisz (1999) describe an S-shaped pattern of earnings elasticity for Canada using kernel density techniques. They find that the intergenerational earnings association is very low for children whose parents are at the bottom of the distribution, and becomes stronger towards the middle of the distribution. A similar pattern is also described in Grawe (2004) using regression splines. Another study by Bratsberg et al. (2007) presents a cross-country examination of non-linear patterns for the US, the UK, Denmark, Finland and Norway. They find evidence that the intergenerational earnings association follows a convex pattern in Nordic countries, while a linear relationship is observed for the US and the UK.

Many studies have also examined non-linearities across the parental income distribution using transition matrices (Blanden et al., 2004; Corak and Heisz, 1999; Couch and Lillard, 2004; Dearden et al., 1997). This approach affords the opportunity to examine how child-parent pairs are positioned across the earnings/income distribution and how this position changes from one generation to the other. Further, it is possible to observe the direction of movement between quantiles. Results from previous studies suggest a ‘stickier’ relationship between children’s and parent’s incomes for individuals whose parents are at the bottom and top of the income distribution in the US, UK, Canada, Denmark, Sweden, Finland and Norway (Corak and Heisz, 1999; Jäntti et al., 2006).

For the UK, most work on mobility levels across the parental income distribution relies on the estimates from transition matrices (Blanden et al., 2004; Dearden et al., 1997). The study by Dearden et al. (1997) using data for children from the NCDS cohort reveals that, for fathers in the top quartile, a very large proportion of sons (52%) and daughters (48%) remain in the same quartile. The authors also note an asymmetry, with upward mobility out of the bottom quartile being more likely than downward mobility out of the top quartile. Similar results are also obtained by Blanden et al. (2004) for the NCDS and BCS sons and daughters. In addition, Blanden and colleagues observe that the proportion of children in the same quartile as their parents increased between the two cohorts, reinforcing their findings of declining mobility between 1958 and 1970. Transition matrices can reveal important asymmetries in the probabilities of quantile transitions and can be indicative of heterogeneity in mobility across the income distribution and also of non-linearities in the intergenerational mobility function. However, the main drawback of this approach is that only movements between quantiles are registered, meaning that important information about mobility within these blocks are obscured.

Recent methodological developments related to the increasing use of rank-based measures of mobility have promoted new ways of examining directional movements focusing on the relationship between the rank position of children and their parents. With the aim of expanding the analysis beyond that of transition probabilities between quantiles, Bhattacharya and Mazumder (2011) have developed a framework to examine directional rank mobility at different points of the income distribution in a more systematic manner. This approach consists of comparing the positional ranks of both generations, and examining the direction and size of the movement in ranks between generations. Estimators of directional mobility have been recently used by Corak et al. (2014) to compare earnings mobility between the US, Canada, and Sweden. One key advantage of this rank-based approach is that it is well suited for comparisons across subgroups of the population.

A second, smaller strand of this literature has focused on how intergenerational mobility varies across the distribution of children's incomes or earnings in adulthood (Eide and Showalter, 1999; Grawe, 2004; Gregg et al., 2019; Palomino et al., 2018). These studies usually employ the conditional quantile regression (QR) technique developed by Koenker and Bassett (1978). While OLS minimises squared residuals and fits the linear model providing estimates at the mean of the distribution, conditional QR fits the model for quantiles, minimizing absolute residuals at any specific quantile of the conditional distribution (Koenker and Bassett, 1978; Koenker and Hallock, 2001). Other recent studies have also used unconditional quantile regressions (Gregg et al., 2019; Palomino et al., 2018; Schnitzlein, 2016), which allow for an interpretation of marginal effects of covariates on the unconditional distribution of the dependent variable and are useful for interpreting coefficients from different model specifications.[57]

Eide and Showalter (1999) find that mobility is lower at the bottom than at the top of the son's conditional earnings distribution in the US. That is, family earnings are a less important factor in explaining son's earnings at the top of the distribution than at the bottom of the distribution. The recent study by Palomino et al. (2018) uses conditional and unconditional QR and find a U-shaped pattern for the US using both methods, with higher income persistence at both tails of the distribution. For the UK, Gregg et al. (2019) estimate unconditional QR using the BCS data for the cohort of individuals born in 1970. They report a J-shaped pattern in the relationship between parental income and sons' earnings, although this seems to be driven by a stronger association at the very top of the distribution at the 90th percentile.

I expand the analysis of intergenerational income mobility to examine these issues and provide a deeper understanding of relative mobility in the UK. I begin by examining heterogeneities in relative income mobility across the parental income distribution using transition matrices and estimators of directional rank mobility. Linking to the previous discussion, I also examine these patterns by gender. I then expand the analysis to examine differences across the children's income distribution using quantile regressions techniques. By taking both approaches, I am able to build a more comprehensive picture since these complementary methods address the question of who experiences mobility from different perspectives.

[57]More details on quantile regressions are provided later in section 3.4.2.

3.2.3 The UK postcode lottery

In the final part of this chapter, I examine heterogeneities in the degree of intergenerational income mobility across regions of the UK. This question is motivated by the observation that the UK is considered one of the most regionally unequal countries in the group of developed countries. The OECD reports that the UK has the 6th highest regional economic disparities of GDP per capita among 30 OCED countries, and experienced the 4th largest increase in disparities between 2000 and 2016 (OECD, 2018*a*).

Historically, British society has been marked by many different types of divides (e.g. of class, gender, ethnicity), with those of a geographical nature also being important. Striking and growing disparities within and between regions have been widely observed by researchers with respect to various socio-economic indicators, such as income, economic activity, health services, educational achievement and the availability of opportunities.

A historical north/south divide has been widely described in the literature, contrasting the richer, dynamic regions in the south, especially in the South East and London, with the poorer, more stagnant regions in the north. Notably, income and economic activity are very unevenly distributed across the country. Recent statistics for regional gross disposable household income per capita for 1997-2017 released by the Office for National Statistics (2019*a*) reveal that income is much more concentrated in the south of the country. These statistics also reveal that the patterns of concentration have remained mostly unchanged during the last two decades.

Many of these regional disparities are product of long-term processes that reinforce deep-rooted persistent inequalities. The highly centralised administrative system in the UK also contributes to regional imbalances by often failing to design effective policies that address the bigger issues while accounting for regional particularities. As Andy Haldane recently pointed out, “Although the UK has a long history of regional disparities, they are currently at their highest levels in more than a century. Chopping and changing UK regional policies has contributed to these widening differences" Haldane (2020).[58] Not only is there a lack of consistency and continuity in regional policy, but the levels of investment made by local government are relatively low. In 2016, only 34% of total public investment was made by subnational governments in the UK, compared to the OECD average of 57% (OECD, 2018*a*).

Deep regional inequalities in the UK exist beyond the simple north/south or rural/urban divisions (Bell et al., 2018; Friedman and Macmillan, 2017; OECD, 2018*a*). While many people feel 'left behind', others experience the results of growth and economic dynamism. These growing disparities can contribute to social disruption and political polarisation and have become frequent topics in political agendas in the UK, as demonstrated by the Brexit vote and with pledges from Boris Johnson's government to 'level up' under-performing regions and reduce regional imbalances. High levels of regional inequality are not only harmful to the regions that lag behind and do not fulfil their potential, but also to the richer regions, which face increased pressures on prices and to provide goods and services to support the dynamic economy.

Recent work by the Social Mobility Commission (SMC) has emphasised the existence of striking regional disparities in social mobility indicators such as educational attainment, labour market outcomes and home ownership (Social Mobility Commission, 2016*b*, 2017*a*). Specifically, using the so called Social Mobility Index (SMI), the SMC emphasises differences in opportunities across local authority areas (LAs) in the UK by looking at a range of educational outcomes for children and young people from disadvantaged backgrounds, as well as the standing of the local job and housing market in those areas.[59] [60] According to their performance in the Social Mobility Index, LAs are classified in coldspots and hotspots

[58] Haldane is the current Chief Economist at the Bank of England and Chair of the Industrial Strategy Council.

[59] Local Authority areas (LAs), are a geographical sub-national division of England - there are 324 LAs in England. According to the ONS, this is the general term used for a body administering local government.

[60] More specifically, the Social Mobility Index includes: (i) the educational outcomes of those of poorer backgrounds (eligible to free school meals) in each area, including early years, primary and secondary school, post-16 outcomes and higher education participation; and (ii) outcomes achieved by adults in the area, including average income, prevalence of low paid work, availability of professional jobs, home ownership and house affordability. For more information, see Social Mobility Commission (2016*a*).

of social mobility, emphasising regional differences in the performance of disadvantaged children and in the availability of opportunities for adults.

Figure 3.1 shows the regional distribution of LAs according to their performance in the Social Mobility Index in 2017. The vast majority of LAs in the top quintile can be found in London (45%) and the South East of England (25%), followed by the East of England (14%). These areas perform extremely well in comparison to the rest of the country. The Commission also identifies a large number of coldspots of social mobility particularly in the Midlands, with 31% of LAs in in the bottom quintile of the SMI being located in the East Midlands and 17% in the West Midlands. These areas provide the worst social progress opportunities for individuals from disadvantaged backgrounds.

As the SMC points out, these findings reinforce the idea that

> “[...]Britain’s social mobility problem is not just one of income or class background. It is increasingly one of geography. A stark social mobility postcode lottery exists today, where the chances of someone from a disadvantaged background getting on in life is closely linked to where they grow up and choose to make a life for themselves.” (Social Mobility Commission, 2017*a*, p. 2)

This idea is echoed in people’s beliefs about social mobility in the UK. The Social Mobility Barometer[61] reports that the majority of people (77%) think that there is a large gap between social classes in Britain, with almost half saying that where you end up is determined by your background. Most also believe in the existence of a postcode lottery, with around 76% of people saying that there are ‘fairly or very large’ differences in opportunities available in different parts of Britain (Social Mobility Commission, 2019).

[61] The Social Mobility Barometer is an annual survey organised by the Social Mobility Commission. They interview around 5000 people each year since 2017, asking questions related to attitudes to social mobility in the UK.

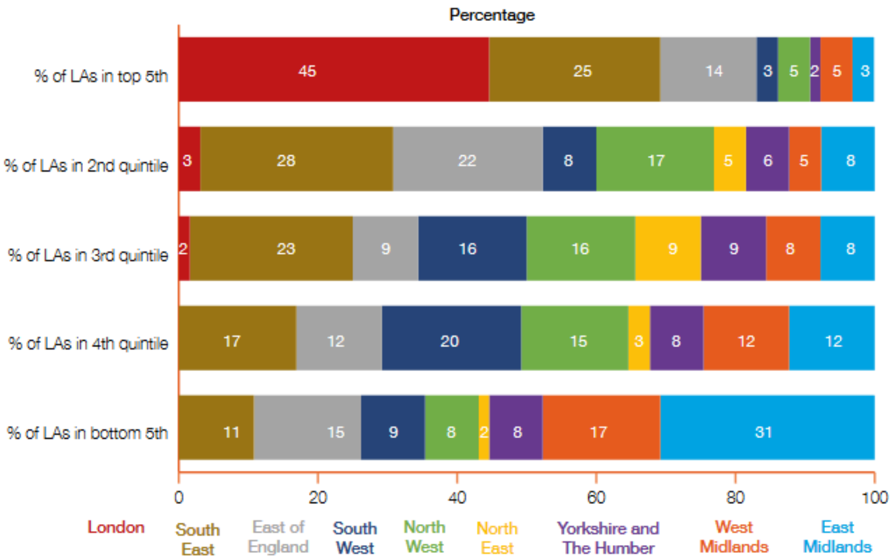


Figure 3.1 Percentage of local authorities (by region) within each Social Mobility Index performance quintile

Source: Social Mobility Commission (2017*a*).

This work by the Social Mobility Commission points to the existence of several regional disparities across the country in terms of social mobility. However, the Social Mobility Index focuses on the *intragenerational* dimension of social mobility, and as the SMC acknowledges, it is only a limited proxy of spatial differences in*intergenerational* mobility. Whilst the SMC reports that there is ongoing research using administrative data to estimate income mobility by regions, with the exception of the studies by Friedman and Macmillan (2017) and Bell et al. (2018), regional differences in intergenerational mobility in the UK remain largely unresearched. To the best of my knowledge, only these two studies examine regional variation in the degree of intergenerational mobility in the UK. Differently to this chapter, they focus on occupational, educational and housing mobility.

Despite the scarcity of evidence on regional variations in intergenerational mobility for the UK, this has been increasingly the focus of recent studies of intergenerational mobility in the international literature. Looking beyond national estimates of social mobility is seen as an important way to shed light on potential mechanisms that underlie the intergenerational association.

This literature started with the influential work by Raj Chetty and colleagues (Chetty et al., 2014) who examine regional differences in intergenerational earnings mobility the US. Chetty et al. (2014) describe the US as being a collection of societies, some of which are ‘lands of opportunity’ and display high levels of mobility, while in others there are few opportunities available for disadvantaged children. Using a very large administrative data set containing tax records for more than 40 million children and their parents, Chetty et al. (2014) estimate intergenerational mobility elasticities and rank coefficients across small areas called commuting zones[62] in the US. The study finds substantial variation in relative and absolute mobility across these small geographical areas. Their findings suggest that children who grow up in certain states and cities have better odds of experiencing upwards mobility than similar children born elsewhere.

Following this study for the US, a growing body of research has examined regional variation in intergenerational mobility for other countries. Among these studies, Bratberg et al. (2017) have observed variations in intergenerational rank mobility between big regions in Germany, Norway and Sweden. For Sweden, the paper by Heidrich (2017) uses administrative data to estimate income elasticities, rank coefficients and absolute upward mobility across regions. In contrast to Chetty et al. (2014), Heidrich (2017) use a multilevel model approach to understand regional variation in income mobility, in order to account for differences in population size across Swedish regions. She finds that many regions differ statistically significantly from the Swedish average in terms of absolute mobility, but only a few differ when examining relative mobility. Other recent papers by Güell et al. (2018) and Acciari et al. (2019) investigate regional differences in intergenerational income mobility in Italy. Using administrative data on tax returns for more than half a million parents-child pairs, Acciari et al. (2019) find that children who grow up in provinces in the North of Italy, the richest area of the country, have more opportunities. This region has 3-4 times more upward mobility than the South. Similar regional variation within Italy is observed by Güell et al. (2018), although

[62]Commuting zones are geographical aggregations of counties, similar to metro areas, but covering the entire country (Chetty et al., 2014).

these authors depart from the usual income and earnings data and use informational content of surnames and individual incomes from tax records to examine intergenerational mobility.[63]

The observation of regional variation in mobility within countries can provide key insights of the underlying local characteristics which are present within regions with higher and lower mobility. The work by Chetty et al. (2014) indicates that high relative mobility is correlated with high absolute upwards mobility and also with less residential segregation by race and income, less income inequality, better primary schools, greater social capital and greater family stability. Undertaking a similar exercise, other recent studies have found that intergenerational mobility is positively correlated with economic activity and labour market conditions, education, indicators of family instability and social capital (Acciari et al., 2019; Eriksen and Munk, 2020; Güell et al., 2018).

For the UK, the evidence on regional disparities in intergenerational mobility is very scarce. Friedman and Macmillan (2017) are the first to examine regional differences in occupational mobility in England, Scotland and Wales. They find an interesting result, which brings into question the presentation of London as the engine-room of social mobility. Instead, they find that it has the lowest rate of upward mobility and the highest rate of downward mobility. Another recent study by Bell et al. (2018) examines regional variation in intergenerational occupational, educational and housing mobility, using the decennial census data from the Longitudinal Study of England and Wales (LS). They find substantial regional differences, with occupational mobility being very high in London and rather lower for ex-industrial and mining areas. London also has the highest educational mobility, while Yorkshire and Humberside in the North has the lowest. When looking at housing mobility, they observe a different pattern, with London having the lowest mobility, and Wales the highest mobility, which indicates that home ownership mobility is negatively related to average house prices across regions and is not correlated with occupational mobility.

[63] As described in Güell et al. (2015), this method relies on indications of family linkages based on rare surnames. They generate estimates of intergenerational mobility by examining the 'inheritance' of economic outcomes across generations of individuals who share a surname that is relatively rare i.e not shared by many people in the region.

Looking at these differences in the measures of mobility, Bell et al. (2018) emphasise that geographical comparisons based on one dimension of mobility do not necessarily reflect those based on alternative measures. Therefore, an important gap in the literature persists, namely, on the geography of income mobility in the UK. The BHPS and Understanding Society data provides information that allows me to observe income and earnings for multiple cohorts and, simultaneously, focus on the geography of income mobility, complementing the previous analysis undertaken by Friedman and Macmillan, and Bell and colleagues. I add to this literature, focusing on measures of income mobility using household income, which reflect the broader availability of economic resources within households across the country. To my knowledge, this chapter provides the first regional estimates of intergenerational income mobility for the UK.

Taken together, the evidence on regional variations in other countries, the work on social mobility by the UK's SMC and the recent evidence pointing to regional differences in occupational, educational and housing mobility by Friedman and Macmillan (2017) and Bell et al. (2018) suggest that regional differences in intergenerational income mobility could also be present in the UK. The existence of these differences in how economic advantage is transmitted across generations *within* the country would reinforce the evidence pointing to the existence of a 'postcode lottery', suggesting that the importance of family background for living standards in adulthood also depends on the region of upbringing. Evidence of this would complement the analysis currently provided by the Social Mobility Commission and would be key for informing effective policy decisions. If, in some regions, it is easier or harder for disadvantaged individuals to climb the social ladder, this variation could bring insights into local conditions and other causal factors that can promote more equal opportunities.

3.3 Data

Using the BHPS+UKHLS data, I consider the main sample of all pairs of parents matched to their children, using the same sample selection criteria defined in Chapter 2.

For the analysis of intergenerational income mobility by gender, it is noteworthy that this sample is almost evenly split by gender. In addition, Table 3.1 reports differences in sample means by gender for the main variables used in the analysis. A more complete table of summary statistics splitting the sample by gender is presented in Table B.1 in Appendix B. From Table 3.1, daughters have noticeably lower levels of household income in adulthood, on average £272 per month lower than that of sons, although this income is observed at similar ages for sons and daughters, at around age 30. Personal and labour incomes are lower for daughters, on average, possibly due to women's reduced participation in the labour force. The average parental household income is more similar, although still slightly lower for daughters, as well as the age at which parental income is observed. Finally, daughters in the sample are more likely to live with a partner (62%) than sons (57%), and among those who have a live-in partner, they are also more likely to be officially married (37.5%).

Table 3.1 Differences in means by gender

Variable	Sons	Daughters	Diff.
Parental household income	3660.0	3580.1	79.9**
Parental age when income observed	42.3	41.8	0.5***
Adult household income	4290.3	4017.9	272.4***
Adult personal income	2262.9	1712.0	550.9***
Adult labour income	2129.5	1376.6	752.9***
Age when household income observed	30.2	30.1	0.1
Child age when parental income observed	13.0	12.9	0.1**
Has live-in partner (%)	57	64	-7***
N	997	1129	

Notes: This table presents the difference in sample means for the main variables splitting the sample by gender. Significance levels are shown as *p<0.1, **p<0.05, ***p<0.01. No corrections for multiple comparisons applied. All income variables are pre-tax income measured in real pounds per month and constructed as described in the text in section 2.3.2. Source: BHPS and Understanding Society.

For the regional analysis, I split the main matched sample of pairs of parents/children according to childhood location, which is defined as the region where the children grow up.[64] In contrast to Chetty et al. (2014), who assign childhood location based on the location where individuals lived at age 16, I define childhood location as the region where a child lived for the

[64] I have no reason to expect that the pattern of attrition and non-response in the data would affect households non-randomly by regions and so this should not bias the results of comparisons between regions.

longest time before the age of 18 (and parental income was observed), which better captures region of upbringing.

3.4 Estimation and Results

3.4.1 Intergenerational income mobility by gender

Estimation

The estimation of income mobility by gender follows the same methodology employed to estimate intergenerational income mobility for the UK at the national level in Chapter 2. The intergenerational income elasticity is estimated by using an OLS regression of log income for children (observed in adulthood) on the log of income for parents (observed in childhood). The income for each generation is obtained by averaging income across multiple periods to alleviate the biases from measurement error. The OLS approach has the drawback that, in order to control for age effects, the ages must be also averaged for the whole period of observation. Alternatively, as I propose in Chapter 2, a measure of long-run income for each generation adjusted for age and time effects can be obtained by using the TSRA approach.

In order to obtain the estimates of income mobility for sons and daughters, the models are estimated separately by gender. In the TSRA approach, this means estimating the two stages separately by gender.[65] Furthermore, when estimating rank coefficients, the ranks of sons and daughters are based on the national distribution (and not only individuals of the same gender subgroup).[66] This generates comparable results for these subgroups based on a common rank distribution.

The analysis of assortative mating is based on the framework developed by Chadwick and Solon (2002). Here, I focus on the intergenerational elasticities estimated for sons

[65] Although the results obtained are very similar when estimating only the second stage equations separately by gender.

[66] This is also true when using the TSRA approach, since the ranks are defined after the first stage regressions.

and daughters with partners. Varying the income measure used as the dependent variable (e.g. personal income, partner's income, couple's income) while keeping constant the main explanatory variable (parent's household income), I am able to analyse the extent of assortative mating on income by estimating the association between the incomes of parents and their children's partners.

Finally, as discussed previously in Chapter 2, the usual TSRA standard errors should be adjusted to take into account the sampling variance of generated outcomes (particularly, regressors) from the first stage. The results presented in Chapter 2 reveal, however, that the bootstrapped standard errors were generally smaller than the corresponding robust standard errors clustered on parents. For this reason, the results in this chapter present the most 'pessimistic' robust standard errors, not bootstrapped.

Results

I begin by estimating intergenerational income elasticities and rank coefficients separately for sons and daughters, with respect to their own parents' income, using two different measures of income as the dependent variable. The results are displayed in Table 3.2. Panel A presents the intergenerational income elasticities for sons and daughters. The estimated elasticity for sons is 0.252 and for daughters it is 0.261 obtained by the TSRA method (Column 2). These point estimates are remarkably similar and as demonstrated by the test of equality of coefficients shown in the third row of the table, I fail to reject the null hypothesis that these estimates are equal. Then, looking at the results for personal income, I obtain an elasticity of 0.331 for sons and of 0.217 for daughters using TSRA (Column 4). These coefficients are statistically significantly different only at the 10% level.

Panel B presents the estimates of rank coefficients by gender. Considering household income, the estimated rank coefficient is 0.312 for sons and 0.275 for daughters using TSRA (Column 2). Here too the small difference between the point estimates was not significant. Reinforcing these findings, the rank coefficients obtained using personal income are slightly lower for daughters (0.244) than for sons (0.288) (Column 4), indicating that daughters are slightly more mobile, but again the difference between these estimates is small and falls short of statistical significance.

Table 3.2 Estimates of intergenerational income elasticities and rank coefficients by gender

	Household income			Personal income		
	OLS	TSRA	N	OLS	TSRA	N
	(1)	(2)		(3)	(4)	
Panel A: IGE						
Sons	0.266***	0.252***	997	0.339***	0.331***	981
	(0.044)	(0.031)		(0.051)	(0.045)	
Daughters	0.269***	0.261***	1129	0.232***	0.217***	1118
	(0.035)	(0.030)		(0.034)	(0.036)	
Sons = Daughters (p-value)	0.966	0.874		0.080	0.057	
Panel B: Rank Coefficient						
Sons	0.310***	0.312***	997	0.271***	0.288***	988
	(0.032)	(0.033)		(0.033)	(0.032)	
Daughters	0.292***	0.275***	1129	0.240***	0.244***	1126
	(0.034)	(0.031)		(0.030)	(0.030)	
Sons = Daughters (p-value)	0.695	0.348		0.481	0.423	

Notes: Robust clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Child income and parental income are household income. These measures are described in the text in section 2.3.2. OLS models controls for the average age of parents and children, their average age squared and and birth year dummies. TSRA models control for a quartic in age of parents and children, as well as year dummies. The ranks of sons and daughters are based on the national distribution. The p-value for the t-test of equality of coefficients also reported in the table. This was obtained with the fully interacted model including the female dummy. Source: BHPS and Understanding Society.

Overall, I find a similar level of income mobility for sons and daughters in the UK. This suggests that family background is an equally strong predictor of economic status in adulthood, regardless of gender. These are the first estimates of income mobility by gender using household income as a measure of economic status for the UK, and the results obtained mirror the findings from other countries, such as the US, Sweden and Mexico (Chadwick and Solon, 2002; Hirvonen, 2008; Raaum et al., 2008; Torche, 2015), although differences in the levels of mobility between these countries and the UK are observed. When mobility is estimated using personal income at the individual level, I find that British women are slightly more mobile than men, although this difference is mostly not significant. This is also in line with findings of previous research for the UK on earnings mobility (Blanden et al., 2004; Dearden et al., 1997).

When examining the estimates based on personal income, it is notable that these point estimates are more different by gender that the corresponding estimates based on household income. While it is not possible to reject the equality of the rank coefficients, the results indicate that the IGEs are different at 10%. The conceptual differences between the two measures of mobility may provide key insights into the reasons for this apparent contrast. While the rank coefficients are standardised, IGEs take into account the ratio of standard deviations of the income distributions of the two generations. The standard deviation of parent's household income is similar for both genders, but in the offspring's generation it is higher for sons than for daughters when considering personal income (see Table B.1 in Appendix B).[67] This means that the observed difference in IGEs is likely driven by differences in the variance of the distributions of personal income between sons and daughters, rather than by a difference in the linear association between child and parental incomes by gender.

Here, it is important to acknowledge that further research is needed to obtain more comparable measures of long-run income for sons and daughters at the individual level. This is required for a better understanding of whether differences in mobility by gender (or lack thereof) are genuine or if the patterns observed are a consequence of estimates that are not directly comparable. As labour earnings are usually the main source of personal income, this suggests that the patterns of labour market participation of women likely play a key role in the estimation of intergenerational income mobility by gender. For example, considering only women with positive earnings (e.g. employed) is likely to over-estimate income mobility of daughters. Even with the increased participation of women in the labour force in recent years, they might still have different earnings profiles than men due to periods out of employment related to child bearing and caring around the age at which income variables are typically observed (Jäntti and Jenkins, 2015). Therefore, in order to obtain more comparable proxies of long-run income for men and women, it is important, not only, to account for selection into employment, for example using the Heckman procedure as shown in Blanden (2005*a*), but to also account for differences in the patterns of full-time and part-time work by gender. This

[67]Note that this is not the case for labour earnings, here the standard deviation is higher for daughters.

could be done, for example, by adjusting the measures of income or earnings for hours worked. Since these labour supply decisions are usually related to fertility choices, the presence of children in the household and age, additional research is also needed to fully comprehend the consequences of the life cycle bias for the estimation of income mobility for women. Finally, as the intergenerational literature is very focused on earnings mobility, it frequently neglects the importance of non-labour income sources. Another research avenue that deserves more attention is the study of intergenerational associations considering these alternative sources of income.

The role of assortative mating

Complementing this analysis I also examine the role of assortative mating for the intergenerational mobility of partnered sons and daughters. Bringing children's partners into the analysis can provide additional insights into the relationship between sons' and daughters' household incomes and that of their parents. Since partner's earnings and income contribute to household income, the degree of intergenerational income persistence observed considering household income will depend on the magnitude of the partner's contribution to the household and on the extent of assortative mating, that is, how closely couples match on traits that are correlated with their parents' incomes.

To examine how mating on income contributes to intergenerational mobility, I now focus on the sample of sons and daughters with cohabiting partners, which corresponds to a total of 567 sons and 727 daughters. Since the analysis will require information on partner's income, the cases in which partner's personal income is missing or zero must be dropped.[68] Thus, I am left with a sample of 454 sons and 609 daughters.

Table 3.3 provides the estimates of intergenerational persistence of income for sons and daughters, their partners and for the couple as a whole. Firstly, I estimate the IGEs for this subsample of sons and daughters. The estimate of income elasticity is 0.193 for sons and

[68] Similarly to the sample restrictions applied to the children in my sample, partner's income is only captured between the ages of 25-60 and while they are a partner of an individual (i.e. not after separation/divorce).

0.196 for daughters when the dependent variable is household income (Row 1).[69] For personal income, it is 0.300 for sons and 0.245 for daughters (Row 2).

Next, to examine the degree of assortative mating on income, I estimate IGEs using the partner's personal income (Row 3) and the combined couple's personal income (Row 3) as dependent variables. The first interesting result is that the relationship between partner's income and in-laws incomes is positive and significant, 0.109 for sons and 0.149 for daughters (Row 3). Among daughters with partners, partner's personal income is considerably elastic with respect to the parent's income. This provides evidence of intergenerational assortative mating based on income. In addition, comparing the estimates obtained for cohabiting partnered couples and married couples in Row 3, the IGEs for married couples are larger (although less precisely estimated, due to a reduction in the sample size). This suggests that the relationship between partner's income and their in-laws' income is stronger for married couples, especially for daughters.

The elasticity of a couple's combined income to parents-in-law income is influenced by the share of contribution from the income of the sample member and that of their partner, as has been shown in other studies (Blanden, 2005*b*; Chadwick and Solon, 2002). By definition, the elasticity of the couple's combined income (in Row 4) can be decomposed, and written as the weighted average of the elasticity of the individual with respect to their parents' income (Row 2) and the elasticity of the partner's personal income with respect to their in-laws income (Row 3). When the share of partner's contribution is higher, then the association observed between partner and parents-in-law will be stronger. This explains why the elasticities of partner's income with in-laws is closer to those of daughters' own income and their parents. This also suggests that among the sample of partnered sons, sons and their partners contribute almost equally to household income. On the other hand, among daughters with partners, the

[69]These results are similar to those obtained for the sample of all partnered sons and daughters without restricting the sample to sons and daughters whose partners have positive income - the IGE for all 567 sons with partners is 0.198 and for all 727 daughters with partners it is 0.180. This is reassuring that the estimates presented in the table are most likely not biased by restricting the sample to those with positive income.

contribution of partner's income is bigger than the contribution of daughter's income. These results indicate that, on average, daughter's partners contribute the larger share of income.

Moreover, I observe that the relationship between partner's income and parent's income is stronger for married couples than for partnered couples who live together. A possible explanation for this is that partnership could be considered a less established relationship, a lighter version of marriage, for which, on average, matching on socio-economic characteristics is exhibited to a lesser degree. However, as the individuals in my sample are relatively young, it is not possible to disentangle the reasons why couples decide to live together or to get married and if the number of married couples will increase as individuals get a bit older. If this is the case, assortative mating on income would be even stronger if we were to observe these individuals at later ages.

Table 3.3 Intergenerational income mobility and assortative mating

	IGE (β)			
Dependent variable	Sons w/ partners	Daughters w/ partners	Married sons	Married daughters
(1) Household income	0.193***	0.196***	0.176***	0.174***
	(0.044)	(0.033)	(0.059)	(0.045)
(2) Personal income	0.300***	0.245***	0.273***	0.202***
	(0.057)	(0.051)	(0.060)	(0.066)
(3) Partner's personal income	0.109*	0.149***	0.134	0.202***
	(0.063)	(0.060)	(0.085)	(0.060)
(4) Couple's personal income	0.212***	0.234***	0.222***	0.224***
	(0.048)	(0.044)	(0.059)	(0.047)
N	454	609	292	390

Notes: Robust clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Income for the younger generation (child/couple/partner) varies across the rows of this table. For parents, income is household income. The sample is restricted to individuals with data on their partner's personal income and excludes the cases for which partner's income is zero. IGEs are estimated using TSRA models that control for a quartic in age of parents and children, as well as year dummies. Source: BHPS and Understanding Society.

The analysis of intergenerational mobility for a subsample of partnered and married individuals provides evidence of assortative mating on income. I find a positive and significant relationship between partner's personal income and the income of their in-laws and the results using couple's combined income demonstrate that income persistence between parents and partners does contribute to intergenerational persistence. This indicates that the matching of individuals of similar socio-economic status is an important factor contributing to the persistence of inequalities in British society. This is in line with the findings of studies from the international literature, Chadwick and Solon (2002) for the US, Hirvonen (2008) for Sweden and Blanden (2005*b*) for Canada. For the UK, a previous study by Blanden (2005*a*) reports an elasticity between partner's earnings and in-laws of 0.306 for sons with partners and 0.239 for daughters with partners using data from the BCS cohort. The extent of mating based on occupational mobility reported by Ermisch et al. (2006) is also larger than what I find. These studies have found a considerably stronger degree of assortative mating than what is reported here, which could be a product of genuine cohort differences (e.g. in the timing or importance of income for finding a partner) or of differences in the age at which sons and daughters are observed in adulthood. As discussed previously, observing income data at younger ages would not only provide noisy information on long-run income but could also introduce selection in the sample. This is because age is related to other characteristics that can influence labour supply decisions and therefore the generation of income, such as cohabiting with a partner, getting married or having children. The relatively younger sample of the BHPS+UKHLS is less likely to be partnered, to be legally married and to have completed their fertility.[70] This implies that it is possible that the results presented here underestimate the real extent of assortative mating and that its contribution for the intergenerational transmission of economic status would be even greater.

When income mobility is based on household income, this measures the extent to which 'living standards' or welfare are associated across generations. This approach assumes that all resources are pooled within the household and that labour supply and consumption are the observable result of the maximization of household well-being preferences, constrained by a household budget restriction. Household well-being is therefore a product of endogenous decisions by its members. It is important to acknowledge that this approach implicitly relies on the assumption that a household, even if it consists of multiple members, acts as a single decision-making unit. This implies that the source of income does not play a role in the household's allocation with regard to labour supply and consumption. While this does not

[70] The proportion of couples who are legally married in the BCS was 60% for men and 69% for women (Blanden, 2005*a*). These shares are similar to those observed here for sons and daughters who have a cohabiting partner, but the proportion of married couples is considerably smaller.

mean that household arrangements and decisions are completely free of gender-based roles, the existence of economies of scale, household public goods and income pooling means that household income likely provides a better indicator of welfare than aggregating individual incomes. However, if these assumptions do not hold, e.g. if household members do not have equal bargaining power to influence intra-household decisions, this complicates the interpretation of the estimates of income mobility by gender. In this case, the estimates would not be directly comparable because the pooled household income would probably translate into different well-being levels for sons and daughters.

In addition to marriage decisions, especially when examining mobility for daughters, it is important to keep in mind that fertility choices may also influence labour supply decisions and how women contribute to household income. As discussed previously, the choice of participating in the labour force and of choosing between full- or part-time work is closely associated with child bearing and caring, which occurs around the age at which income variables are usually observed. This means that the all these factors may influence my results for the intergenerational persistence of daughters and wives, particularly because income is observed at relatively young ages for the whole sample. Further research is needed to fully comprehend the life cycle bias for the estimation of income mobility for women. Improved estimates could be obtained by observing the income for the second generation i.e. sons and daughters at later ages and for longer periods and by adjusting the estimates for selection into employment and into part time work. This is a limitation of this study and accounting for these issues would provide more comparable estimates by gender. We still know very little about the income mobility of daughters and the implications of different patterns of labour force participation by gender to the study of income mobility and assortative mating. This investigation remains an interesting avenue for future research, as well as exploring the extent of assortative mating based on other characteristics, such as educational levels.

3.4.2 Income mobility at different points of the income distribution

Transition matrices and directional rank mobility

Transition matrices for parental and children's average income provide a good starting point to examine the strength of the intergenerational mobility income association at different points of the parental income distribution. By dividing the distributions of children's and parent's incomes(or earnings) into groups of equal size, usually quartiles or quintiles, one can calculate the probability that a child born at a given quantile will end up at the same or at a different quantile as an adult. In other words, quantile transition matrices show the proportion of children (as adults) who are in each quantile, relative to the quantile distribution of their parents. In one extreme case, if there is complete immobility across generations, all children would be in the lead diagonal of the matrix, and the parental income distribution would completely determine the distribution of children's outcomes. At the opposite extreme, if there is complete mobility, the parental income would be irrelevant in determining children's outcomes and all elements of the matrix would be equal to $1/n$, where n is the number of quantiles. These two extreme cases are rarely observed in practice, therefore quantile transition matrices can provide important information about non-linearities of the intergenerational income association and the direction of movements.

While transition matrices are useful to see how children move upwards or downwards in the income distribution, compared to their parents' quantiles, this approach is limited by its capacity to only observe movements between quantiles rather than within quantiles. Directional rank mobility measures aim to summarise and consolidate the information in a more systematic manner and provide an advantage over traditional transition matrices in that they allow the observation of mobility at a finer scale across the income distribution. Here, I use the estimators of directional rank mobility developed by Bhattacharya and Mazumder (2011) to examine differences in the direction of mobility by gender. Where transition probabilities depend on an arbitrary discretisation of the income distribution i.e. which quantile is chosen as the threshold, directional rank mobility measures do not suffer from the same limitation.[71]

Following Bhattacharya and Mazumder (2011), I measure upward rank mobility (UP) as the probability that a child will rank higher in the child income distribution than their parents' rank position in the parental income distribution. The probability that a child's rank (*Rc*) is higher than their parents' rank (*Rp*) by a fixed amount τ, conditional on the parent being at or below a particular percentile p in the parent's income distribution, is defined as:

$$UP_{\tau,p} = Pr(Rc - Rp > \tau | Rp \leq p) \tag{3.1}$$

where τ is a chosen threshold that represents the number of ranks that a child has to surpass the parental rank to be regarded as having experienced upward mobility. If $\tau = 0$, UP is just the probability that the child ranks higher than their parents in the respective income distribution. Values of $\tau > 0$ afford the opportunity to look at the magnitude of the gain in rank positions across generations. For example, $UP_{10,0-20}$ would represent the probability that the child ranks at least 10 positions higher than their parents, conditional on parent income being in the bottom quintile of the income rank distribution.

An analogous expression can be used to look at downward mobility (DOWN), representing the probability that a child will rank lower in the child income distribution that their parent's position in the parental income distribution. Downward rank mobility is defined as:

$$DOWN_{\tau,p} = Pr(Rp - Rc > \tau | Rp \geq p) \tag{3.2}$$

[71] Analysing transition probabilities based on a transition matrix dividing the distributions in quartiles, mobility would be observed if there is a movement from the 24th to the 26th percentile, for example. However, any movement within the first quartile, for example between the 1st and 24th percentiles would not be captured.

For example, $DOWN_{0,81-100}$ would represent the probability that a child ranks lower than their parents, conditional on parent income being in the top quintile of the income distribution.[72]

Results: Differences across the parental income distribution

Figure 3.2 presents a visual transition matrix for the UK. This provides a non-parametric description of income mobility at the national level, showing the joint income distribution of children and their parents and illustrating movements between quantiles. This figure depicts the distribution of household income for parents and children, ranked within each generation after adjusting income for age and time effects in the first stage of the TSRA estimation. Figure B.1 in Appendix B presents the national transition matrix using unadjusted household income for both generations, and the results remain qualitatively unchanged.[73]

Consistent with previous studies for the UK (Blanden et al., 2004; Dearden et al., 1997), I identify a high degree of persistence at the extremes of the income distribution. 29.8% of all children born from parents located in the first quintile of the parental income distribution (Q1) are themselves located in the first quintile of the child income distribution as adults. The transition probability[74] out of the bottom quintile is 70.2%, but only 9.4% of these children make it to the top quintile of the income distribution. At the upper end of the distribution, persistence is also very high. 35.5% of all children from the richest parents (Q5) are themselves located in the top quintile of the child income distribution. The transition probability out of the top quintile is 64.5%, and around 10% of children move downwards to the bottom quintile of the income distribution in adulthood. The results obtained using personal income

[72]The results on upward and downward rank mobility are presented in quintiles of the parental percentile rank distribution (i.e. $p = 0-20, 21-40, \ldots, 81-100$).

[73]It is unclear how estimates of directional rank mobility and transition probabilities are affected by measurement error (Corak et al., 2014). However, as long as ranks within each generation are preserved, errors in the measurement of income would not affect the results (Bhattacharya and Mazumder, 2011).

[74]The transition probability for each quintile equals 1 minus the probability of children staying in the same quintile as their parents.

for the younger generation reveal similar stickiness at the tails of the distribution (Figure B.2 in Appendix B).

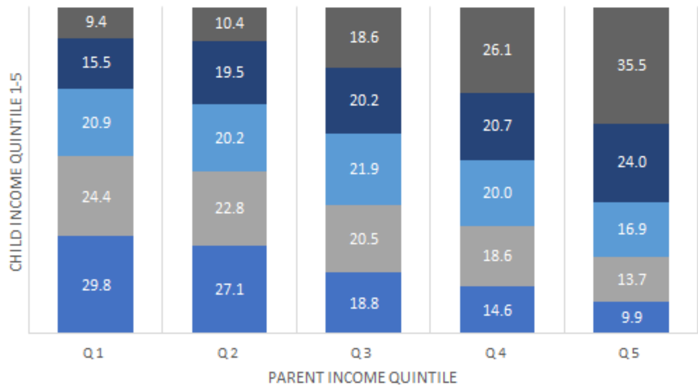
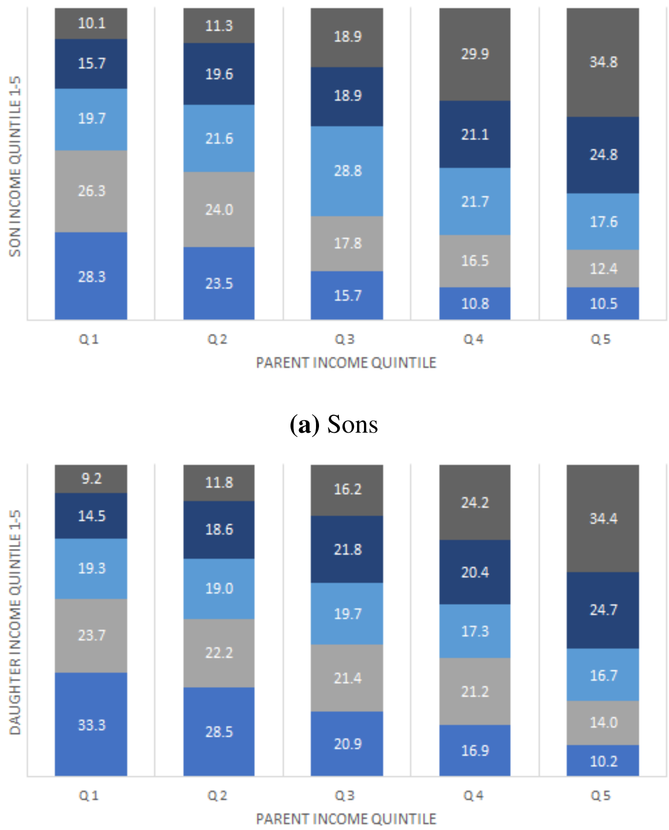


Figure 3.2 National transition matrix

Notes: This figure shows the distribution of children (in child income quintiles) for each parent income quantile. For example, the first column Q1 indicates that 29.8% of all children with parents in the bottom quintile of the parental income distribution (Q1) belong to the bottom quintile themselves in adulthood, while 24.4% move to the second quintile and 20.9% to the third. By definition, this matrix is bi-stochastic i.e. the sum of all values in each column equals 100% and all values of the same colour across columns also add up to 100% - any difference is due to rounding. Income is household income for both generations. Ranks are based on the national income distribution and constructed from the average adjusted income obtained after taking into account age and time effects in the first stage of TSRA. Total sample size is 2126. Source: Own elaboration, using BHPS and Understanding Society.

Figure 3.3 presents the national transition matrices separately by gender. Overall, I observe the same general pattern for sons and daughters and only some small gender differences, in line with the results for the UK from Dearden et al. (1997) and Blanden et al. (2004). The transition probability out of the bottom quintile is 71.7% for sons and slightly lower for daughters, 66.7%. On the other hand, the transition probability out of the top quintile is around 34% for both genders. Small gender differences are observed in the middle of the distribution, with higher persistence for sons in the third quintile.

These results from the national transition matrices reveal a considerable degree of mobility, which follows an asymmetric pattern across the parental income distribution. Income persistence is higher at the tails of this distribution for both household and personal income. This indicates that mobility is lower for the children of the relatively richest and poorest parents. This is consistent with what has been observed for other countries, such as the US (Chetty



(a) Sons

(b) Daughters

Figure 3.3 National transition matrix by gender

Notes: This figure shows the distribution of children (in child income quintiles) for each parent income quantile separately for (a) sons and (b) daughters. Income is household income for both generations. Ranks are based on the average adjusted income obtained after taking into account age and time effects in the first stage of TSRA. Sample size is 997 sons and 1129 daughters. Source: Own elaboration, using BHPS and Understanding Society.

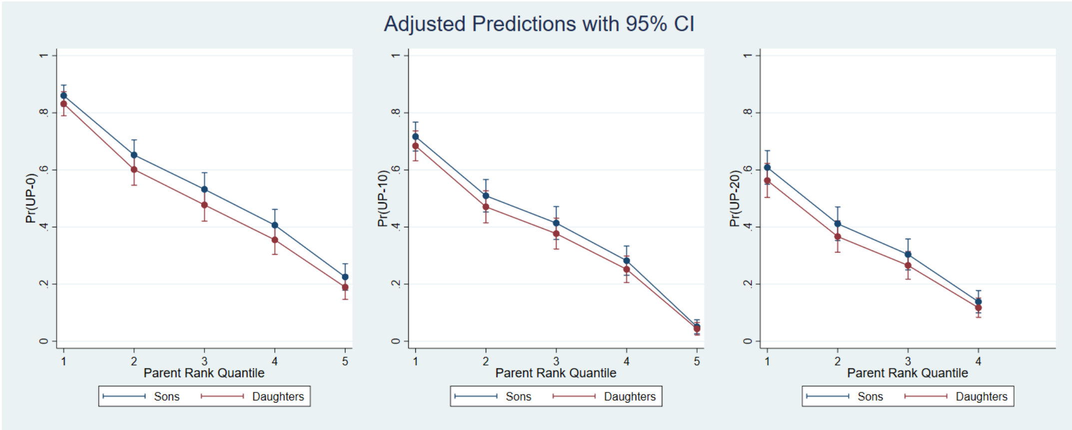
et al., 2014), European countries (Heidrich, 2017) and by previous studies for the UK for the NCDS and BCS cohorts (Blanden et al., 2004; Dearden et al., 1997). I find that around 30% of children with parents in the bottom quintile of the income distribution remain in that quintile, furthermore, this proportion is even higher at the top of the income distribution, where over 35% of children remain in the same quintile as their parents. Estimates of (im)mobility levels of children with parents in the bottom and top quintiles were strikingly similar to those reported for the US. Chetty et al. (2014), who examined family income for both generations, report that 33.7% of children with parents in the bottom quintile of the income distribution remain in that same quintile, and this proportion is even higher at the top of the distribution, where 36.5% of children remain in the same quintile as their parents.

Complementing this analysis, I also examine patterns of upwards and downward rank mobility by gender (Figure 3.4). Figure 3.4a illustrates the patterns of upward rank mobility based on household income. On the left hand side, the probabilities of children ranking higher than their parents are presented, whereas the middle and the right panels show the probabilities of children ranking at least 10 and 20 ranks higher than their parents, respectively. This analysis reveals a considerable degree of mobility within quintiles of the parental *rank* distribution, which are clouded when examining only larger scale movements between income quintiles based on the transition matrices. As expected, the general pattern of a monotonically decreasing probability as we move up the parental income distribution holds for both sons and daughters, as income gains needed to move ranks become larger. Rank mobility estimates do not account for the size of the gains, meaning that the movement from a child whose parents are on the 5th percentile is registered as upward mobility, regardless of whether they move to the 6th percentile or to the 90th percentile. Although children from the poorest rank quantiles may exhibit high levels of mobility, the actual income gains from this mobility might not be economically meaningful.

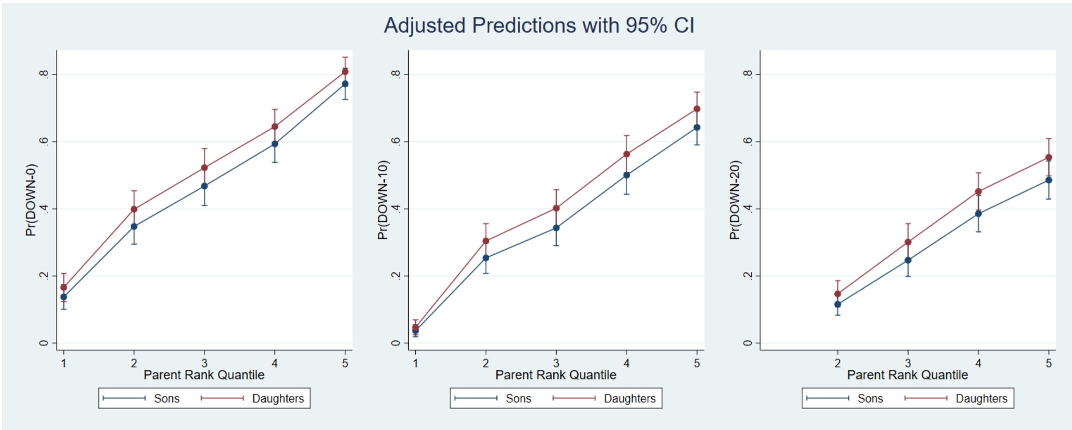
The most interesting finding here is that across all types of parents and values of τ examined, sons have a slightly higher probability of being upwardly mobile, i.e. more likely to achieve higher ranks than their parents than daughters do. Estimates for downward rank mobility by gender are illustrated in Figure 3.4b. Interestingly, daughters are more downwardly mobile than sons across the whole distribution of parental incomes. Although the estimated probabilities are indicative of lower upward mobility for daughters, due to small sample sizes, it is not possible to draw strong conclusions about these gender differences. An examination of rank mobility considering personal income for children (shown in Figure B.3 in Appendix B) instead of household income, adds to these findings, with larger gender differences observed.

The aggregate rank estimates of income mobility presented in the first part of this chapter have suggested that daughters and sons have similar levels of mobility. The results from an examination of differences in the degree of mobility across the income distribution bring new light to these findings and provide tentative evidence of differences in the type of mobility experienced by each gender. The transition matrices by gender reveal that even though overall daughters seem more mobile, the transition probabilities out of the bottom quintile are lower for daughters than for sons. This suggests that it might be more difficult for daughters of disadvantaged families to move upwardly. The directional rank mobility estimates reinforce these findings, while also bringing to light to fine-scale rank movements within income quintiles. Daughters are more likely to be downwardly mobile across all parental income quintiles, while sons are more upwardly mobile considering small but also larger movements. These results provide (moderate) evidence that the higher mobility for daughters might be driven by downward movements along the whole distribution of parent's incomes.

The transition matrices reveal a considerable degree of persistence, which follows an asymmetric pattern and is higher at the tails of the parental income distribution. The results for directional mobility suggest that a fine-scale mobility within quantiles is often missed when only larger scale re-ranking between income quintiles are considered. Further research is needed to better understand the extent of the importance of movements within quantiles and to confirm the tentative evidence on differences in the directional rank mobility by gender and investigate its reasons. Following the method proposed by Bhattacharya and Mazumder (2011) when analysing racial gaps in mobility in the US, one suggestion is to examine more thoroughly the potential underlying mechanisms for such differences by gender. For example, to investigate the role of human capital, one could examine whether there are gender gaps in the probabilities of moving upwards or downwards, conditional on education or qualifications.



(a) Upward rank mobility by gender



(b) Downward rank mobility by gender

Figure 3.4 Directional rank mobility by gender

This figure presents the predicted probabilities of being (a) upwardly and (b) downwardly mobile conditional on parents' income rank quintile, by gender. See text for a description of the estimator. The left panels show the probabilities when $\tau = 0$, the middle panels show the results for $\tau = 10$ and the right panels show the probabilities when $\tau = 20$), where τ represents various amounts of gain or loss in percentile ranks. Child income and parental income are household income. Ranks are based on the average adjusted income of each generation obtained after taking into account age and time effects in the first stage of TSRA. Source: Own elaboration, using BHPS and Understanding Society.

Quantile regressions

After examining how income mobility varies across the distribution of income of the first generation, I now focus on differences in mobility across the distribution of income of the second generation using quantile regressions (QR) techniques. The QR framework provides a way to examine the differential impact of explanatory variables along the distribution of an outcome. In the context of intergenerational mobility, this means allowing the coefficient of parental income i.e. point estimates of intergenerational persistence to vary at different quantiles q_τ of the distribution of children's incomes. This allows for the possibility that parental income plays a bigger role for children who are, for example, at the 10th percentile of the income distribution than for those who are at the 90th percentile.

The *conditional* quantile regression method, as proposed by Koenker and Bassett (1978), is the most common QR approach and aims to estimate the effect of a covariate on a quantile of the distribution of the dependent variable conditional on specific values of other covariates. Here, this means estimating the extent of the intergenerational association for quantiles of the conditional distribution of children's incomes. A limitation of this approach is that, unlike the OLS estimator, conditional QR cannot be used to represent the estimates at the unconditional quantiles of the distribution.[75] This is because the τ^{th} unconditional quantile of distribution y might not be the same as the τ^{th} conditional quantile of $y|x$: $q_\tau(y|x) \neq q_\tau(y)$. Thus, conditional QR may generate results that are not generalisable or easy to interpret when multiple independent variables are added to the model (Gregg et al., 2019).

To address this issue, recent studies of income and earnings mobility (for example Palomino et al. 2018 and Gregg et al. 2019) have employed *unconditional* quantile regressions, based on the method proposed by Firpo et al. (2009). This approach allows the estimation of the marginalised (unconditional) effects of a covariate at the unconditional quantiles of the

[75] The OLS estimator is valid for both conditional and unconditional distributions by the law of iterated expectations, so $\mathbb{E}(y) = \mathbb{E}_x[\mathbb{E}(y|x)]$.

distribution of the dependent variable.[76] In practice, Firpo et al. (2009) utilise the concepts of influence function (IF) and re-centred influence function (RIF) to recover the unconditional effects. Influence functions are an analytical tool used to obtain robust estimates and assess the influence of adding or removing an individual observation on a distributional statistic of interest (Borah and Basu, 2013; Firpo et al., 2009). The RIF is obtained by adding the distributional statistic of interest back to the influence function. The unconditional quantile regression method consists of running a standard least squares regression where the dependent variable is replaced by a RIF of the statistic of interest i.e. quantile (Firpo et al., 2009).[77] This transformation and the fact that the conditional expectation of the RIF is equal to the unconditional expectation of the RIF, $\mathbb{E}[RIF(y;q_\tau)] = \mathbb{E}_x[\mathbb{E}(RIF(y;q_\tau)|x)]$, make it possible to interpret of the estimates obtained by the RIF-regression as unconditional effects. In this chapter, I present the results using unconditional quantile regressions in the main text, and the results using conditional quantile regressions are reported in Appendix B.

Results: Differences across the children's income distribution

Figure 3.5 plots the estimated IGEs from unconditional QR for household income and personal income, displaying the income association at the 10th, 30th, 50th, 70th and 90th percentiles of the children's income distribution, rather than at the mean obtained by OLS or TSRA. In Figure 3.5a, the OLS IGE estimates using household income (continuous blue line) reveal slightly higher persistence at the very bottom of the children's distribution, and lower at the top of the distribution. For personal income (dotted orange line), however, income mobility is very similar at the tails of the distribution, and slightly lower than in the middle.

Figure 3.5b shows the unconditional quantile estimates of IGE based on household income and personal income adjusted for age and time effects using the TSRA method. In both figures, I observe a relatively flat line, indicating that intergenerational income persistence varies

[76] Note that if no additional covariates are added to the model, conditional and unconditional quantile regressions would yield the same results. However, this is not the case here because of the addition of controls for age and cohort effects to address the life cycle bias.

[77] More detailed information on the implementation of this method can be found in the original paper by Firpo and colleagues, Rios-Avila (2020) and Gregg et al. (2019).

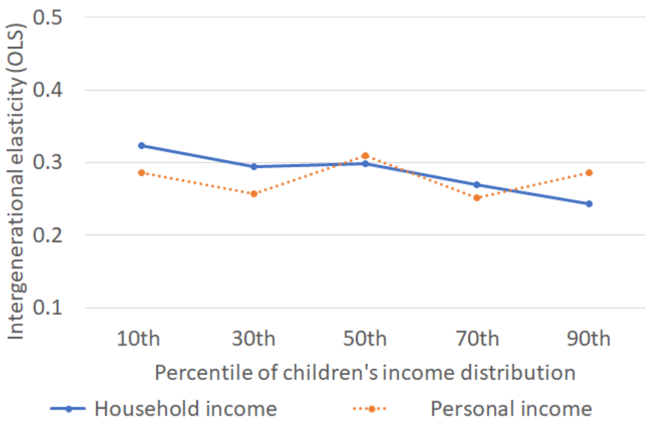
little across the children's income distribution. Table 3.4 presents complete results for IGEs estimated at different percentiles of the children's income distribution using unconditional QR. The first column presents the original OLS and TSRA estimates (i.e. at the mean) for comparison. These numbers reveal that the small differences in IGE estimates across the distribution of children's incomes - for example, a slightly higher degree of mobility observed at the very top of the distribution - are not statistically significant. This is the case for the two measures of income considered, household income and personal income. The results obtained using conditional quantile regressions, presented in Table B.2 in Appendix B, reveal a similar pattern.

My results here reveal only small non-significant variation in income mobility across quantiles of the children's income distribution. This contrasts slightly with the findings of the recent study by Gregg et al. (2019), who were the first to explore non-linearities along the distribution of income of the second-generation measures for the UK. The authors describe a 'J-shaped' curve between son's earnings and parental income for the BCS cohort. Specifically, Gregg et al. (2019) have found a much stronger association at the 90th percentile; they report an IGE of 0.445 for sons' earnings measured at age 30. Meanwhile, the estimates are smaller and much more similar at other points across the distribution. This would indicate that parental income is a strong predictor of earnings potential for sons at the bottom and middle, but even stronger the top of the earnings distribution. It is notable that, although this stronger persistence is observed at the top of the distribution, the relationship is otherwise linear and their estimates are remarkably close to those reported here for personal income, which are not significantly different from each other. In general, my results show a uniform influence of parental income across the children's income distribution and do not provide this evidence of higher persistence at the top.

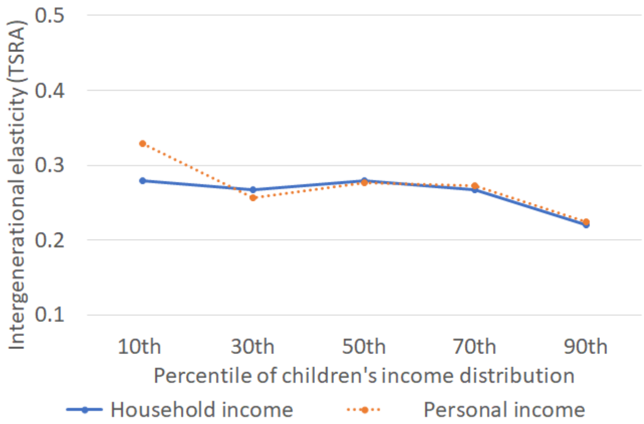
The contrasting results for the 90th percentile could be due to several reasons. One plausible explanation is that this difference could be related to the distinct income measures analysed in each study. Gregg et al. (2019) focus on son's earnings for the second generation. As discussed before, the estimates based on household income are not directly comparable to those based on individual measures of income because of the distinct nature of the association observed. Household income focuses on the transmission of living standards/welfare at the household level and is directly influenced by household composition. Even for the more comparable results, obtained using personal income, differences in the concepts of income could be driving the observed differences. This is because personal income incorporates other sources of income in addition to earnings at the individual level, such as returns to investment, transfers and benefits. This would imply that a weaker association is observed once these other sources of income are accounted for. A second possible explanation is that the observed differences relate to the inclusion of women in my sample. However, it would be interesting that this would only have an effect at that part of the distribution. Third, differences could appear due to differences in the methodology for approximating long-run income, although it is unlikely that this would only generate such a stark difference only at the very top of the distribution. Finally, it is not possible to reject the possibility that this difference arises due to genuine cohorts differences, or of changes in the institutional and policy environment.

All these are plausible explanations, and being only the second study to obtain estimates of income mobility across the income distribution of the second generation for the UK, disentangling these explanations is beyond the scope of this analysis. Recent studies for Germany and the US have also provided evidence suggestive of a higher persistence at the top of the distribution (Palomino et al., 2018; Schnitzlein, 2016), usually attributed to differences in the amount of education provided to children. More evidence is needed to fully comprehend these issues and remains a topic for future research. Specifically, efforts should focus on obtaining additional evidence using alternative measures of income and on expanding the analyses to include daughters.[78] One interesting methodological suggestion is provided by Palomino et al. (2018). Using the PSID, they suggest the pooling of the income data available to estimate income mobility for percentiles of the distribution (as opposed to only specific points or quantiles) and obtain accurate estimations at the tails.

[78]Here, it is important to keep in mind the same limitations that apply for the gender comparison. To obtain estimates based on individual measures of income, it is important to be able to account for the complexities of female labour force participation.



(a) OLS



(b) TSRA

Figure 3.5 Unconditional quantile regression estimates of IGE

These figures illustrate the estimates of intergenerational income elasticities by unconditional quantiles of the children's income distribution, based on the results displayed in Table 3.4. Source: Own elaboration, using BHPS and Understanding Society.

Table 3.4 Estimates of IGEs by quantiles

Unconditional QR		Percentiles				
Dependent variable	Mean	10th	30th	50th	70th	90th
Panel A: OLS						
Household income	0.269***	0.324***	0.294***	0.298***	0.270***	0.243***
	(0.028)	[0.059]	[0.035]	[0.026]	[0.026]	[0.035]
Equal with 10th percentile (p-value)	-	-	0.4499	0.5600	0.2419	0.1197
Equal with 90th percentile (p-value)	-	0.1197	0.2089	0.1200	0.3632	-
Personal income	0.293***	0.286***	0.257***	0.310***	0.253***	0.286***
	(0.031)	[0.050]	[0.034]	[0.028]	[0.024]	[0.031]
Equal with 10th percentile (p-value)	-	-	0.6061	0.7044	0.6127	0.9929
Equal with 90th percentile (p-value)	-	0.9929	0.4952	0.5115	0.2992	-
Panel B: TSRA						
Household income	0.255***	0.279***	0.268***	0.280***	0.267***	0.221***
	(0.027)	[0.050]	[0.034]	[0.028]	[0.024]	[0.031]
Equal with 10th percentile (p-value)	-	-	0.8009	0.9854	0.8029	0.2824
Equal with 90th percentile (p-value)	-	0.2824	0.2649	0.1418	0.1353	-
Personal income	0.270***	0.329***	0.257***	0.277***	0.273***	0.224***
	(0.028)	[0.083]	[0.036]	[0.026]	[0.028]	[0.029]
Equal with 10th percentile (p-value)	-	-	0.3193	0.5030	0.5267	0.2342
Equal with 90th percentile (p-value)	-	0.2342	0.4581	0.1399	0.1188	-

Notes: Robust clustered standard errors in parentheses. Bootstrapped standard errors in brackets (100 reps), * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Child income is either household income or personal income, as mentioned in the table, and parental income is household income. These income measures are described in the text in section 2.3.2. Total sample size is 2126 for household income and 2099 for personal income. OLS models controls for the average age of parents and children, their average age squared and birth year dummies. TSRA models control for a quartic in age of parents and children, as well as year dummies. The table also shows the p-value for the t-test of equality of coefficients with those estimated for the 10th and 90th percentiles. Source: BHPS and Understanding Society.

3.4.3 Regional variations in income mobility

Estimation

I use two different concepts to describe intergenerational income mobility at the regional level: relative and absolute mobility. *Relative mobility* is the focus of most intergenerational studies, and it compares the relative outcomes of children from different backgrounds. It is informative about the size of the gap between the average incomes of children from high- and low-income families in each region. In this sense, it is also a measure of income inequality by region. Here, I examine relative mobility by estimating intergenerational income elasticities and rank coefficients for each region. In practice, similar to the analysis by gender, this is easily implemented by estimating different equations for each region. The region-specific rank coefficients (γ_r) can be obtained by:

$$R_{ir} = \alpha_r + \gamma_r R_i^P + \varepsilon_{ir} \tag{3.3}$$

where R_{ir} denotes the rank in the national income distribution of children for a child i who grew up in region r, R_i^P denotes the parent's rank in the parental national income distribution and α_r denotes a constant term for each region.

In addition, I examine the degree of *absolute mobility* across regions. Absolute mobility represents the outcomes of children from parents at a given income or rank in the parental income distribution. In practice, I construct a measure of Absolute Upward Mobility (AUM), following the approach described by Chetty et al. (2014). The R25 reflects the mean rank (as adults) of children whose parents are at the 25th percentile of the (national) income distribution. The choice of comparing absolute mobility at the 25th percentile is motivated by the need to understand the prospects of upwards mobility for children from poorer families, and to facilitate comparisons with regional estimates from other studies.[79] When the rank

[79] Such as the work by Chetty et al. (2014) for the US, Heidrich (2017) for Sweden and Acciari et al. (2019) for Italy.

relationship is linear, the average rank of children with below-median parental income equals the average rank of children at the 25th percentile (p=25) of the national income distribution ($\mathbb{E}[R_i|R_i^P = 25]$). A visual representation of the association between children's and parent's percentile ranks for the UK is presented in Figure B.5 in the Appendix, and in the next subsection for England's regions (Figure 3.7). As long as linearity holds, the measure of absolute upward mobility for each region can be calculated as:

$$R_r^{25} = \alpha_r + 25.\gamma_r \tag{3.4}$$

where α_r is the constant from the rank model estimated separately for each region r (see above equation 3.3), which measures the expected rank of a child who grew up in region r and was born from parents at the bottom of the national income distribution, and the slope γ_r is the rank coefficient estimated for each region.

Rank-based measures are particularly useful for making 'apples-to-apples' comparisons of relative mobility among subgroups of the population when based on the ranks of the national income distribution (Deutscher and Mazumder, 2019; Mazumder, 2016). Following Chetty et al. (2014), I use this property of rank coefficients to compare mobility across regions in the UK and always assign the rank to parents and children based on their respective national income distribution (rather than the distribution within each region). If regional ranks were used instead, what would it mean that children from disadvantaged families in Yorkshire reach on average the 35th percentile (within Yorkshire) while children from disadvantaged families in the South West reach on average the 40th percentile rank (within the South West). By using a national scale, regions across the country can be directly compared to one another, since the rank mobility estimates measure mobility with respect to the national distribution, thus having the same meaning in all places.[80] Therefore, using the 25th percentile of the parental national income distribution avoids the problem that in poor regions the expected

[80] An intergenerational elasticity estimated within regions, on the other hand, would be informative of mobility with respect to the group specific mean, not the national mean (Acciari et al., 2019).

rank of children from parents at the 25th percentile could be much lower than in rich regions (Acciari et al., 2019).

To summarise, the conceptual differences between relative and absolute mobility are illustrated in Figure 3.6. On the left panel, both relative mobility and absolute mobility at the 25th percentile of parental income distribution are represented. The slope of the line represents relative mobility, in this case, the rank coefficient, representing the association between child and parent income ranks. Meanwhile, the dotted line represents the mean rank of children whose parents are on the 25th percentile ($p = 25$), depicting the degree of absolute upward mobility at this point of the income distribution. On the right panel, using three regions as an example, this difference between the concepts of relative and absolute mobility are further emphasised. Region 1 and Region 3 have a similar level of relative mobility, as the slopes of these lines are the same. However, these regions have different levels of absolute mobility, with children from poorer families in Region 1 achieving lower ranks than those from Region 3. Region 1 and Region 2 have the same level of absolute mobility at $p = 25$, however Region 2 has lower relative mobility (or higher degree of persistence), as shown by the steeper rank slope. A steeper rank slope is indicative of a higher variance of children's ranks in this region i.e. there is a larger distance between children from top and bottom ranked parents.

Regional estimates

Table 3.5 contains the main estimates of relative and absolute intergenerational income mobility for England, and splitting the sample in those children who grew up in the North or in the South of the country.[81] Although this division between the North and the South of England is not official, it is commonly referred to, for example in the media, and is a very useful way to describe geographical disparities in Britain. For the purposes of this study, the South is comprised of the South East and South West, Greater London and the East of

[81] Limited data for Scotland, Wales and Northern Ireland constricted the sub-national regional analysis to England only.

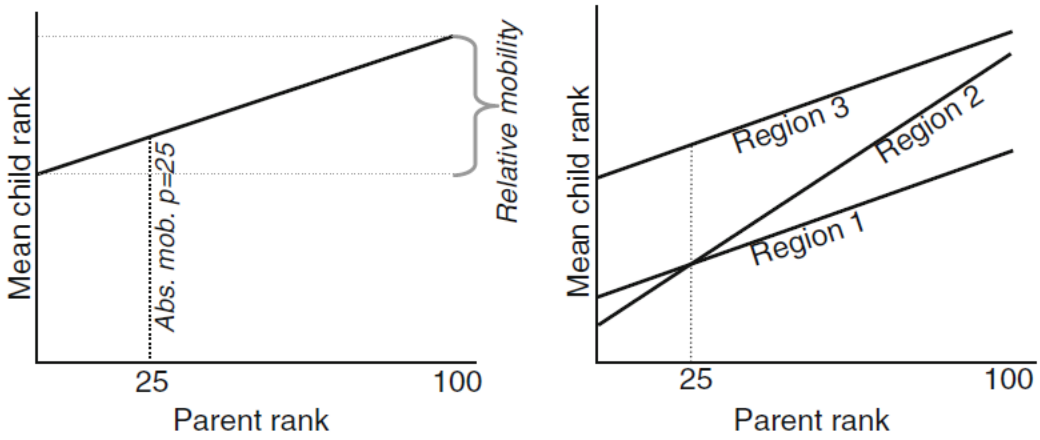


Figure 3.6 Relative and absolute mobility

Notes: This diagram illustrates the differences between the concepts of relative and absolute mobility. Source: Figure taken from Heidrich (2017).

England. The North includes the North East, Yorkshire and the Humber, the North West and the Midlands. An auxiliary map that accompanies this text is provided in Figure B.4 in the Appendix.

Examining first the rank coefficients and intergenerational elasticities, which measure relative mobility, it is notable that the estimates for England are similar to those obtained previously for the whole of the UK. Interestingly, however, the IGE estimates for the South are strikingly lower than those for the North. This indicates that children who grow up in the South of England experience higher levels of relative income mobility. The IGE estimate is 0.360 for the North and 0.134 for the South - almost three times smaller for the South than for the North. This means that a 10% differential in parental income translates into a 3.6% differential in child adult income if the child grew up in the North and 1.34% if the child grew up in the South of England. Rank coefficients for the South are also smaller than for the North, around half the size, indicating lower rank persistence. These differences between the North and the South of England are highly statistically significant.

The association between the ranks of parents and children is illustrated in Figure 3.7, which presents binned scatter plots of child income rank on parent's income rank. Each point

Table 3.5 Absolute and relative mobility in the North and South of England

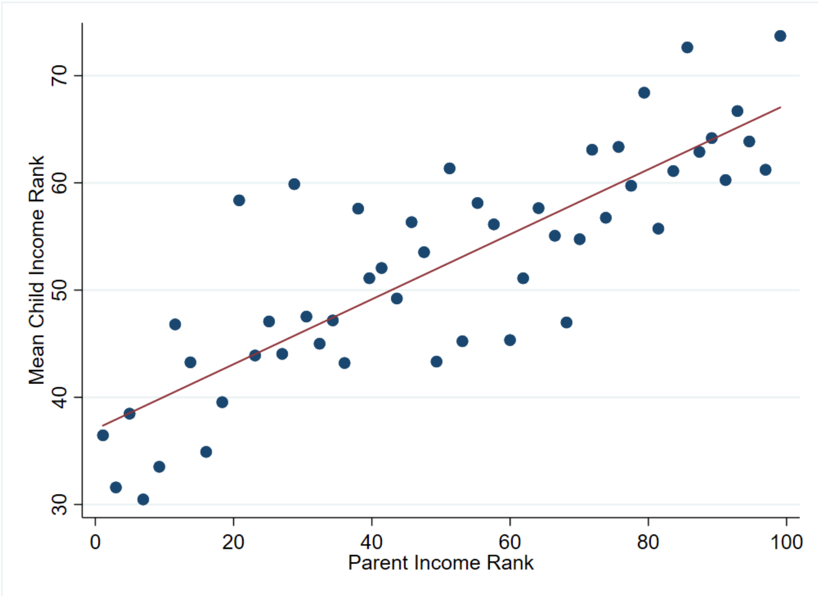
		Region		
Mobility measure	England	North	South	South (-) London
Panel A: Relative Mobility				
IGE (OLS)	0.275***	0.351***	0.173***	0.182***
	(0.031)	(0.046)	(0.039)	(0.040)
IGE (TSRA)	0.255***	0.360***	0.134***	0.132***
	(0.032)	(0.046)	(0.039)	(0.040)
IGE North = IGE South (p-value)	0.003			
Rank coefficient (OLS)	0.278***	0.333***	0.203***	0.219***
	(0.031)	(0.046)	(0.042)	(0.044)
Rank coefficient (TSRA)	0.272***	0.329***	0.195***	0.203***
	(0.030)	(0.042)	(0.041)	(0.044)
Rank North = Rank South (p-value)	0.022			
Panel B: Absolute Upward Mobility				
R25	44.17	38.16	51.08	49.53
N	1379	695	684	552

Notes: Robust clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Children's and parent's ranks were based on the national household income distribution in each generation. These income measures are described in the text in section 2.3.2. OLS models controls for the average age of parents and children, their average age squared and birth year dummies. TSRA models control for a quartic in age of parents and children, as well as year dummies. The measures of relative and absolute mobility are explained in the text in section 3.4.3. The South of England comprises the South East and South West, Greater London and the East of England. The North includes the North East, Yorkshire and the Humber, the North West and the Midlands. South(-)London includes South East, South West and East of England. The p-values for the t-test of equality of IGE and Rank coefficients for the North and South estimated using TSRA are presented in the table. Source: BHPS and Understanding Society.

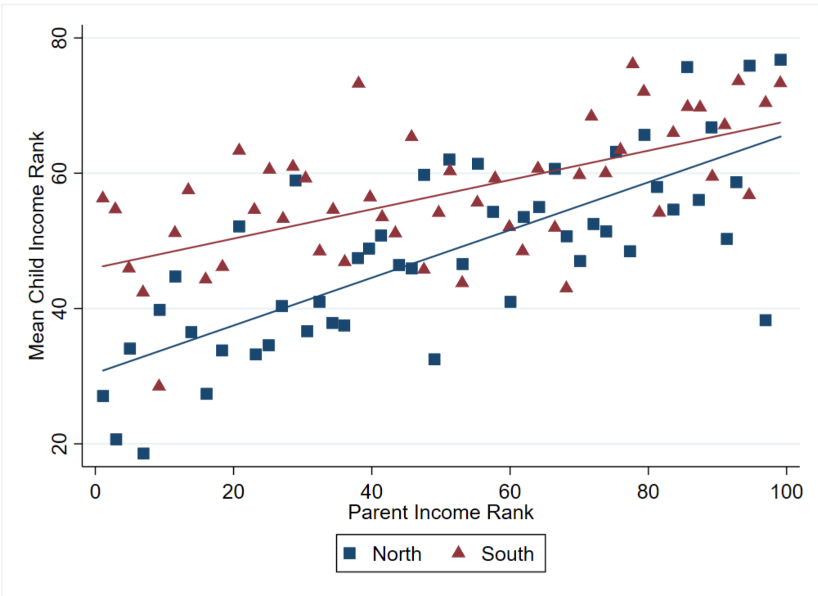
represents the expected income rank of children and parents within one of the 50 quantiles of the parental income rank distribution. A strong positive relationship is observed for England in Figure 3.7a (and for the UK as well, as shown in Figure B.5 in the Appendix). The rank coefficient for England (by OLS) indicates that a 10 percentile rank increase in parent income is associated with a 2.78 higher rank for children, as reported in Table 3.5. Furthermore, Figure 3.7b compares the rank-rank relationship in the North and in the South of England. The slopes of these lines reflect the lower levels of relative income mobility in the North (steeper slope) and higher levels in the South (flatter slope). These figures provide a graphical evidence that justifies the adoption of a linear specification in the regression models used to estimate the rank coefficients.

It is important to emphasise, however, that these differences observed in relative mobility do not necessarily mean that children from disadvantaged parents who grow up in the South do better than those who grow up in the North in absolute terms. In order to understand whether this is the case, we must turn to the measure of absolute mobility. The R25, which represents the mean rank of children whose parents are at the 25th percentile of the income distribution, reveals interesting differences between the North and the South of the country. The estimates presented in Table 3.5 also indicate a lower degree of absolute upward mobility for individuals who grow up in the North. For England as a whole, the R25 equals 44.17, which means that a child of parents with income below the median (on average, from the 25th percentile) is expected to end up in the 44th percentile of their income distribution. That is, this child is expected to move upwards but fall short of reaching the median income. Examining north/south differences, a child from parents with income below the median who is raised in the North of England is expected to reach the 38th percentile, whereas a child raised in the South of England could be expected to reach the 51st percentile - i.e. marginally above the median of the distribution.

Other research has identified a divide between London and the rest of the country (Friedman and Macmillan, 2017; Social Mobility Commission, 2017*a*). Looking at the results from the Social Mobility Index in 2017, London accounts for more than two thirds of the social mobility hotspots in the country (Social Mobility Commission, 2017*a*). The last column of Table 3.5 shows results for the South excluding London, in order to examine the so called 'London effect'. The results reveal that excluding London only slightly increases the coefficient for the South, and the South of England is still substantially more mobile than the North. This suggests that London is not driving my results on the differences between the North and the South of the country. Unfortunately, the smaller sample size when looking at London separately does not allow us to draw strong conclusions about its differences from the rest of the South with precision.



(a) England



(b) North and South of England

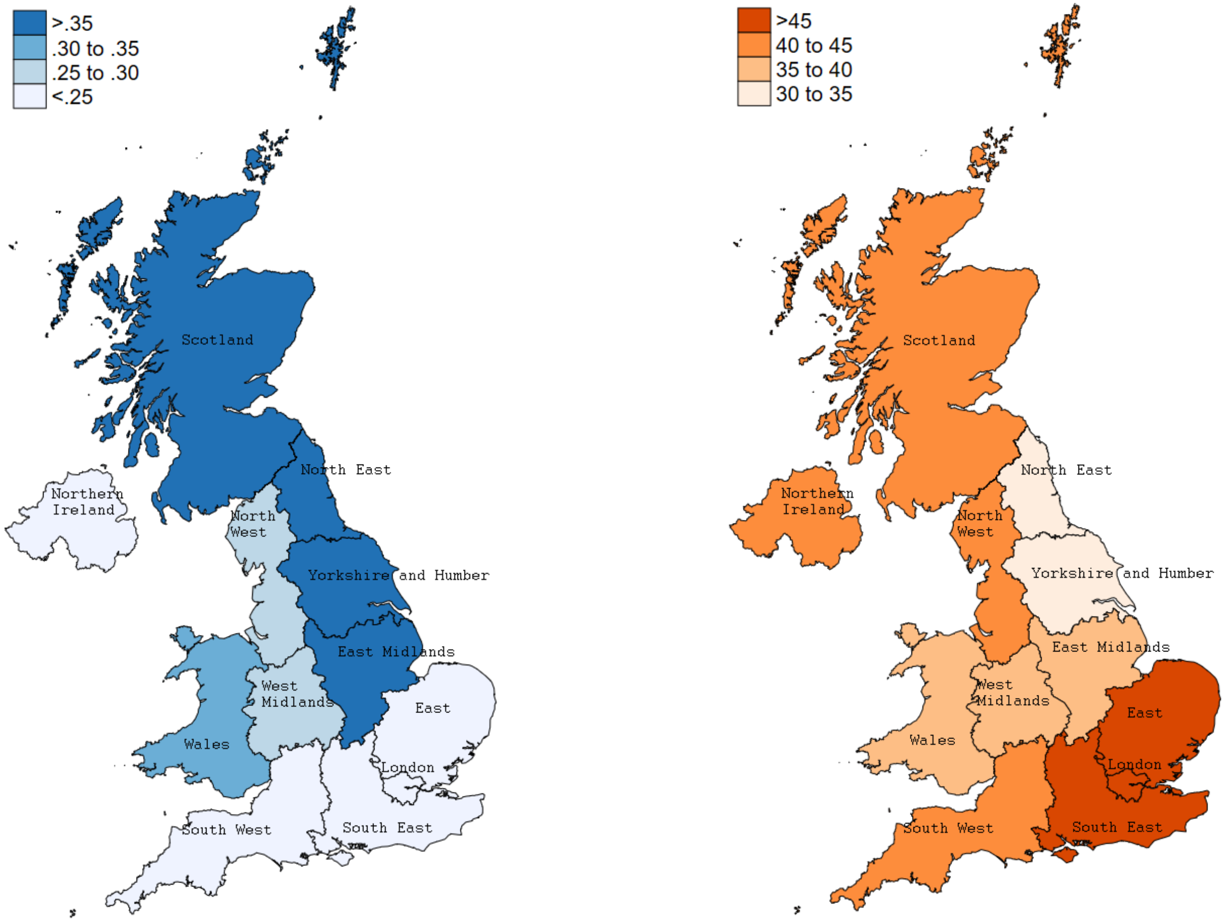
Figure 3.7 Association between children's and parent's percentile ranks in England

Notes: This figure presents non-parametric binned scatter plots of the relationship between children's and parent's percentile ranks. Panel (a) presents the series in circles using data for England, and Panel (b) presents the series separately by regions in the North (in squares) and South (in triangles) of England. This plot is based on the main matched sample and illustrates the mean child percentile rank within each in 50 quantiles of parent's income, considering the household income distributions in both generations. No controls are included. Source: Own elaboration, using data from BHPS and Understanding Society.

Next, I examine intergenerational income mobility at a more disaggregated regional level. I document mobility at the regional level using the same definitions of parental and child income as before, with the focus on household income. The complete estimates of relative and absolute mobility by region are presented in Table B.3 in Appendix B. Figure 3.8a provides a visual representation of the regional variations in relative income mobility measured by the rank coefficient across the 9 regions of England, plus Wales, Scotland and Northern Ireland.[82] In this chloropleth map, darker blues correspond to regions with higher intergenerational rank coefficients i.e. less mobile areas. Two broad patterns emerge from this figure. First, it suggests that there is considerable regional variation in relative mobility. The rank coefficient ranges from 0.164 in the South East of England, the most mobile region according to this measure, to 0.446 in Yorkshire and Humber, the region with the least relative income mobility. Second, it confirms the clear North-South gradient that I have observed before. In addition, when looking at the North of England, an additional interesting East-West divide emerges, with regions in the North-East having the lowest relative mobility.

Now turning to an examination of absolute mobility, Figure 3.8b summarises the geographical variation in absolute upward mobility, measured by the mean rank of children with parents at the 25th percentile of the distribution. Here, darker oranges represent areas with high absolute upward mobility, that is, more mobile areas. Interestingly, as with relative mobility, upward mobility also has a North/South gradient and the regions with highest mobility are in the South of England. The top region for absolute upward mobility (highest R25) is London, with 56.46. The region with lowest R25 is Yorkshire and the Humber, with 31.79. It is not possible to draw strong conclusions about the statistical differences between these regional estimates because sample sizes become quite small when looking at the separate regions, generating imprecise estimates. However, these results hint at differences that qualitatively would be very meaningful and translate to significant gaps in children's incomes as adults. More data is needed to fully explore these variations.

[82] Similar patterns are observed when relative mobility is measured by the intergenerational income elasticity instead. The estimates of IGE for each region are presented in Table B.3 in Appendix B.



(a) Relative mobility (rank coefficient)

(b) Absolute upward mobility (R25)

Figure 3.8 Regional variation in intergenerational income mobility

Notes: These figures illustrate the geography of intergenerational income mobility in the UK. Figure (a) presents the results for relative mobility considering the estimated rank coefficients (by TSRA). Darker blues represent areas with high rank coefficients, that is, with low intergenerational relative income mobility. Figure (b) presents the results for absolute mobility, showing the estimates of Absolute Upward Mobility as measured by the R25. Darker oranges represent areas with high R25, which have high absolute mobility. The underlying estimates for each region are presented in Table B.3 in Appendix B. Source: Own elaboration, using data from BHPS and Understanding Society.

The principal finding here is that stark differences in absolute and relative mobility exist across regions in the UK. Consistent with recent evidence for the US and other European countries (Acciari et al., 2019; Chetty et al., 2014; Güell et al., 2018; Heidrich, 2017), this result substantiates the notion that a UK 'postcode lottery' does exist, whereby the importance of family background for future economic outcomes also depends on location. One of the most striking results is the presence of a clear north-south divide in England. The strongest association (lower mobility) between child and parent income ranks is observed in the North of England, where the relative income difference between children from low- and high-income families is around 33 percentile ranks. In the South, on the other hand, this difference is about 20 percentile ranks. The estimated rank coefficients are almost half the size for the South, while the estimated income elasticities are even smaller, by a factor of three. These results indicate that family background is a much stronger predictor of success (or failure) in adulthood for individuals who grow up in the North of England. Moreover, regions in the North of England provide fewer opportunities of upward mobility for disadvantaged children; absolute upward mobility is lower in the North than in the South of England. The mean income rank of a child whose parent is at the 25th percentile income distribution is 38.2 in the North of England and 51.1 in the South of England.

As explained in the estimation section, at the national level, the rank coefficient and the absolute upward mobility measure are mechanically related, but this is not necessarily the case within each region At the regional level, a large R25 could be consequence of a high intercept for the region or of a steep rank slope (see equation 3.4). In my data, I observe that much of the variation in absolute upward mobility between the North and the South (and also across smaller regions) comes from the constant terms, an indication that there is a strong region effect in addition to the role of family background. The 'North' effect is moving every child raised in the region down in the national income distribution, independently of their parental rank.

Considerable geographical variation is also observed when examining the estimates of relative and absolute mobility across smaller regions at a higher level of regional disaggregation.

My findings shine light on the existence of this spatial variation, which was previously unaccounted for when focusing on national aggregate estimates of IGE and rank coefficients. Some regions in the UK have relative mobility comparable to the highest mobility countries, such as Denmark and Norway, while others have substantially lower levels of mobility. Regarding relative mobility, the South East and London stand out as regions with higher rank income mobility, with a relative income difference of around 16 and 20 percentile ranks, respectively. In these regions, being born to relatively disadvantaged parents does not prevent children from having higher income in adulthood to the same extent as in other regions. Conversely, particularly low relative mobility is observed in Yorkshire and Humber, where the relative outcome difference is more than 44 percentile ranks. Consistent with the evidence for Italy and the US (Acciari et al., 2019; Chetty et al., 2014), I observe that regions with high relative mobility tend to exhibit high absolute upward mobility. For example, in London and the South East, children with parents at the 25th percentile of the distribution are expected to end up near the median, with mean percentile ranks of 56.5 and 53.7, respectively. In Yorkshire and Humber, on the other hand, these children are expected to reach only around the 32nd percentile rank. These rank differences translate to large income differences.

These findings resonate with analyses of differences in the availability of opportunities across regions conducted by the Social Mobility Commission using data from the Social Mobility Index (Social Mobility Commission, 2016*b*, 2017*a*). My results complement these analyses and provide further evidence that more opportunities for social mobility exist in the South. My findings are also in line with those by Bell et al. (2018) - regions with more relative income mobility tend to correspond to those that exhibit more educational mobility.

Overall, this study supports the notion that not only does a UK postcode lottery exist, but that regional differences particularly impact the outcomes of disadvantaged children and the potential for upward mobility. In the South of the England, children from advantaged and disadvantaged backgrounds are less different in their economic outcomes as adults, and children from more disadvantaged parents have better opportunities for upward mobility than those who grow up in the North. Although income mobile and immobile regions differ considerably in regard to several demographic and economic characteristics, the exact reasons for these regional variations in income mobility across the UK are not yet known. Previous research for the US and Italy suggests that specific regional characteristics are more associated to a higher degree of intergenerational mobility. In these countries, it has been shown that high relative and absolute mobility areas tend to display less residential racial and financial segregation, less income inequality, less family instability, better primary schools and more social capital (Acciari et al., 2019; Chetty et al., 2014; Connolly et al., 2019; Güell et al., 2018). Despite my key findings on regional variation of income mobility, one shortcoming is that this study is not able to disentangle the reasons for the observed UK regional variation. Explaining this heterogeneity is one of the main challenges of this literature. This is an important avenue for future research which would contribute to our understanding of the local socio-economic conditions associated with better opportunities for social mobility. In addition to identifying key correlates of mobility that explain cross-regional variation, future work could use regional estimates to obtain insights into the causal drivers of intergenerational mobility. A recent example of how this could be done is presented by Derenoncourt (2019) using US data - this is an important avenue of research to understand how social mobility can be improved through policy action.

3.5 Conclusion

In this chapter, I have used the BHPS and Understanding Society data to draw a picture of intergenerational income mobility in the UK through unpacking mobility across three key dimensions. Specifically, I examine the heterogeneity of income mobility by gender, across the income distribution and between the regions in the UK. In line with the previous literature and my first empirical chapter, I have focused on measures of household income to capture the association of living standards across generations and to also consider mothers and daughters in my analyses. These measures were constructed by averaging income observed over multiple years to obtain average permanent income, while also accounting for life-cycle and time effects. Adjusted incomes and ranks are highly comparable for both generations since I account for the ages and time when income is observed.

The UK is considered to be a country with relatively low levels of intergenerational mobility. In this analysis I have demonstrated that even when focusing on household income which is a broader indicator of socio-economic status, this categorisation might be over-simplistic and hide subtle but important differences. My results reveal substantial differences across regions and by parent's position in the income distribution. This demonstrates the importance of interpreting any cross-country comparisons with great caution. Whenever possible, summary measures of relative mobility should be examined in greater detail in order to observe nuances in the direction of the movements and possible differences across the distribution. When the main policy interest is in promoting upward mobility for disadvantaged individuals, these measures should also be complemented by measures of absolute mobility to understand differences in outcome levels that would otherwise go unnoticed. Moreover, my results from the examination of gender differences highlight the importance of carefully considering the income measure being used in order to understand the nature of the intergenerational association.

Many of the heterogeneities unpacked in this chapter identify groups for which family background is strongly associated with socio-economic status in adulthood. My research points to variations across the parent's income distribution; notably a strong intergenerational persistence for children from the richest and poorest families. Moreover, it shows that when considering mobility for women, it is important to account for the pooling of resources at the household level and for differences in mobility at the individual level that may be attributed to labour market participation patterns. I also providence evidence that a postcode lottery exists, and that opportunities vary substantially across the regions of the UK.

National measures of intergenerational income mobility potentially say little about the state of mobility for specific subgroups of the population and for different regions in the country. The main findings from this chapter demonstrate the importance of looking beyond aggregate measures of income mobility. The observed variations across subgroups identify groups of people and regions that experience less income mobility. However, we still know very little about why some of these differences exist and further research is needed to fully comprehend some of the patterns observed. For example, in the gender discussion, it is of crucial importance to obtain comparable estimates of mobility based on individual measures of income (e.g. personal income, earnings) for sons and daughters. The estimates presented in this chapter do not adjust for differential patterns of labour force participation and hours worked, and this could be an important explanation for the small differences observed. In addition, especially for women, more research is needed to understand the potential impact of life cycle bias. The age issue might be even more problematic for obtaining proxies of long-run income for daughters because heterogeneity in age income profiles might be related to decisions of having children/ childcare and to labour supply decisions. More research is also needed to ascertain whether there are differences in income mobility across the income distribution. The current findings for the UK are incredibly scarce and inconclusive and more evidence is needed to improve the targeting of policies aiming to promote mobility. Finally, it is important to continue the investigation into regional variations in mobility to obtain additional evidence of these patterns at a more local scale and to be able to make the link with regional characteristics that may be favourable to more social mobility.

Appendix B

B.1 Supplementary figures

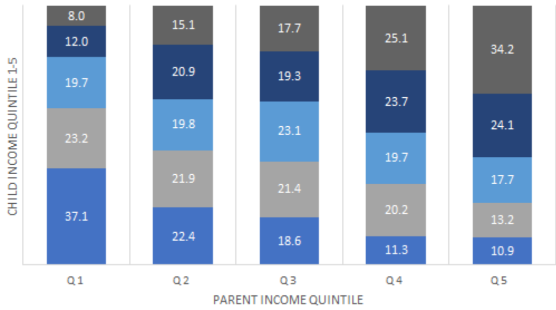


Figure B.1 National transition matrix - unadjusted household income

This figure shows the distribution of children (in child income quintiles) for each parent income quantile. By definition, the sum of all blocks in each column equals 100% and blocks of the same colour across columns also add up to 100% - any difference is due to rounding. Income is household income for both generations. Ranks are based on the multi-year average income without controlling for age and time when income is observed. Total sample size is 2126. Source: Own elaboration, based on data from BHPS and Understanding Society.

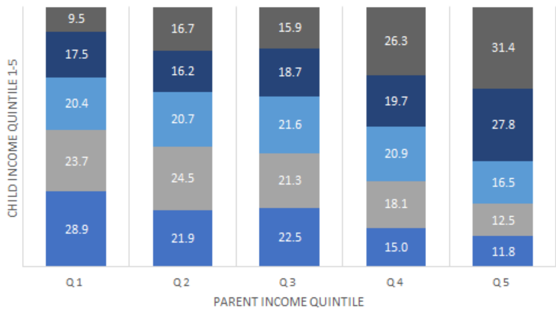
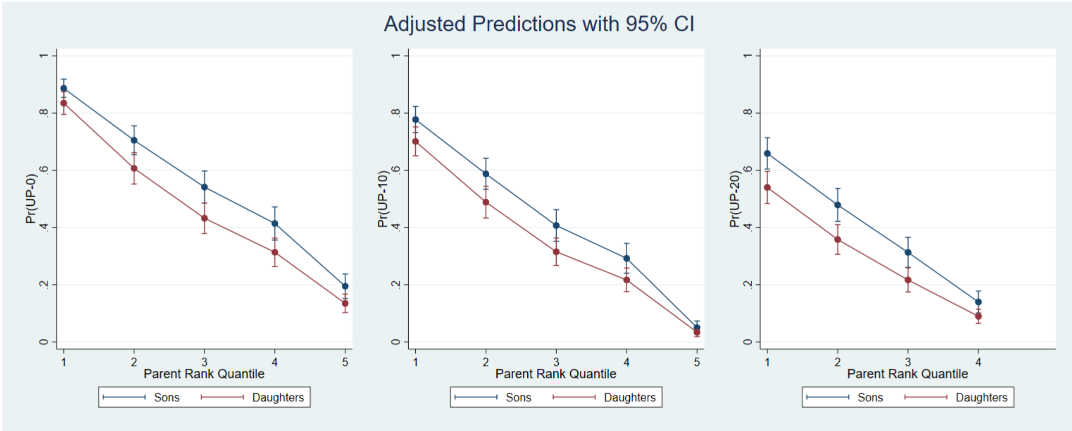
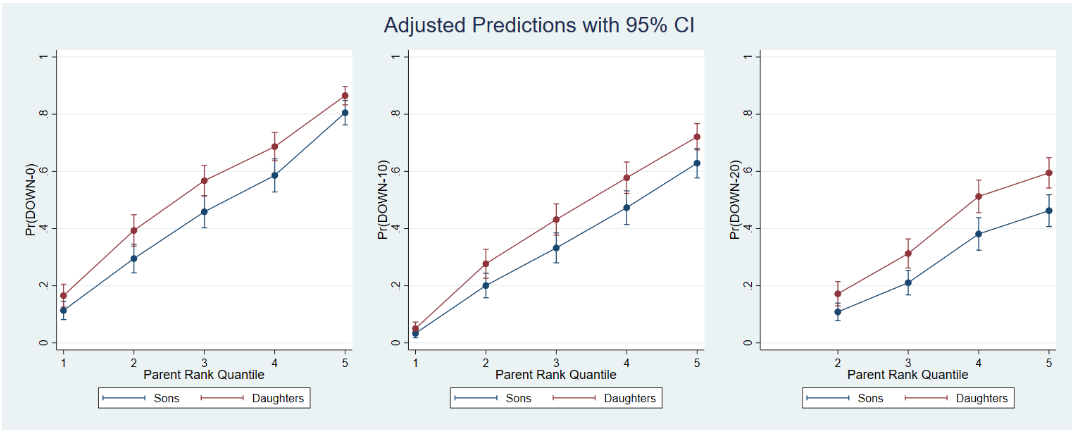


Figure B.2 National transition matrix (personal income)

This figure shows the distribution of children (in child income quintiles) for each parent income quantile. By definition, the sum of all blocks in each column equals 100% and blocks of the same colour across columns also add up to 100% - any difference is due to rounding. Income is household income for the parents' generation and personal income for the children. Ranks are based on the national income distribution and constructed from the average adjusted income obtained after taking into account age and time effects in the first stage of TSRA. Total sample size is 2114. Source: Own elaboration, based on data from BHPS and Understanding Society.



(a) Upward rank mobility by gender



(b) Downward rank mobility by gender

Figure B.3 Directional rank mobility by gender (personal income)

This figure presents the predicted probabilities of being (a) upwardly and (b) downwardly mobile conditional on parents' income rank quintile, by gender. The left panels show the probabilities when $\tau = 0$, the middle panels show the results for $\tau = 10$ and the right panels show the probabilities when $\tau = 20$), where τ represents various amounts of gain or loss in percentile ranks. Child income is personal income and parental income is household income. Ranks are based on the average adjusted income of each generation obtained after taking into account age and time effects in the first stage of TSRA. Source: Own elaboration, using BHPS and Understanding Society.

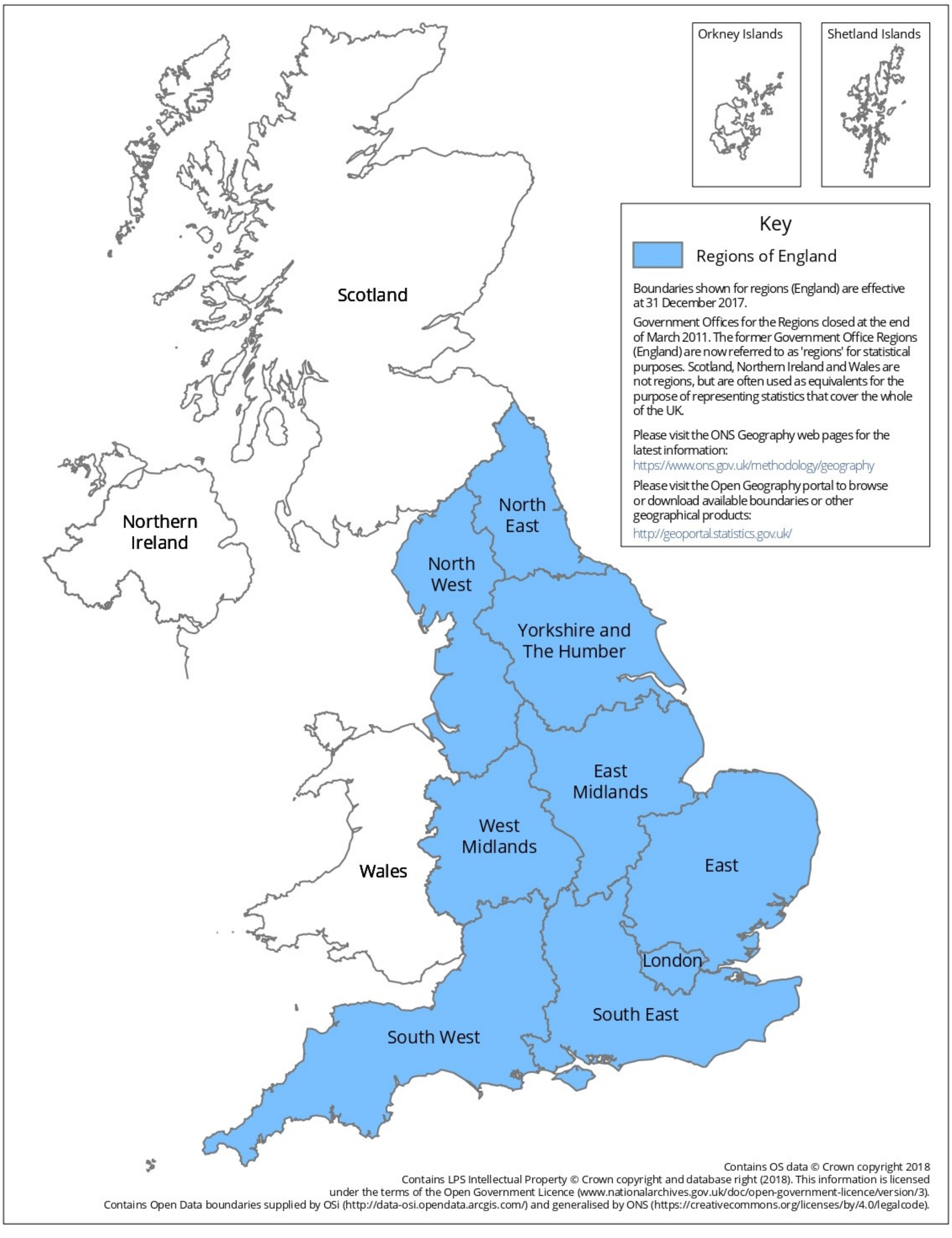


Figure B.4 Map of the UK and regions of England

Source: ONS Geography Open Data

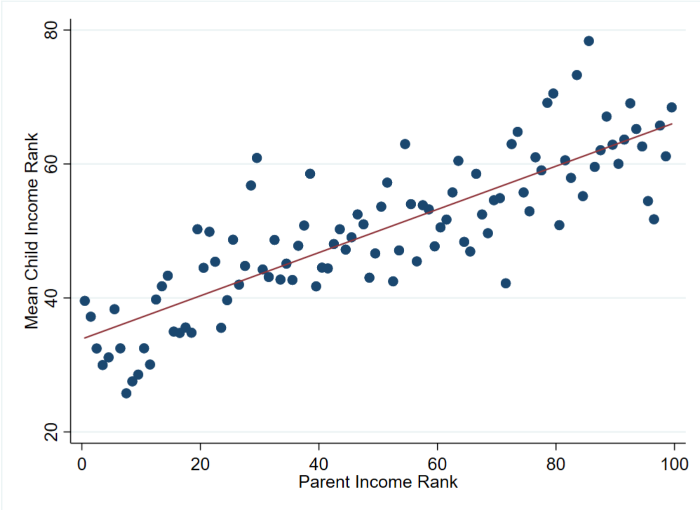


Figure B.5 Association between children's and parent's percentile ranks in the UK

This figure presents the non-parametric binned scatter plot of the relationship between children's and parent's percentile ranks. This plot is based on the UK matched sample and illustrates the mean child percentile rank within each parent percentile rank bin, considering the household income distributions in both generations. No controls are included. Source: Author's calculations based on data from BHPS and Understanding Society.

B.2 Supplementary results

Table B.1 Summary statistics by gender

Panel A: Sons (N=997)	Obs	Mean	SD	Min	Max
Sons' Parents					
Parental household income	7143	3660.00	(2699.72)	15.29	82309.13
Parental age when income observed	7143	42.29	(6.97)	25	65
Sons					
Adult household income	6390	4290.31	(2399.32)	1.74	24941.51
Adult personal income	6351	2262.93	(1562.38)	0	18337.41
Adult labour earnings	6350	2129.53	(1569.14)	0	18337.41
Age when household income observed	6390	30.15	(4.45)	25	44
Child age when parental income observed	7143	13.03	(4.14)	0	18

Panel B: Daughters (N=1129)	Obs	Mean	SD	Min	Max
Daughters' Parents					
Parental household income	8488	3580.14	(2732.01)	66.21	102426.22
Parental age when income observed	8488	41.80	(6.99)	25	65
Daughters					
Adult household income	7089	4017.87	(2445.06)	4.34	41794.38
Adult personal income	7069	1712.04	(1135.30)	0	14583.00
Adult labour earnings	7062	1376.60	(1206.27)	0	9021.56
Age when household income observed	7089	30.05	(4.42)	25	44
Child age when parental income observed	8488	12.91	(4.15)	0	18

Notes: Unweighted summary statistics based on 6390 observations of sons in adulthood, and 7143 observations of sons' parents, corresponding to 997 pairs of parents-sons in total. For daughters, this is based on 7089 observations in adulthood, and 8488 observations on daughters' parents, corresponding to 1129 pairs of parents-daughters. All income variables are pre-tax income in (£/month) and constructed as explained in the text in section 2.3.2. Source: BHPS and Understanding Society.

Table B.2 Estimates of IGEs by quantiles - conditional quantile regressions

Conditional QR				Percentiles		
Dependent variable	Mean (OLS)	10th	30th	50th	70th	90th
Household income	0.269***	0.285***	0.275***	0.303***	0.273***	0.271***
	(0.028)	[0.049]	[0.028]	[0.025]	[0.026]	[0.037]
Equal with 10th percentile (p-value)	-	-	0.8305	0.7029	0.8153	0.8227
Equal with 90th percentile (p-value)	-	0.8227	0.9218	0.4216	0.9385	-
Personal income	0.293***	0.304***	0.335***	0.271***	0.262***	0.253***
	(0.031)	[0.076]	[0.036]	[0.028]	[0.027]	[0.032]
Equal with 10th percentile (p-value)	-	-	0.6449	0.6485	0.5768	0.5214
Equal with 90th percentile (p-value)	-	0.5214	0.0631	0.6132	0.7658	-

Notes: Robust clustered standard errors in parentheses. Bootstrapped standard errors in brackets (100 reps), * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Explanatory variable is parental household income. OLS model controls for the average age of parents and children and birth cohort dummies. Source: BHPS and Understanding Society.

Table B.3 Estimates of relative and absolute mobility by regions

	Relative Mobility				Absolute Mobility	
	IGE		Rank		R25	
Region	OLS	TSRA	OLS	TSRA		N
North East	0.298	0.326**	0.370**	0.385***	35.04	65
	(0.18)	(0.15)	(0.18)	(0.12)		
North West	0.266***	0.319***	0.273***	0.260***	42.94	201
	(0.066)	(0.073)	(0.076)	(0.075)		
Yorkshire and Humber	0.385***	0.445***	0.403***	0.446***	31.79	150
	(0.11)	(0.098)	(0.10)	(0.085)		
East Midlands	0.466***	0.407***	0.432***	0.383***	37.43	153
	(0.12)	(0.11)	(0.11)	(0.11)		
West Midlands	0.355***	0.336***	0.286***	0.268***	38.84	126
	(0.11)	(0.12)	(0.093)	(0.093)		
East of England	0.207**	0.114	0.279***	0.224**	49.33	151
	(0.83)	(0.079)	(0.099)	(0.098)		
London	0.231*	0.166	0.248**	0.199*	56.46	132
	(0.12)	(0.11)	(0.12)	(0.10)		
South East	0.148**	0.129**	0.150**	0.164***	53.73	259
	(0.063)	(0.059)	(0.066)	(0.061)		
South West	0.123	0.104	0.195**	0.226***	42.90	142
	(0.077)	(0.071)	(0.094)	(0.069)		
Wales	0.220**	0.264***	0.334***	0.322***	38.11	298
	(0.10)	(0.072)	(0.061)	(0.057)		
Scotland	0.317***	0.260***	0.413***	0.364***	40.40	256
	(0.082)	(0.091)	(0.072)	(0.069)		
Northern Ireland	0.272***	0.237***	0.259***	0.229***	44.93	192
	(0.078)	(0.080)	(0.079)	(0.076)		

Notes: Robust clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Ranks were based on the national household income distribution in each generation. These income measures are described in the text in section 2.3.2. OLS models controls for the average age of parents and children, their average age squared and birth year dummies. TSRA models control for a quartic in age of parents and children, as well as year dummies. The measures of relative and absolute mobility are explained in the text in section 3.4.3. Source: BHPS and Understanding Society.

Chapter 4

Anticipated Discrimination and Educational Achievement: a study of ethnic minorities in England

4.1 Introduction

The accumulation of human capital remains a topic of substantial interest to academics and policy makers, as it is one of the key factors for economic success. Educational achievement and, more broadly, the levels of cognitive skills brought to the labour market have important implications for long-term outcomes and opportunities for social mobility.[83] There is a large literature on the determinants of educational attainment of ethnic minorities students, including the role of family background and socio-economic status, of pupil's characteristics, attitudes and aspirations and the role of the local environment through opportunities in the labour market and schooling factors (Burgess and Heller-Sahlgren, 2018; Demie and Strand, 2006; DfES, 2005; Heath et al., 2008; Modood, 2003; Strand, 2014; Strand et al., 2015; Wilson et al., 2011). Despite growing evidence on a multitude of factors affecting the decisions of human capital acquisition, the anticipation of facing discrimination in the labour market remains an understudied aspect.

[83]Cognitive skills represent a broad set of skills, competencies and expertise that can be acquired and developed with human capital investment (i.e. more education or training). Although the focus here is on cognitive skills, gaps in non-cognitive skills (motivation, self-control, time preference, social skills) also emerge in early years and are important determinants of socio-economic status in adulthood (Carneiro et al., 2005; Heckman et al., 2006).

As various theoretical models of discrimination have put forth, discrimination in the labour market can influence individuals' incentives to invest in human capital, via expectations of future returns to those investments (Arcidiacono et al., 2010; Coate and Loury, 1993*b*; Lang and Manove, 2011; Lundberg and Startz, 1983). Drawing on these models, it is expected that the effect of anticipating discrimination could go in two opposite directions. On the one hand, early models of discrimination suggest that expecting discrimination could reduce the incentives to invest in human capital if ethnic minorities believe that these investments would not be rewarded fairly and would attract lower relative returns in the labour market. This could occur, for example, if investments in human capital are imperfectly observed by the market and education is a less reliable signal of productivity for members of minority groups (Lundberg and Startz, 1983; Phelps, 1972), or if employers hold negative stereotypes about ethnic minorities (Arrow, 1973; Coate and Loury, 1993*b*). Meanwhile, more recent models suggest that anticipating discrimination could lead to an increase in investment in human capital if these investments can be directly observed, indicating that some ethnic minorities may have a greater incentive to signal their ability via educational achievement (Arcidiacono et al., 2010; Lang and Manove, 2011).

Whether or not discrimination in the labour market influences investment in human capital in practice is therefore an empirical question, which I aim to address in this chapter. While some previous studies have sought evidence of the influence of discrimination on educational decisions by conducting indirect tests and examining ethnic differences in the labour market returns to education and ability (Arcidiacono et al., 2010; Carneiro et al., 2005; Lang and Manove, 2011; Neal, 2006), very little evidence exists on the direct role of anticipated discrimination. This knowledge gap persists, primarily due to the scarcity and difficulties in obtaining the necessary data on individual expectations. I address this research question by examining the relationship between anticipated discrimination and the educational performance of British ethnic minority adolescents. Particularly, I consider their exam performance at the end of compulsory schooling (so-called Key Stage 4), which marks a crucial point within the educational system.

The UK makes an interesting case for the study of ethnic minorities due to the sizeable and incredibly diverse ethnic composition of its population. The evidence on ethnic gaps in wages, employment and other labour market outcomes (Berthoud, 2000; Cabinet Office, 2017; Heath and Cheung, 2006; Office for National Statistics, 2019*b*) make it plausible that the expectations of facing discrimination in the labour market could also vary with the specific experiences of each ethnic group and therefore it is also important to examine this heterogeneity. The UK evidence also reveals variations in the educational achievements of different ethnic minority groups, although interestingly with an unexpected pattern of high academic attainment. With the exception of pupils of Chinese origin, most ethnic minority pupils lag considerably behind their White British counterparts at the beginning of primary school (DfES, 2005; Dustmann et al., 2010). Against many odds, most of these pupils (with the exception of the Black Caribbean group) catch up during secondary school and display, on average, higher achievement at later stages of schooling (DfES, 2005; Dustmann et al., 2010; Strand, 2014; Strand et al., 2015; Wilson et al., 2011) and higher rates of participation in post-compulsory education (Khattab, 2018; Leslie and Drinkwater, 1999; Modood, 2003) than the White British students.

To address the key knowledge gap of how anticipated discrimination influences actual educational outcomes related to investment in human capital, I use data from Next Steps, a unique dataset comprising participant responses on their expectations of facing future discrimination in the labour market. This provides a suitable measure to investigate the direct role of anticipated discrimination. Next Steps surveyed a cohort of English students born in 1989-1990, providing a wealth of information on pupils' individual characteristics, family background and parental characteristics. Further, this dataset can be matched to the National Pupil Database (NPD), which contains sensitive data on educational attainment of pupils at different key schooling stages. While other studies group all individuals of non-white descent together as Black, Asian and Minority Ethnic (BAME), a generalisation that likely hides important distinctions, here I also examine heterogeneity across different ethnicities. The over-sampling of pupils from ethnic minority backgrounds in Next Steps allows detailed ethnic comparisons on the relationship between anticipated discrimination and educational outcomes of these pupils. In terms of estimation, I use least squares to adjust for a series of observable factors, using an extensive set of controls to reduce the potential omitted variable bias. While a bias from omitting unobserved variables may remain a concern, I employ the method developed by Oster (2019) to assess the potential selection on unobservables by estimating bias-adjusted upper and lower bounds for the coefficient of anticipated discrimination. Finally, to address the possibility that expectations of facing discrimination may be endogenous to academic potential, I consider value-added models accounting for the earliest measure of achievement available in the data i.e. test scores at ages 10-11 (Key Stage 2). However, it remains possible that even at this younger age, test scores could be influenced by expectations of facing discrimination, so these results may under-estimate the effect of anticipating discrimination.

To summarise the results, I find a positive and significant association between anticipating labour market discrimination and better educational outcomes among ethnic minority students, after controlling for an extensive set of individual and family background characteristics. Students who anticipate discrimination are more likely to achieve five or more A*-C GCSEs including English and Maths, the GCSE 'gold standard'. They also achieve a higher total GCSE score and more standard passes (A*-C) in English or Maths. Interestingly, the magnitude and sign of this association is not homogeneous across different ethnic minority groups. I provide evidence that the 'push' effect from anticipating discrimination is stronger for pupils of South Asian origin, who exhibit particularly higher probability of achieving high Maths scores and the GCSE gold standard when anticipating discrimination. On the other hand, a weaker influence of anticipated discrimination is observed for pupils of Black Caribbean and Black African ethnicity. These results are not seriously threatened by a potential selection on unobserved variables and, as expected, are weaker when taking into account prior academic achievement.

These findings mark significant contributions to the literature in two main ways. Firstly, they provide empirical evidence that anticipated discrimination on ethnic grounds can influence the educational outcomes of ethnic minority pupils. This is an important finding that is in line with the theoretical contributions of Lang and Manove (2011) and Arcidiacono et al. (2010). Secondly, the study adds to the debate on understanding the factors behind the educational achievement of ethnic minorities in the UK and how they differ across ethnic minority groups.

In the next section, I briefly review the relevant theories of labour market discrimination and present a discussion of how discrimination may influence human capital investment decisions. Section 4.3 presents an overview of the situation of ethnic minorities in the UK. Section 4.4 describes the data and the strategy employed to estimate the link between anticipated discrimination and educational achievement. Section 4.5 presents and discusses the main results and sensitivity checks. I end with conclusions in Section 4.6.

4.2 Theoretical framework: Labour market discrimination

This research resonates with a wider theoretical literature on labour market discrimination, which describes how discrimination arises and how ethnic differentials in labour market outcomes can be generated in the presence of discrimination. For the purposes of this chapter, it is critical to understand how discrimination and the expectations of facing discrimination in the future might influence the development of productive characteristics, in particular the decision to acquire human capital, which affects the formation of cognitive skills. In this section, I first review the assumptions and predictions of the main theoretical models of discrimination. Then, I discuss how the theory considers the endogenous role of discrimination for human capital formation.

The origins of labour market discrimination

Economic theories offer various explanations of how discrimination arises in the labour market. These have been recently reviewed in detail by Fang and Moro (2011), Lang and

Lehmann (2012) and Guryan and Charles (2013). Broadly speaking, the theoretical models of discrimination can be divided in two categories: taste-based and statistical discrimination.

Prejudice (or taste-based) models consider that certain individuals have a negative preference, or dislike, towards members of another group. According to the seminal model of taste-based discrimination developed by Becker (1957), differences in the treatment between groups in the labour market may arise when at least some employers, customers or workers have personal prejudice or dislike towards interacting with members of a particular group, and are willing to pay a price to avoid this interaction. Becker further argues that this will produce incentives for segregation in the labour market and that preferences for certain groups are unrelated to preferences for more productive workers.

A second strand of theories describes another possible origin of labour market discrimination using models of statistical discrimination, which refers to "the phenomenon of a decision-maker using observable characteristics of individuals as a proxy for unobservable, but outcome-relevant, characteristics" (Fang and Moro, 2011, p.134). In the case of ethnic wage and employability differences in the labour market, statistical discrimination is thought to arise from imperfect information (or imperfect observability) regarding an individuals' productivity in the labour market. As a consequence, employers resort to making assumptions about the productivity of potential workers using observable characteristics (e.g. race, ethnicity, nationality) and group statistics as proxies.

The social origins of taste-based discrimination have been largely studied by psychologists and sociologists, while economists have mostly focused on and proposed different explanations for why statistical discrimination arises. The seminal models by Phelps (1972) and Arrow (1973) explain why minorities might be treated differently in the labour market even in the absence of explicit preferences against this group. In the model of statistical discrimination developed by Phelps (1972), the source of discrimination is either some unexplained exogenous ex-ante difference or some difference in the reliability of the productivity signal between two groups of workers. In this model, employers observe a worker's group membership (e.g. ethnicity or gender) and some signal of their productivity obtained via an imperfect test.[84] With the signal being generally less strong or reliable for one group, members of this group with high, but less reliable, signals receive lower wages than those of members from the other group with high, but more reliable, signals in equilibrium.[85]

Another strand of models, established by Arrow (1973), considers the emergence of group differences endogenously in equilibrium, as a product of ‘self-fulfilling prophecies’. Here, the two groups are assumed to be equal ex-ante and there is a shift in explaining the origin of statistical discrimination from group differences to employers’ beliefs. In contrast to Phelps, Arrow (1973) assumes that employers have equal difficulty observing the productivity of all groups. Employers incur costs to determine a worker’s true productivity and have some idea (preconception) of the distribution of productivity of each of the workers types. Thus, they rely on observable characteristics (e.g. group membership) to infer the worker’s skill level, and as before, a noisy signal of qualification (e.g. a test, interview), which implies that the employer’s differential judgement about the probability of a worker being qualified will be reflected in the wages. If workers from one group are believed to be qualified with a lower probability or to be less productive, employers will hold negative stereotypes about this group and its workers will be paid lower wages in equilibrium.[86]

Discrimination and investment in human capital

In this chapter, the main interest relates to the consequences of labour market discrimination (either statistical or taste-based) for human capital investment decisions through the presence of individual expectations and beliefs about this phenomenon. Here, I discuss how the main

[84] The employer is able to assess an applicant’s performance with some kind of test, which measures the applicant’s promise of degree of qualification (with some error). This test score is then used as a signal of the worker’s potential skill level.

[85] This difference in reliability of the signal can be generated, for example, if the error term in a performance test used to signal productivity is related to group membership (e.g. to race).

[86] In a later formalisation of Arrow’s model, Coate and Loury (1993*b*) assume that wages are determined exogenously to the model (i.e. workers receive equal pay for equal work), which is different from Arrow’s assumption that minorities receive lower wages in equilibrium. This change in assumption shifts the focus to discrimination in job assignment (e.g. hiring), rather than wages.

models of discrimination explain the link between labour market discrimination and workers' decisions of acquiring human capital prior to entering the market.

In the theoretical literature of discrimination, the decision to invest in human capital is an endogenous choice, which means that the presence of discrimination in the labour market can influence this decision as part of a pre-market process. A crucial assumption of the traditional models of discrimination is that employers only have imperfect information about productivity and receive no direct information about investments in human capital. From these models that assume employers use a noisy productivity signal or prior beliefs in labour market decisions, it follows that the separating discriminatory equilibrium creates incentives for members of the ethnic minority group to invest less in human capital.

Arrow (1973) and Coate and Loury (1993*b*) describe how asymmetric beliefs between groups and negative stereotypes can be perpetuated through self-fulfilling prophecies. To understand how this happens, it is important to consider the assumptions of the model by Coate and Loury (1993*b*), who build on the earlier model by Arrow (1973). This model assumes that there are two types of workers: skilled workers are paid higher wages but need to incur a skill investment cost before being able to work and unskilled workers who do not need to invest in skills and receive lower wages. The cost of investing in skills does not vary between groups. Human capital investments are made prior to entry in the market and are not directly observed by the employer. If skilled workers from a certain group do not receive the appropriate wages or job assignment, this might lower the incentive for other workers from this group to invest in human capital in order to raise their productivity. Hence, employers' prior negative beliefs about the group are confirmed in equilibrium. This model therefore predicts that ethnic minorities will under-invest in education if they are believed to be on average less qualified.

The result will be similar if discrimination is assumed to originate from exogenous differences in the reliability of the productivity signal between groups, rather than asymmetric beliefs, as in the framework put forward by Phelps (1972) and further developed by Lundberg and Startz (1983). Workers from the group with the noisier signal will, again, have lower incentives to invest in human capital. In their model, investment in human capital is a costly choice made prior to entry to the labour market that influences a worker's acquired levels of skills.[87] Workers know their own productive abilities, innate and acquired, and decide the optimum level of investment in human capital to maximize wages net of education costs. Employers, on the other hand, only observe a worker's productivity signal, which is a function of the worker's marginal product and group membership, but they do not observe endowed or acquired human capital directly.[88] Assuming that the productivity signal is less reliable for minorities, the model predicts that employers will rationally discriminate between the two groups, offering separate wage schedules in equilibrium.[89] [90] Workers respond to the separate wage equilibria, with workers from the minority group investing less in human capital, becoming the low-wage/low-skilled group.

While most of this literature has focused on statistical discrimination, Coate and Loury (1993*a*) developed a taste-based discrimination model that predicts that prejudice against some minority group might also reduce incentives to invest in skills. Under this model, employers experience a negative pay-off when hiring workers from a minority group, thus discrimination translates into firms hiring less workers from this group, lowering their incentive to invest in skills. Minority workers are therefore under-represented among workers hired, with respect to their representation in the population and in the group of skilled workers.[91]

[87] For Lundberg and Startz, human capital investment is not only comprised of formal schooling, but also of other (unobserved) skills acquired in the process.

[88] The authors describe that this signal could be in the form of a test score, as suggested by Phelps, or simply a measure of all information acquired by the employer in the hiring process (Lundberg and Startz, 1983, p.342).

[89] To describe the discriminatory equilibria, Lundberg and Startz assume that employers know the density function describing the distribution of workers' productive characteristics (innate ability and acquired human capital) in the population. These characteristics are randomly distributed in the population, and both groups have identical mean innate ability, mean error in the test scores and the same test variance. The difference between the two groups arises because of the assumption that the majority group has a higher variance in the innate ability distribution (more heterogeneous) and a lower variance in the testing ability distribution (more homogeneous), when compared to the minority group (Lundberg and Startz, 1983, p.344).

[90] Because the worker's wage is determined by the expected marginal product conditional on the signal and group membership (ethnicity), workers with the less reliable signal will get paid lower wages than the other group, in equilibrium.

[91] As Coate and Loury (1993*a*) further explain, in this setting, these disparities will not necessarily be eliminated by statistical enforcement policies such as quotas. Workers from the group targeted by quotas might

Building on traditional models of discrimination, other authors have questioned the prediction that ethnic minorities will always under-invest in skill accumulation (Arcidiacono et al., 2010; Lang and Manove, 2011). Notably, using a model of statistical discrimination and educational sorting, Lang and Manove (2011) (hereafter LM) predict that some minorities might over-invest in education with the aim of increasing their signal of productivity for the market.[92] A central difference to the earlier models of discrimination is the critical assumption made by LM; at least some aspect of human capital investment is directly observable by the market.[93] Following the literature on traditional signalling models (Spence, 1973; Weiss, 1995), LM assume that observable investments in education (i.e. specific test scores or degree certificates) can signal productivity.

In their signalling model with two types of workers, LM show that the presence of statistical discrimination in the labour market creates incentives for minority workers to signal their ability using education. In this model, education is a more valuable signal when the direct observation of a worker's productivity is less reliable.[94] In addition, the ability of firms to assess a worker's productivity improves with increased education for both groups, and at higher education levels firms are equally able to assess workers from the two groups.

LM describe the possible direct and indirect effects of investing in more education. The *direct* effects of additional education on the employer's inference of productivity (keeping inferred ability constant) works through two main pathways: (i) additional education leads the employer to infer higher productivity; and (ii) additional education increases the expected value of the signal. The *indirect* (signalling) effect of additional education occurs because

be persuaded to make lower investments in human capital, but if some members of some group still acquire skills, firms will prioritise hiring these workers and a separating equilibrium will happen.

[92] Lang and Manove (2011) assume that employers observe group membership (in this case, race) freely. In addition, workers of a same ethnic group may differ in ability and educational attainment, and both these factors are complementary inputs in the creation of productivity. Only the worker can observe his own ability, but skill investments (education attainment) are observable by the market precisely, and productivity is observed directly but with some error.

[93] For instance, in the models by Lundberg and Startz (1983) and Coate and Loury (1993*b*) productivity is only imperfectly observed by the market and no direct information about investment in human capital is available to employers.

[94] LM assume that the worker cannot deduce his productivity from knowledge of his ability and education alone, and that there is a random, unobserved element. As a consequence, employers will rely on their own direct observation of a worker's productivity, which has a higher expected value for high ability workers.

it increases the employer's inference of a worker's ability. Therefore, assuming it is more difficult for employers to directly observe the productivity of ethnic minority workers, firms will put more weight on education and less weight on observed productivity for minority workers.[95] The assertion here, is that education is a more valuable signal of ability for minority workers than for White workers. Consequently, minority workers of intermediate ability will have an incentive to obtain more education than White workers of the same ability, and will make observable investments in education, for example through post-compulsory schooling and obtaining degrees and certificates.[96] Testing the main predictions of their model empirically, LM find evidence of differences in observable investments in education prior to entry in the labour market, to the extent that African Americans have lower education than Whites, on average, and that conditional on ability (measured by AFQT scores), they obtain more education than Whites (in years of schooling). This latter observation is consistent with African Americans over-investing in observable aspects of human capital, which they attribute to being a response to discrimination.

Arcidiacono et al. (2010) relate closely to these ideas. Using a model of statistical discrimination and dynamic learning, the authors argue that education plays more than just an initial signalling role in wage determination.[97] In fact, they argue that college graduation plays a key role in the direct revelation of ability to the market because, upon being presented with information on college attendance, grades and completion, employers learn about a potential worker's ability. Their model predicts that college education symmetrically improves the precision of the signal for workers from both ethnic groups ('black' and 'white' workers). However, the value of this increased precision is greater for minority workers because

[95]LM justify this assumption by referring to a body of work that suggests that employers usually find it more difficult to evaluate the productivity of ethnic minority candidates, especially of those with lower and intermediate levels of education. This is because of a number of factors, including networks (or lack thereof) and differences in communication cues, among others.

[96]Minority workers with high ability obtain the same level of education as comparable White workers. Meanwhile, workers with the lowest ability will obtain the lowest level of education, which is efficient and not influenced by signalling in a well-behaved separating equilibrium.

[97]Traditionally, in models of employer dynamic learning in the lines of Farber and Gibbons (1996), education serves as a tool for workers to signal ability but the role of this signal in determining wages decreases with experience, as further information is revealed to the employer.

employers have prior beliefs about ethnic differences in the distribution of ability that drives statistical discrimination in the labour market for high school graduates.[98] This creates a greater incentive for ethnic minority workers on the college-high school margin to attend college. Once again, this supports the observation that, conditional on ability, minority workers have an incentive to more education.

In line with the various theoretical models discussed in this section, this literature suggests that the existence of labour market discrimination might influence human capital investment decisions both before and after individuals enter the labour market, making it harder to distinguish between market and pre-market processes. Perceived discrimination and the expectations of facing discrimination in the future could influence pre-market educational outcomes, general well-being, as well as other behaviours and decisions, and thus accentuate existing differences between groups. According to these theories, discrimination in the labour market could have contrasting effects on human capital investment. On the one hand, if human capital investments are unobservable to employers, anticipating discrimination (and low future returns) could lower the incentives to invest in the acquisition of additional skills and education. On the other hand, if these investments are observable to employers, the incentive to invest in human capital is greater as education can be used to signal higher productivity in an attempt to counteract the 'ethnic penalty'. This over-investment behaviour might be further reinforced by parental or cultural values that perceive formal education as the means to achieve a better life.

Empirical investigations of anticipated discrimination

Despite the large body of literature focussing on measuring the extent of labour market discrimination and on investigating the role of various pre-labour market factors, only a

[98] In this model, in contrast to Lang and Manove, Arcidiacono et al. (2010) do not assume racial differences in the precision of additional education as a productivity signal. Instead, they argue that statistical discrimination arises because employers have a prior belief about racial differences in the distribution of ability, as in the model proposed by Arrow (1973). The authors argue that in the high school market, where ability is initially unobserved by employers, there are greater incentives to statistically discriminate on the basis of race. In contrast, in the college graduates market, ability is more directly revealed for both race groups.

handful of studies have examined the role of expectations of facing future discrimination on pre-labour market outcomes empirically. To the best of my knowledge, all (but one) have done so indirectly. This is primarily due to difficulties in obtaining data on expectations of facing discrimination, which prevents direct tests of how expected labour market discrimination shapes human capital formation.

The study by Fernández-Reino (2016) is, to the best of my knowledge, the only study that exploits expectations of discrimination directly. This study examines the influence of anticipated discrimination on post-16 education continuation decisions of ethnic minority students relative to White students in England, finding only modest, and mostly non-significant, effects of anticipating discrimination. However, as the author controls for prior achievement in all models, which could, itself be affected by expectations of discrimination, it is possible that these results are affected by endogeneity problems.

Indirect tests, in the likes of analyses by Neal and Johnson (1996), Lang and Manove (2011) and Arcidiacono et al. (2010), are more common. These three papers examine race-specific returns to cognitive skills in the US. Neal and Johnson (1996) demonstrate that controlling for a measure of cognitive ability (the AFQT score)[99] in a reduced-form model of wage determination considerably reduces the observed racial wage gap. The authors argue that cognitive ability measured by this test in late teenage years is a pre-market factor, and is not affected by expectations or experiences of discrimination in the labour market.[100] They report that differences in the AFQT test scores explain three quarters of the wage gap between black and white young men and almost all of the racial gap for young women. This result implies that ethnic wage gaps are primarily driven by pre-market ethnic disparities in productive

[99]The AFQT are a subset of tests administered as part of the Armed Services Vocational Aptitude Battery in the US. They are used for enlistment screening and job assignment within the US military services. Neal and Johnson (1996) claim that the AFQT is a racially unbiased measure of basic skills that helps predict actual job performance. Analysing respondents who were 18 or younger when they took the AFQT test means that most had not yet entered the labour market full-time nor started post-secondary schooling. However, other studies have called attention to the lack of consensus about what AFQT scores measure (Darity and Mason, 1998).

[100]Here, pre-market factors are factors determined outside of the market but that play an important role in accounting for ethnic differentials - in wages, employment prospects and other outcomes - in the labour market (Carneiro et al., 2005; Neal and Johnson, 1996).

characteristics (in this case, an ethnic cognitive ability gap), rather than discrimination (Neal, 2006).

Lang and Manove (2011) have demonstrated that controlling for both AFQT scores and schooling at the time of the test, the racial wage gap reappears. Thus, adjusting the AFQT scores for ethnic differences in years of schooling considerably reduces their role in explaining the wage gap. This indicates that part of the AFQT gap observed by Neal and Johnson is actually explained by the generally lower levels of schooling of ethnic minorities.[101] These results also imply that ethnic minorities have more years of schooling than white individuals of same ability (as measured by the AFQT). Lang and Manove (2011) argue that discrimination might increase the incentives for ethnic minorities to make observable investments education. In line with this argument, Arcidiacono et al. (2010) show that, conditional on ability, ethnic minorities have an extra incentive to attend college. They show that returns to observed ability (measured by test scores) are large for college graduates and considerably lower for high school graduates, whose earnings only rise gradually with experience and employer learning. Since ethnic minorities face a higher ethnic penalty in the high school market but not so much in the college graduates market, their incentive to obtain a college degree is higher.

Although individual expectations of facing discrimination are not directly observed, ethnic differences in returns to education when controlling for ability, as those found by Lang and Manove (2011), may be suggestive that minority individuals respond to (expected) discrimination by changing their investment in human capital prior to market entry. Other studies have looked at more general expectations. For example, Carneiro et al. (2005) show that ethnic minority children in the US as young as age 10 already expect to achieve lower educational levels. If these pessimistic expectations are a result of perceived labour market discrimination, the subsequent lower levels of investment would also be attributable to discrimination.[102]

[101] However, these gaps in test scores could also be explained by lowered academic effort in anticipation of future discrimination or adverse environments (Carneiro et al., 2005).

[102] There could be other possible reasons for these more pessimistic expectations, for example lower school quality or other adverse environments.

In this chapter, I aim to disentangle the role of anticipated discrimination using a comprehensive dataset, which allows the direct role of student's expectations to be tested empirically. Before describing the data and estimation strategy used in my main analysis, the next section contextualises the situation of ethnic minorities in the UK in terms of labour market and educational outcomes, and provides more detailed information about the educational system in England.

4.3 Background: Ethnic Minorities in the UK

4.3.1 Ethnic minorities in the labour market

In the UK, individuals from ethnic minority groups comprise a large and growing share of the population. In line with a large international literature on ethnic gaps, the examination of socio-economic outcomes for these groups is of great public interest. The latest Race Disparity Audit commissioned by the UK government identified ethnic differences in several socio-economic outcomes, including living standards, housing, work, policing and health (Cabinet Office, 2017).

The unique ethnic composition of its population makes the UK an interesting case study of ethnic minorities. Alongside a White British majority, the population is comprised of individuals from another six large ethnic minority groups, namely Black Caribbean, Black African, Indian, Pakistani, Bangladeshi and Chinese.[103] According to the most recent Census (2011), 19.5% of the population in England and Wales classified themselves as being from these ethnic minority groups, corresponding to more than 11 million people (Office for National Statistics, 2018).[104] [105] This also represents a large increase from 12.6% in the 2001

[103]These groups are also known as BAME, referring to Black, Asian and Minority Ethnic people.

[104]Considering the combined censuses of Scotland, Northern Ireland, England and Wales, 13% of the population are from ethnic minority groups. Numbers from the ONS 'Ethnicity Facts and Figures' https://www.ethnicity-facts-figures.service.gov.uk/

[105]Reflecting the large fraction of individuals of ethnic minority origins in the British population, the Department for Education (DfE) also reports a large participation of pupils from the main ethnic minority groups in maintained schools in England, especially those from Indian, Pakistani and Black African origins (DfES, 2005).

Census. The presence of these diverse groups in the country is linked to immigration flows that took place for various reasons across different periods following the 1950s (Dustmann and Theodoropoulos, 2010).[106]

Ethnic minorities also make up a considerable share of the British labour force. According to the ONS, ethnic minority groups comprise 20.3% of Britain's workforce (Office for National Statistics, 2019*b*). These minorities are the focus of public surrounding the disadvantages and challenges faced by members of these communities. Several studies have documented the presence of ethnic gaps in the labour market, including disparities in wages , occupations and employment (Berthoud, 2000; Cabinet Office, 2017; Heath and Cheung, 2006; Office for National Statistics, 2019*b*).

A recent report on ethnic pay gaps by the Office for National Statistics (2019*b*) emphasises the large wage differences between ethnic groups in Britain. When compared to White British workers, employees from certain ethnic groups fare better, while others do worse. For instance, workers of Chinese and Indian heritage earn on average 30.9% and 12% more than White British workers, respectively. Meanwhile, workers of Pakistani and Bangladeshi origin earn on average 16.9% and 20.2% less, respectively (Office for National Statistics, 2019*b*).[107] Alongside the ethnic differences in pay, the report also describes a large variation in employment and unemployment rates and in labour force participation between ethnic groups (Heath and Cheung, 2006; Longhi and Brynin, 2017; Office for National Statistics, 2019*b*). People from the Pakistani and Bangladeshi ethnic groups are more likely to be unemployed and more likely to work in low-skilled jobs and receive the lowest hourly pay (Cabinet Office, 2017). This picture becomes even more complex when comparing women

[106]Most immigrants from the Caribbean arrived in the period from 1955 and 1965, while most immigrants from Indian, Pakistani and Black African origin arrived between 1865 and 1974, and most of Bangladeshi origin arrived in the early 1980s (Dustmann and Theodoropoulos, 2010).

[107]Office for National Statistics (2019*b*), based on data from the Annual Population Survey 2018.

from different ethnic groups and when considering the distinction between UK-born and immigrants.[108] [109]

Raw (unadjusted) ethnic gaps usually narrow after controlling for certain individual productivity-related characteristics, suggesting that potentially endogenous factors, including occupational choice, age, experience and education acquired prior to entering the labour market are significant drivers of the ethnic pay gaps in Britain (Berthoud, 2000; Blackaby et al., 1999; Brynin and Güveli, 2012; Dustmann et al., 2003; Dustmann and Theodoropoulos, 2010; Heath and Cheung, 2006; Li and Heath, 2018; Metcalf, 2009). This means that there are substantial differences in these characteristics across ethnic groups. For example, differential occupational choices have been shown to matter considerably, as some ethnic minorities are much more likely to work in low-skilled, low job occupations (Brynin and Güveli, 2012; Cabinet Office, 2017; Elliott and Lindley, 2008; Longhi et al., 2012). However, in contrast to other countries, such as the US, studies for the UK have shown that controlling for the level of qualifications increases the general ethnic wage gap (Blackaby et al., 2002; Heath and Cheung, 2006; Metcalf, 2009). This is because individuals from certain ethnic minority groups have, on average, higher educational attainment and qualifications than their White British counterparts. Thus, these minorities obtain relatively lower returns to qualifications (Heath and Cheung, 2006; Metcalf, 2009).

Nevertheless, even after controlling for individual differences in education levels and other productive characteristics, there is evidence that some unexplained ethnic gaps remain

[108] Among women, ethnic pay gaps are much smaller and some groups have a pay advantage, earning on average more than women from the White British group. However, large differences are observed regarding women's labour market participation, with evidence of women from certain ethnicities experiencing a 'double disadvantage' in the labour market. Participation rates are considerably lower for women of certain ethnic groups, for example among women of Pakistani and Bangladeshi origin (below 30%) (Cabinet Office, 2017; Longhi and Brynin, 2017).

[109] Country of birth is one important factor that explains the ethnic pay gap in the UK (Longhi and Brynin, 2017). Newly arrived immigrants might encounter several difficulties in the host country, including problems with the language and culture, recognition of qualifications and lack of a established social network (Blackaby et al., 2002; Lindley, 2002). Nevertheless, significant ethnic differences in pay are observed even for UK-born minorities (Longhi et al., 2012).

(Berthoud, 2000; Blackaby et al., 2002; Brynin and Güveli, 2012; Longhi et al., 2012).[110] As discussed before, this unexplained part of the gap is typically attributed to discrimination.

4.3.2 The educational performance of ethnic minorities

An overview of the educational system in England

Schooling in England is compulsory between the ages of 5 and 16, the current statutory leaving age.[111] One of the main features of the English educational system is the centralised assessment of pupils. Pupils are tested in 4 Key Stages (KS), as summarised in Table 4.1, with the first being the Key Stage 1 in years 1 and 2 and the last stage in compulsory education being the Key Stage 4 in years 10 and 11. For each stage, the government sets specific achievement targets for pupils, and students take nation-wide exams graded by external examiners. From KS1-KS3, pupils are assessed in the core disciplines, Maths, English and Sciences. The exams at the end of KS4 are known as the General Certificate in Secondary Education (GCSE) and they are comprised of Maths and English as core subjects, plus a variety of optional subjects. Since there is little or no grade repetition in England, pupils who enter school in the same year take their KS exams in the same year. Between the ages of 16-18, students have to stay in full-time education or training (apprenticeship), although until 2008, in the time frame relevant to this analysis, students had also the option to leave education after the age of 16. Pupils who decide to stay in full-time education study in a sixth form college (A-levels) or further education college, which corresponds to Key Stage 5.

GCSEs are considered high stakes examinations and results are used by the Department for Education, policy makers and academics as a benchmark to measure pupils' educational achievement, their progress and to compare the quality of schools. During the time frame relevant to this analysis, GCSEs were graded on a letter scale from A*-G, with A* being the

[110] This is not only the case of ethnic wage gaps but also for other labour market outcomes, such as gaps in employment prospects and occupation (Carmichael and Woods, 2000), as well as unemployment (Longhi and Brynin, 2017).

[111] However, since the Education and Skills Act 2008 students are also required to participate in some form of full-time education or training until the age of 18.

highest possible grade and C being the standard 'good pass' grade that is usually regarded as a basic requirement by employers and universities.[112] Failure to achieve this grade is associated with higher drop-out rates, and a lower probability of entering high-level courses in post-compulsory education and of continuing into higher education (Machin et al., 2018). A good performance in GCSEs is therefore critical to accessing post-compulsory education and the results of these high stakes exams can greatly influence many other long-term outcomes (e.g. employment prospects, wages) (Kingdon and Cassen, 2010; Machin et al., 2018; McIntosh, 2006; McIntosh and Vignoles, 2001).

Table 4.1 The Educational System in England

Key-Stage	Year	Age	
KS1	1-2	5-7	Primary School
KS2	3-6	7-11	Primary School (SATS)
KS3	7-9	11-14	Secondary School
KS4	10-11	14-16	Secondary School (GCSE)
KS5	12-13	16-18	College, Sixth-Form (A-levels)

Ethnic differences in educational attainment

Given the strong relationship between educational outcomes, particularly of GCSE results, and other future labour market outcomes, a large literature has focused on ethnic attainment gaps and sought to explain their main determinants. Research shows that ethnic educational gaps are present in all stages of schooling. Pupils from ethnic minority backgrounds initially have much lower educational performance but catch up during their school years (DfES, 2005; Dustmann et al., 2010).[113] According to Dustmann et al. (2010), language proficiency is one of the main factors that contributes to this catch-up during the initial school years, while poverty works against it.

[112] The official GCSE grading system has undergone recent reforms in the 2010s, and exams are now graded from 1-9 according to the new rules.

[113] Dustmann et al. (2010) document that students from ethnic minority groups lag substantially behind their White counterparts during primary school (KS1). In the following years, this gap declines (and even reverses, for some ethnic groups) during secondary school.

Even with this catch-up, by the end of secondary school, significant ethnic attainment gaps are observed. Pupils from certain minority groups (Black Caribbean, Black, Pakistani, Bangladeshi) perform worse in GCSEs relative to White pupils, while those from other groups, such as those from Chinese and Indian origins, perform much better than the national average (DfES, 2005; Modood, 2003; Strand, 2014; Strand et al., 2015; Wilson et al., 2011). Ethnic gaps for the threshold of achieving five or more GCSE passes at A*–C grades including English and Mathematics by gender are illustrated in Figure 4.1. For the White British group, on average, 50.2% of boys and 60.1% of girls achieved the 'gold standard'. It is notable that pupils from Black Caribbean, Black African, Traveller of Irish Heritage and Gypsy/Roma origins have the worst performance among all pupils, for both genders. For instance, only 33.3% of boys from Black Caribbean backgrounds achieve 5 or more A*-C. On the other hand, pupils from certain groups achieve better GCSE results than White British pupils. Pupils from the Chinese group are the highest achievers, with 77.1% of boys and 85.1% of girls achieving the gold standard, followed by the pupils from Indian background, with 64.8% of boys and 75.8% of girls achieving these results.

Ethnic differences have also been observed in the UK with respect to later educational outcomes, with participation in post-16 education and higher education being notably higher among some ethnic minorities (Khattab, 2018; Leslie and Drinkwater, 1999; Lessard-Phillips et al., 2018; Modood, 2003). The reasons for such ethnic disparities in educational achievement are complex and cannot be attributed to one single factor.

In the literature on discrimination, which draws from the human capital models a la Becker (2009), individuals make decisions to acquire human capital in response to the expected returns of education. In contrast, much of the literature on the economics of education assumes an education production function model, in which the output of the educational process - typically the educational achievement of students - is conceptualised as a function of a series of cumulative inputs, including individual, parental and school characteristics, as well as innate endowments or learning capacities (Hanushek, 2010). In this sense, attainment is determined by a series of inputs, rather than being an individual choice. The relative

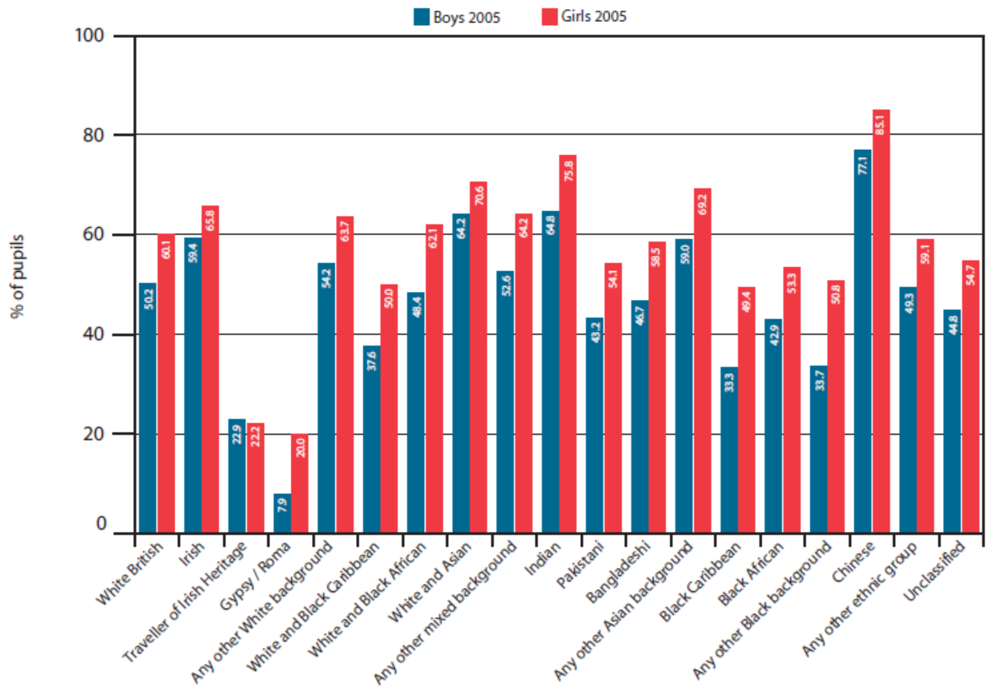


Figure 4.1 Pupils achieving 5+ A*-C GCSE and equivalent including English and Maths by ethnic group and gender (%), England, 2005.

Source:DfES (2005).

importance of some of these inputs may vary between ethnic groups, while others affect educational outcomes uniformly across all groups.

Parental and family socio-economic background is a major determinant of children's educational outcomes - a notion that is central to the literature on intergenerational transmissions (Björklund and Salvanes, 2010). In the UK, large differences in educational achievement by socio-economic background start in primary school and tend to increase as pupils get older, especially during secondary school (Crawford et al., 2017). In examining ethnic educational gaps, a key role is typically attributed to the interplay between family socio-economic background and ethnicity (DfES, 2005; Li and Heath, 2018; Strand, 2014, 2015). For example, studies have shown that among disadvantaged pupils receiving free school meals (FSM), ethnic minority pupils fare better than those from the White British group in GCSEs (Strand, 2014, 2015).[114] The DfE highlights many differences between ethnic groups in terms of deprivation, household characteristics, language and religion (DfES, 2005).

[114] With the exception of middle and high socio-economic status Black Caribbean boys (Strand, 2014).

Socio-economic background is also associated with many other factors that can influence attainment, including parental education, employment, school selection, neighbourhood effects, among others. Using data from Next Steps, Chowdry et al. (2009) show that parental education is a key explanation for differences in educational attainment for students from different socio-economic backgrounds in England. Parental education affects not only the marginal productivity of children's education but is also associated with other (often unobserved) environmental factors that may influence the development of cognitive skills, such as risk preferences, parental attitudes, time use and parenting skills (Björklund and Salvanes, 2010; Goodman et al., 2011).

Individual socio-demographic characteristics are also considered important drivers of educational attainment. The influence of gender on educational attainment has been widely described in the literature, with girls surpassing boys with regards to a variety of outcomes in England, including test scores and completion rates in secondary school, tertiary education completion rates and others (Demie, 2001; OECD, 2019; Strand et al., 2006). This is illustrated in Figure 4.1, which shows that girls were more likely to achieve five or more 5A*-C GCSEs including English and Maths than boys across almost all ethnic groups (the single exception being Travellers of Irish Heritage).

One factor especially relevant for ethnic minorities is language proficiency. Previous studies have shown that children who speak English as an additional language are usually from families where the dominant language is not English. For these pupils, English fluency is a key predictor of test scores, with large educational gaps between native speakers, bilingual students and those with lower English fluency (Demie and Strand, 2006; Strand and Demie, 2005; Strand et al., 2015). Language proficiency is also a key explanatory factor for the faster relative progress of some ethnic minorities throughout school. Dustmann et al. (2010) find that improved fluency of English in the first years of school explains why the gap closes from primary to secondary school for certain ethnic minority groups.

Pupils' attitudes, aspirations and expectations are also important drivers of differences in educational attainment. Heath et al. (2008) and Burgess and Heller-Sahlgren (2018) document differences in attitudes towards education between immigrants and native students, with immigrants being generally more positive and having higher aspirations than native students, even after controlling for a series of individual and family background characteristics. Fernández-Reino (2016) suggests that ethnic minority pupils in England are more likely to continue in post-16 education partly because of the 'immigrant optimism' associated with parental expectations about the likelihood of the child applying to university. Individual aspirations and expectations are often related to parental beliefs, attitudes and aspirations, and a broader set of cultural values that can be passed on to children and may influence educational achievement and their beliefs in the value of education as means of social mobility (Chowdry et al., 2009, 2011; Gregg and Washbrook, 2011). Therefore, they are also likely to vary between ethnic and socio-economic groups. For example, Chowdry et al. (2009) shows that children from advantaged backgrounds are more likely to have higher aspirations and expectations, and this is associated with higher educational attainment.

Recent sociology studies have discussed ethnic educational gaps from the perspective of a compensatory strategy. Lessard-Phillips et al. (2018) find that ethnic minority graduates of Russell Group institutions are less likely to do well in the labour market after graduation and more likely to adopt a compensatory strategy of continuing into post-graduate education to avoid short-term unemployment or underemployment. These ideas are summarised by Khattab (2018), who asserts that the higher motivation of ethnic minorities could be part of a 'defiance strategy' to counterbalance the effect of ethnic penalties in the labour market and that this would partially explain the higher participation rates in higher education among these groups.

Despite these substantial efforts in investigating the role of key determinants of educational achievement for ethnic minorities, research that considers the role of expectations of labour market discrimination remains scarce. Given that discrimination can influence decisions of investing in human capital, it is important to combine these ideas. Expectations of facing future discrimination on the grounds of characteristics related to ethnic identity and skin colour might be especially relevant for ethnic minorities, as they may influence their expectations of labour market opportunities and returns. This study combines the idea of family, school and individual characteristics being inputs in the production of educational attainment with the notion that individuals make decisions of investing in human capital based on the expected future market returns. Specifically, this chapter advances this discussion by examining the influence of ethnic minority adolescent expectations of facing labour market discrimination on a number of GCSE outcomes.

4.4 Data and Methodology

4.4.1 Data: Next Steps

Data from Next Steps, formerly known as the Longitudinal Study of Young People in England (LSYPE), is used in this chapter.[115] Next Steps is a national longitudinal cohort study, which initially followed over 15,000 children born in the UK between 1st September 1989 and 31st August 1990. To date, the dataset comprises of eight waves. Participants were first interviewed in the spring of 2004 (at ages 13 and 14) and then interviewed annually until 2010, resulting in seven waves of data. An additional eighth wave was added in 2015, when the original participants were aged 25.

The original drawn intended sample comprised of 33,000 pupils in year 9 attending schools in February 2004. The final issued sample of wave 1 was comprised of 21,000 pupils, of which 15,770 (74%) were reached. The first stage of sampling is at the school-level, separately for maintained schools, independent schools and Pupil Referral Units (PRUs). Maintained schools were stratified by deprivation status, and deprived schools[116] were over-sampled

[115] Data obtained from University College London, UCL Institute of Education, Centre for Longitudinal Studies (2018).

[116] Deprived schools are defined as those with a high proportion of pupils receiving free school meals, in the top quintile of the distribution.

by a factor of 1.5. Within each deprivation stratum, schools selection probabilities were calculated based on the number of pupils in year 9 from the six major ethnic minority groups (Indian, Pakistani, Bangladeshi, Black African, Black Caribbean and Mixed). Independent schools and PRUs were sampled using the school level annual schools census (SLASC). In the end, 838 maintained schools were selected, 52 independent schools and 2 PRUs. Of all 892 schools selected, 647 (73%) cooperated with the study. In the second stage of sampling, pupils were sampled within schools, depending on their ethnic group as recorded in the Pupil Level Annual Schools Census (PLASC) and on school selection probabilities. Pupils from the main ethnic minority groups were over-sampled to achieve the target issued sample of 1000 pupils in each group and to allow ethnic comparisons. The school and pupil selection probabilities ensured that all pupils within an ethnic group and deprivation stratum had an equal probability of being selected (Department for Education, 2011, p.7).

Next Steps is linked to the National Pupil Database (NPD), a census that contains pupil and school characteristics and attainment for all children. The linked secure dataset provides information on individual characteristics, family background, test scores at different key stages of schooling, as well as data on free school meal eligibility and Special Education Needs (SEN). For the main analysis, this chapter uses data from waves 1 and 2 of Next Steps, collected in 2004 and 2005, when respondents were aged 13-15 and in years 9 and 10 of the English schooling system. The educational progression of pupils from the Next Steps cohort is summarised in Table 4.1.

Next Steps provides a comprehensive dataset for this chapter since it contains unique information that captures pupils' expectations of facing future discrimination in the labour market. This feature of the data is of crucial importance for my main analysis, where previous studies of examining the role of anticipated discrimination have been constrained by data limitations.

In addition, Next Steps provides a very rich set of information on children's social, economic and educational characteristics. In both waves 1 and 2, beyond the young person's questionnaire, there is a parent questionnaire to collect information on a number of topics, such as family situation, household context and parental education and employment.[117] Responses to the parent interview are provided by the main parent or carer, usually the mother. Thus, Next Steps in combination with the NPD, affords the opportunity to analyse the influence of expectations of future discrimination in the labour market on GCSE performance.

Table 4.2 Educational progression of the Next Steps Cohort

Academic Year	Age	Stage	Next Steps Wave
1989-1990	0		
2000-2001	10-11	KS2 exams	
2003-2004	13-14	KS3 exams	w1(2004)
2004-2005	14-15		w2(2005)
2005-2006	15-16	KS4 exams (GCSEs)	w3(2006)
2006+	16+	post-compulsory education	

Sample

The analysis focuses on ethnic minority pupils in England. I start with a total of 4101 ethnic minority pupils interviewed in waves 1 and 2 of Next Steps. After dropping the cases with no data on educational achievement at GCSE and KS2 levels, anticipated discrimination, ethnicity, and who did not fill in the self-completion questionnaire or whose main parent was not interviewed, the sample for my main analysis comprises 3315 individuals.[118]

In order to account for the Next Steps survey design, oversampling of individuals from ethnic minority groups and non-response, I use the sample weights provided in the data to obtain descriptive statistics that represent the population.[119] This restores the original panel and provides representative proportions of pupils from all deprivation levels and ethnic groups in each wave of the data. For the regression analysis, unweighted results are presented

[117]Unfortunately, many characteristics of interest could not be observed in previous years before wave 1, since this information is not available in the data.

[118]No significant differences between pupils in and out of the final sample are observed with respect to anticipating labour market discrimination, gender, special education needs and parental aspirations. However, for some indicators of socio-economic status (e.g. home ownership and parental employment), proportions were slightly higher for pupils in the final sample.

[119]When applicable, survey weights are applied using the weights provided in wave 2.

throughout the text and weighted results in the Appendix C.[120] I note that my results are robust to weighting the sample using the survey weights provided.

Key variables

Outcomes: Educational achievement

The primary aim of this chapter is to analyse the relationship between anticipating discrimination and individual's investment in human capital, by examining performance in GCSE exams in 2006, at ages 15-16. This includes: total GCSE point score, achieving 5 or more A*-C GCSEs; achieving 5 or more A*-C GCSEs including Maths and English; A*-C in English; A*-C in Maths; highest Science score; and achieving any A*-A. Aside from the total GCSE point score and highest Science score, which are continuous variables and are standardised to have mean zero and standard deviation one, all other variables are dichotomous, assuming value 1 when the respective outcome is achieved and 0 otherwise.

Anticipated discrimination

The measure of anticipated labour market discrimination is obtained from the second wave of Next Steps, prior to GCSE examinations. Pupils aged 14-15 are asked: *'Do you think that your skin colour, ethnic origin or religion will make it more difficult for you to get a job after you leave education?'*, to which they respond 'yes', 'no', or 'don't know'. I interpret the answers to this question as reflecting their expectations of facing future discrimination in the labour market. I then create a binary variable that takes a value of 1 if the pupil anticipates labour market discrimination, and 0 otherwise.

It is not possible to know the exact reasons why pupils might respond 'don't know' to this question. This could reflect a variety of factors, from uncertainty to fear or discomfort of

[120] As discussed in Solon et al. (2015), weighting is unnecessary for a consistent estimation of causal effects as long as the sampling is not endogenous i.e. it is independent of the dependent variable conditional on the explanatory variables, which is plausible in this analysis.

responding 'yes', or even not understanding the question. For the purposes of this chapter, I combine 'yes' and 'don't know' responses and compare this with 'no' responses, which makes the assumption that pupils who do not expect discrimination at all respond with a clear 'no'. This assumption is supported by the fact that responses of 'yes' and 'don't know' show very similar effects on GCSE performance (see results presented in Section 4.5.5).[121]

Standard set of controls

The standard set of controls used in the more parsimonious specification of my model comprises of a series of pre-determined characteristics obtained from wave 1 of Next Steps, and motivated by the literature on key drivers of educational achievement discussed previously in subsection 4.3.2. The standard set of controls includes:

- pupil's socio-demographic characteristics: gender, ethnicity, region of residence, Special Education Needs, whether speaks English as first/main language (EFL);
- family characteristics: number of siblings, whether lives in a two-parent household, home ownership, eligibility for free school meals (FSM);
- parental background: parent(s) highest education levels, parent(s) employment status, whether parent(s) work in managerial or professional occupations, whether parent(s) are aged < 40 (the median in the sample), whether parent(s) report to be in good health, whether household income is higher than £21k (the median in the sample); and
- school fixed effects, to control for school quality, peer effects and unobserved differences between students in different schools.[122]

[121] Combining these categories is also helpful to comply with the secure data service non-disclosure restrictions due to small sample sizes in some ethnic groups.

[122] School selection is also associated with socio-economic status and neighbourhood effects (Crawford et al., 2017; Wilson et al., 2011), so school fixed effects may also capture some of these factors.

Extended set of controls

After including the standard set of controls in the baseline model, the second specification includes an extended set of controls that aim to capture a series of factors that could potentially be a source of omitted variable bias in the more parsimonious regression specification. The extended set of controls includes all the variables from the standard set of controls, and also those listed below (see Appendix C for a more detailed description). As with the standard set of controls, these variables are also based on wave 1 of Next Steps. Therefore, these characteristics are observed prior to GCSE results and anticipated discrimination responses and are plausibly exogenous to anticipated discrimination. The extended set of controls includes, in addition to the standard controls:

- parental aspirations and expectations: parent would like child to continue in full-time education (FTE), parent thinks it is likely their child will go to university;
- pupil's future orientation and career perspectives: future-thinking, family orientated, has career ambition, important factors for a job; and
- past bullying experiences: pupil suffered bullying in the past year.

Parental Aspirations and Expectations: A natural concern is that parental educational aspirations and expectations for their children also affect how their children form expectations of discrimination in the labour market. Family and social capital and the parent-child relationship play an important role in shaping children's aspirations, career choices and the probability of getting post-compulsory education (Khattab, 2015). Because of the early influence of parent's values, beliefs and aspirations for their children, it is likely that parental beliefs will also influence the formation of pupils' expectations of facing future discrimination in the labour market. Parental (and pupil) educational aspirations and expectations are also shown to be positively associated with educational achievement (Chowdry et al., 2009, 2011). In the interview, the main parent is asked: *"What would you yourself like [child name] to do when* *[he/she] reaches 16 and can leave school?"* and *"How likely is it do you think that [child name] will go on to university to do a degree at some time in the future?"*. I use these two questions to capture whether parents would like their child to continue in post-compulsory education and whether parents think it is likely that their child will go to university. In addition to this, parents are asked how involved they feel in their child's school life, which can be interpreted as a measure of the family's engagement with education and of the closeness of the parent-child relationship, usually regarded as a positive influence for pupils' achievement. I use this question to create a binary variable that takes a value of 1 if the parent is very involved and 0 otherwise.

Even though pupil's own educational aspirations for post-compulsory education and their expectations of attending university are also available in the data, they are not included in my extended model because these beliefs might be influenced by anticipated discrimination. This could be the case if ethnic minority pupils adapt or modify their aspirations and expectations because of the belief that they might face an ethnic penalty labour market. In this case, aspirations and expectations of attending university would be a channel through which anticipating discrimination influences educational attainment.

Future Orientation and Career Perspectives: An additional concern is that personality traits that could be linked to anticipating discrimination. Personality traits shape an individual's outlook in life and view of the world and therefore may influence the formation of expectations of future discrimination. Unfortunately, Next Steps does not explicitly include information on the "Big Five" personality traits, which are found to influence educational outcomes (Almlund et al., 2011). However, it does include questions that can be used as a proxy to capture some of these non-cognitive traits. Having a future time perspective is a trait related to consciousness and it can be a motivator of current behaviour. For example, research in developmental psychology shows that future-oriented pupils are more academically motivated and this can translate to better educational outcomes (Greene and DeBacker, 2004). At the same time, future orientation is likely to be associated with the formation of expectations of discrimination, insofar as a teenager who thinks about their future, may have thought about the likelihood of facing discrimination in the labour market when evaluating their labour market prospects. In Next Steps, pupils are asked whether they agree or not with the following statements: *"Thinking about the future, I don't really think much about what I might be doing in a few years time"*, which I interpret as being an indicator of future orientation if pupils disagree or strongly disagree with the statement; and *"Raising a family in the future is important to me"*, which I interpret as a measure of family orientation if the pupil agrees/ strongly agrees with it. In addition, pupils are asked whether they have any ideas about the job they want to do after full-time education, and to reflect on the importance of a number of factors for their future job: *"To have a job where I help other people"*, *"To have a job which pays well"*, *"To be my own boss/have my own business"*, *"To have a job that's interesting and not routine"*, *"To have a job where I can get promoted and get ahead"*, *"To have a job with regular hours"*. These questions are used to create a series of binary variables that capture pupils' individual preferences with respect to future work prospects and that take a value of 1 if the respective job characteristic is considered very important, and 0 otherwise.

Past Bullying Experiences: It is possible that past discriminatory experiences also play a relevant role in the formation of expectations of future discrimination. Bullying is a form of oppression, of psychological or physical nature that manifests in many ways (Farrington, 1993). A number of studies have shown that victims of bullying can experience severe consequences in terms of mental health, attitudes towards school and educational achievement (Eriksen et al., 2014; Gorman et al., 2019) and therefore it is important to consider that it might influence anticipated discrimination. If a pupil is bullied because of their skin colour or ethnic origin, they may be more likely to expect to be discriminated against in the labour market on the same grounds. I create a binary variable to indicate that a pupil has not been bullied in the 12 months prior to the wave 1 interview, which is equal to 1 when the pupil has not experienced name calling, exclusion from a group, being made to hand over money or possessions, threatened with being hit, being hit or attacked, and 0 otherwise.

4.4.2 Descriptive statistics

When asked whether they think that skin colour, ethnic origin or religion will make it more difficult to get a job after they leave education, 62.6% of all ethnic minority pupils respond 'no', while the remaining 15.4% and 22% respond 'yes' and 'don't know', respectively.[123] This is further explored in Table 4.3, which displays the share of pupils within each ethnic group that anticipate discrimination. It is notable that pupils from certain groups are much more likely to anticipate labour market discrimination. The fraction of pupils in the Black Caribbean (50.9%) and Black African (52.4%) groups who expect to face discrimination based on skin colour and ethnic origin when looking for a job in the future is particularly large, while this share is the lowest for pupils in the Indian group, at 29.5%. Examining the proportions of 'yes' and 'don't know' responses separately, Table C.1 reveals that the relative proportions vary between these groups. In the Black Caribbean and Black African groups, most pupils who anticipate discrimination respond 'yes', while pupils in the Indian group mostly respond 'don't know'.

Table 4.3 Anticipation of labour market discrimination by ethnic minority group (%)

	Ethnic Minority Group						
Anticipate discrimination (%)	Indian	Pakistani	Bangladeshi	Black Caribbean	Black African	Mixed/ Other	Total
No	70.5	61.4	62.3	49.1	47.6	67.5	62.6
Yes/Don't Know	29.5	38.6	37.7	50.9	52.4	32.5	37.4
Total	100	100	100	100	100	100	100

Notes: Survey weights are applied using the weights provided in wave 2. Source: Next Steps.

Given that the main objective of this chapter is to analyse whether, and to what extent, anticipated discrimination influences educational performance, it is useful to analyse the raw gaps in key GCSE outcomes by anticipated discrimination. Table 4.4 presents the raw gaps in the different measures of educational achievement according to anticipated discrimination for the pooled sample of all ethnic minority students. In general, pupils who anticipate discrimination achieve better GCSE results on average than those who do not, however most

[123] I examine the implications of pooling these two response categories, and provide the results obtained when considering them separately later in the chapter.

of these differences fall short of significance (unweighted differences in means are reported in Table C.3). The only exception is in the proportion of pupils achieving any A*-A, which is marginally higher among those who do not expect discrimination. Notably, the largest raw gap is observed in the proportion of pupils achieving A*-C in GCSE English (of 4.51%), significant at 10%.

Table 4.4 Raw gaps in GCSE performance by anticipated labour market discrimination

		Anticipate Discrimination	
Educational Outcomes	All	Yes/Don't Know	No
Total GCSE points score	380.736	382.901	379.441
Total GCSE score (Std)	-0.063	-0.048	-0.071
Achieved 5 or more GCSEs (proportion)	0.592	0.609	0.581
Achieved 5 or more GCSEs incl. Eng, Maths (proportion)	0.465	0.480	0.456
Achieved A*-C in GCSE English (proportion)	0.611	0.639	0.594**
Achieved A*-C in GCSE Maths (proportion)	0.543	0.555	0.537
Science score (Std)	-0.041	-0.019	-0.054
Achieved any A*-A (proportion)	0.370	0.360	0.376
Unweighted count	3315	1283	2032

Notes: Survey weights applied using the weights provided in wave 2. Unweighted statistics are presented in Table C.3 in the Appendix. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. For the t-test of differences in means between 'Yes/Don't Know' and 'No':*$p < 0.10$,**$p < 0.05$, ***$p < 0.01$. Source: Next Steps.

Table C.4 in Appendix C presents the sample means for the main variables used in the analysis for the pooled sample of ethnic minority pupils, and splitting the sample by anticipated labour market discrimination. This provides evidence of the ethnic composition of the total sample and of how this differs by anticipated discrimination in the sub-samples. This reflects the ethnic composition of the population. Mixed/Other, Indian and Pakistani are the biggest groups in the sample, representing over 70% of all ethnic minority pupils.

These descriptive statistics suggest that the composition of the sub-samples based on anticipated discrimination is surprisingly similar with respect to the other characteristics analysed. For example, around half of the pupils in my total sample are girls, and no significant difference is observed in the gender composition of the sub-samples by anticipated discrimination. Moreover, no significant differences are observed with regard to the fraction of pupils who speak English as a first language or the fraction of pupils with no special education needs (SEN), which represent the majority of pupils in both sub-samples. There is also little evidence of significant differences in the regional composition of the sub-samples by anticipated discrimination, almost half of the ethnic minority pupils reside in London. However, a slightly smaller proportion of pupils who anticipate discrimination reside in the East Midlands.

In the total sample, most pupils (70%) live with both parents and are not eligible (73%) for free school meals (FSM), and 61% of families are homeowners. With regard to parental background, most pupils (59%) have low/unqualified parents, and 36% of families have household income over £21k. Interestingly, the vast majority of parents (94%) would like their child to continue in full time education after the age of 16, while half of the parents (56%) expect their child to attend university. Only a third of parents report to be very involved in school life (33%). Pupils who anticipate discrimination are slightly less likely to come from a two-parent and homeowner families, but these differences are not statistically significant.

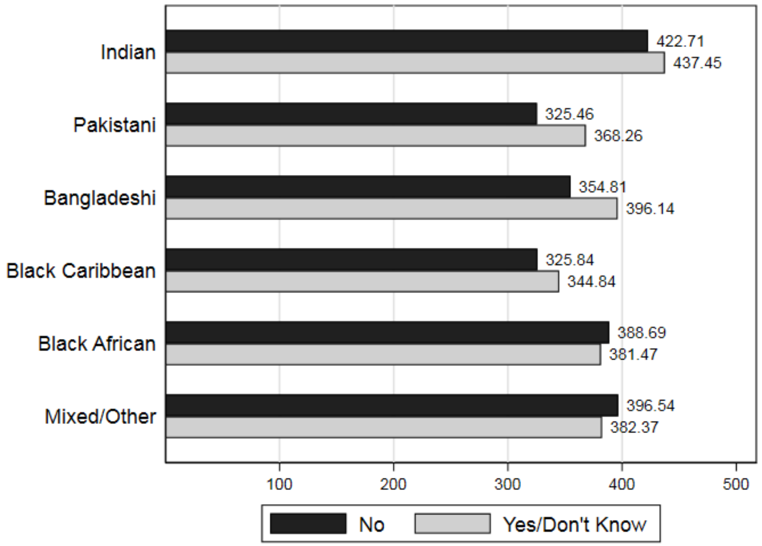
The composition of the sub-samples with respect to family and parental characteristics is strikingly similar, as with the cases of personality traits, future orientation and job perspectives. Most pupils are future oriented (70%), and the vast majority (90%) considers raising a family important. In terms of what is important for a job, most pupils (76%) believe that having a well-paying job is very important and around 70% consider that an interesting job or a job with promotion prospects are very important. Only 35% of pupils consider that being your own boss is very important.

Interestingly, pupils who anticipate discrimination are more likely to have suffered with prior bullying experiences. This is in line with what would be expected, as prior experiences of being bullied might influence one's expectations of being treated differently in the future. Finally, pupils who anticipate discrimination have higher prior achievement at ages 10-11, measured by KS2 total points scores, although this difference is not statistically significant.[124]

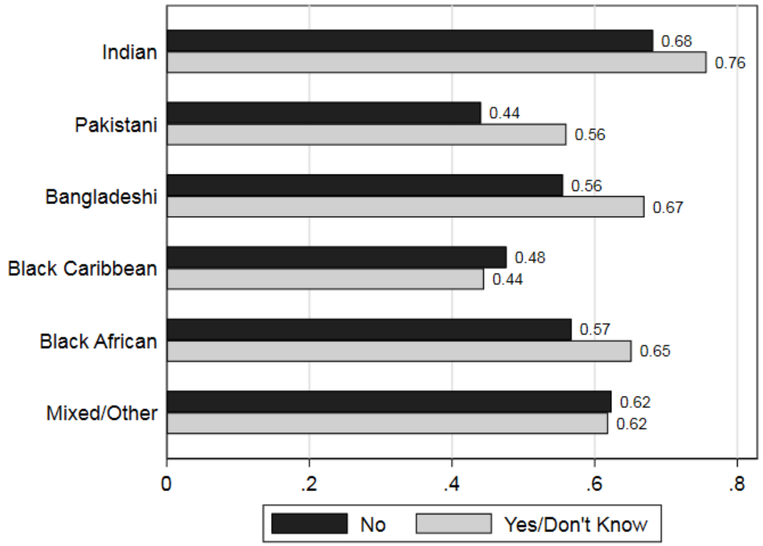
[124]However, it is worth noting that this is the only explanatory variable for which the differences in means become highly statistically significant when examining the unweighted statistics. Nevertheless, this could also be expected a priori, if anticipated discrimination also influences the efforts pupils exert in school already at younger ages.

When considering the specific ethnic minority groups separately, the raw gaps in GCSE achievement by anticipated discrimination become more evident. Moreover, it is notable that the magnitude and direction of these raw gaps vary considerably by ethnicity. Figures 4.2 and 4.3 provide a visual representation of these raw gaps by anticipation of discrimination for various educational outcomes and by ethnicity. On average, pupils from Indian origin are the highest achievers across all measures. By contrast, pupils from the Black Caribbean group are the lowest achievers. . Besides these differences in average achievement between ethnic groups, these numbers also reveal that anticipating discrimination is associated with different levels of achievement within ethnic groups.

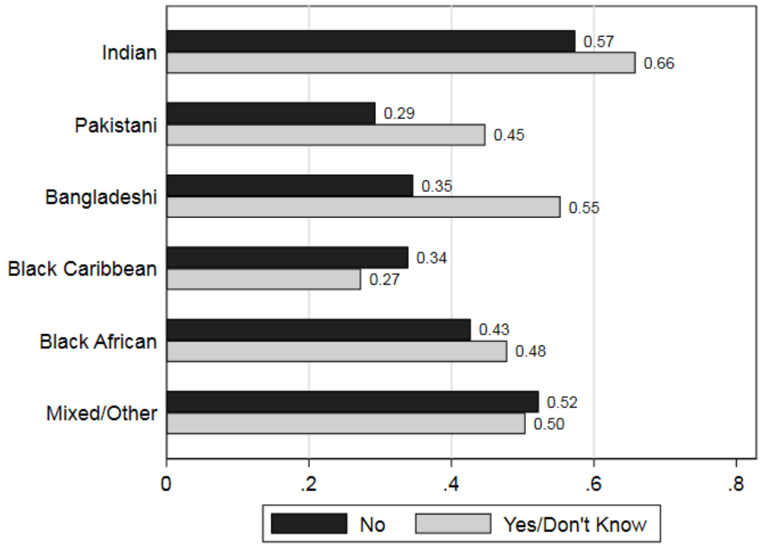
Reflecting the pattern observed for the pooled sample, for most ethnic groups and GCSE outcomes, pupils who anticipate labour market discrimination tend to have better educational attainment on average. These within-group raw gaps by anticipated discrimination are particularly large among pupils of Pakistani and Bangladeshi origin. For example, as shown in Figure 4.2c, the share of pupils who achieve 5 or more A*-C including English and Maths is 16 percentage points higher among pupils of Pakistani origin who anticipate discrimination than those who do not. For pupils of Bangladeshi origin, this raw gap is even larger, at 20 pp. Another very large gap is also observed among pupils of Bangladeshi origin in achieving an A*-C in Maths, 17 pp (Figure 4.3b). Although these figures reveal that pupils who anticipate discrimination tend to have better outcomes on average, some negative gaps are also observed. Figure 4.3a reveals that the share of pupils of Black Caribbean origin who achieve an A*-C in Maths is 6 pp higher among those of this group who do not expect discrimination. Pupils of Black African origin who do not expect discrimination also achieve slightly higher GCSE total points score. Smaller negative gaps are observed for pupils of Mixed/Other origin across many educational outcomes, although this category is an aggregation of quite heterogeneous subgroups but whose sample sizes are too small to be analysed separately.



(a) Total GCSE points score (Mean)

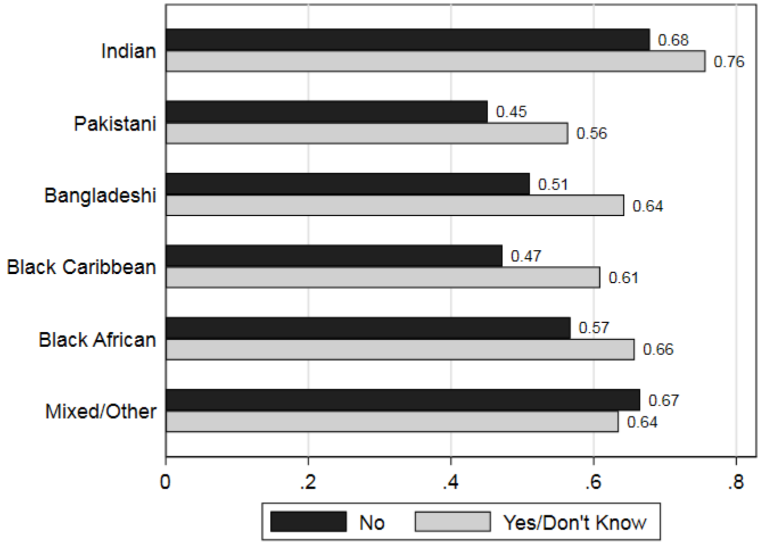


(b) Pupils achieving 5+ A*-C GCSEs (proportion)

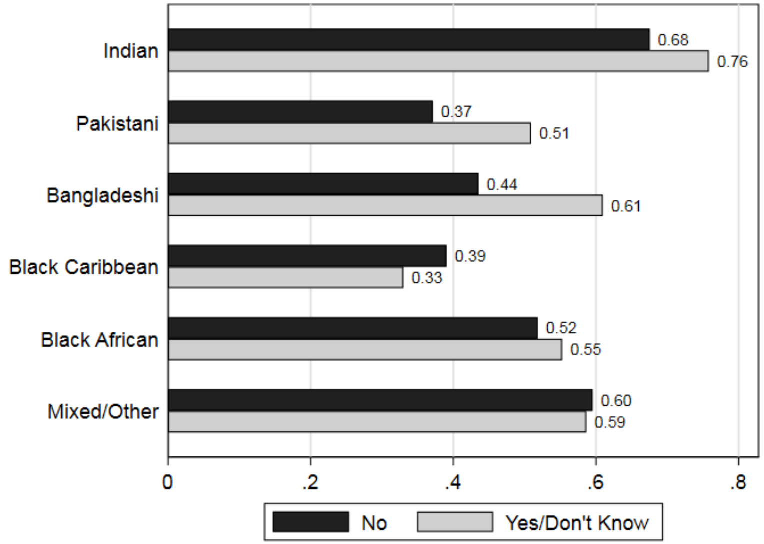


(c) Pupils achieving 5+ A*-C GCSEs incl. English and Maths (proportion)

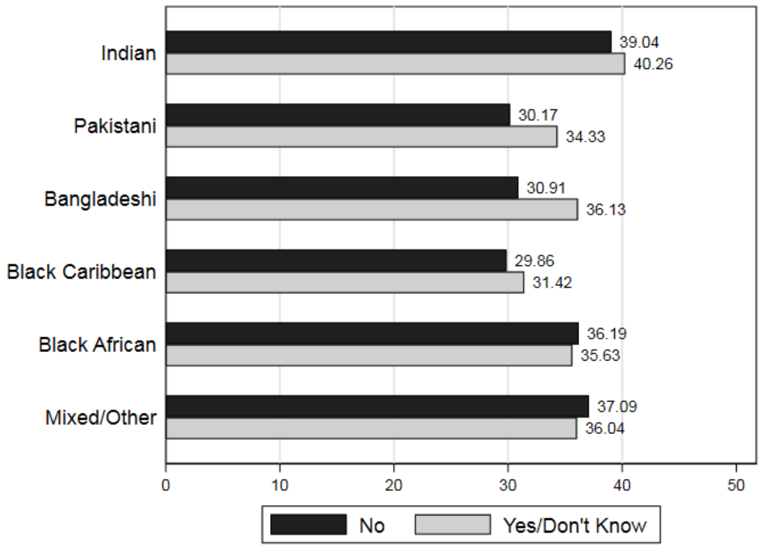
Figure 4.2 Educational achievement by ethnicity and expectation of discrimination in the labour market



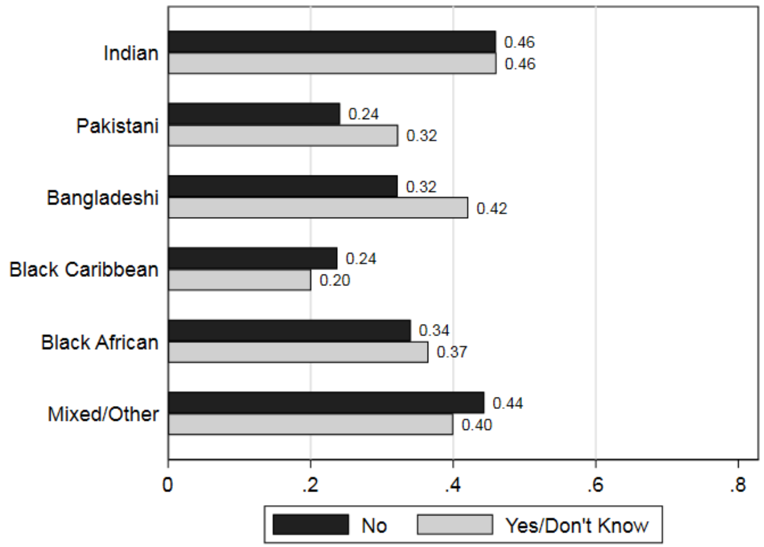
(a) Pupils achieving A*-C in English (proportion)



(b) Pupils achieving A*-C in Maths (proportion)



(c) Highest Science Score (Mean)



(d) Pupils achieving any A*-A (proportion)

Figure 4.3 Educational achievement by ethnicity and expectation of discrimination in the labour market (cont.)

Complete descriptive statistics, presented separately for each ethnic group, are also reported in Appendix C. Similarly to the pooled sample, for most ethnic groups the composition of the sub-samples by anticipated discrimination is largely similar with respect to the main characteristics analysed. Some small differences are observed, with the caveat that it is more difficult to assess the statistical significance of these differences in sample means due to the smaller sample sizes.

4.4.3 Estimation strategy

In the education production function, educational outcomes are determined as a function of various cumulative inputs, including individual, school and parental characteristics (Hanushek, 2010). Educational achievement of individual i, for example test scores, can be represented by the following function:

$$EducAchievement_{it} = f(X_i(t), P_i(t), S_i(t), \alpha_{i_0}, \varepsilon_{it}) \tag{4.1}$$

where $X_i(t), P_i(t), S_i(t)$ are sets of individual, parental and school characteristics, α_{i_0} is a measure of the individual's initial ability endowment and ε_{it} is a random error term.

In this chapter, the main outcome of interest is the educational performance at the end of compulsory schooling in England, measured by pupils' achievement in GCSE exams. I investigate whether educational outcomes are influenced by expectations of facing future discrimination in the labour market, controlling for a large number of individual, family and parental characteristics, as well as school fixed effects. To begin with, I employ a standard least squares approach to the pooled sample of all ethnic minority pupils. The linear reduced-form model can be written as:

$$EducAchievement_i = \beta_0 + \beta_1\, AnticipateDisc_i + X_i'\beta_2 + P_i'\beta_3 + \eta_s + u_i \tag{4.2}$$

where X_i is a vector of individual characteristics (including the ethnicity dummy), P_i captures parental and family characteristics, η_s captures secondary school fixed effects. β_1 is the main coefficient of interest, from the anticipated labour market discrimination variable. When the educational outcome is a dichotomous variable, I estimate the same model using a Linear Probability Model (LPM).[125]

I estimate the above regression equation (4.2) and examine the relationship between anticipated discrimination and educational achievement by progressively adding covariates. In the first, more parsimonious specification, I include the standard set of controls for individual and family background characteristics. In the second specification, I include the extended set of controls. For all regressions, standard errors are clustered at the secondary school level.

Methodological challenges: Endogeneity and the omitted variable bias

OLS and LPM estimates are biased if the reduced-form model is incomplete - that is, when the shorter regression fails to account for other relevant variables that should be present in the full model and when these omitted parts are correlated with the variable of interest - resulting in omitted variables bias (OVB).[126] In order to address the potential OVB, I estimate two versions of the main model, one more parsimonious than the other. The more parsimonious specification includes a standard set of controls, arguably pre-determined variables, and which have been established in the literature as important explanatory factors of educational achievement. With the aim of reducing the potential OVB that could still affect this baseline model, I then include a more extensive set of controls in the second specification of the model. Here I capture additional relevant factors that may have been omitted in the first specification, including future orientation, the importance of career and family and parental aspirations and expectations.

[125] Using a binary probit to model the binary outcome variables does not affect my main findings.

[126] In the words used by Angrist and Pischke (2008), a longer regression with more controls reflecting the complete model has a causal interpretation, while the coefficients from a short regression i.e. reduced-form model might be biased whenever the omitted variables and included variables are correlated.

The process of deciding whether the addition of variables to the model will help getting estimates closer to the parameter of interest is not straightforward in a multivariate regression context, and this approach should be used with caution. Adding more controls is not always better, sometimes the inclusion of certain variables to the model can produce even more biased estimates, acting as a confounding factor in the estimation of the desired effect. It is therefore crucial to distinguish between good controls and bad controls. Good controls are variables that would reduce the potential OVB in the coefficient of interest and make the conditional independence assumption (CIA) more plausible (Angrist and Pischke, 2008).[127] In contrast, bad controls are variables that may amplify the OVB. Variables that could be, for example, a mediator (or even a partial mediator) of the effect of anticipating discrimination on educational achievement, e.g. effort exertion, or an outcome variable from anticipating discrimination such as low self-esteem. The issue with conditioning on bad controls is that they can cause a selection bias, and thus any comparisons of educational outcomes conditional on a bad control would no longer be direct, i.e. 'apples-to-apples' (Angrist and Pischke, 2008). This stems from the fact that anticipating discrimination would change the composition of the sample with respect to this variable. Ultimately, this means that such comparison would not have a causal interpretation.

In line with this definition, all controls in the standard set of controls can be regarded as good controls, as they reflect socio-demographic characteristics, and family and parental background that are pre-determined to the individual and therefore exogenous to anticipating discrimination. In addition, the variables added in the extended set of controls, including some personality-related characteristics, opinions and parents' aspirations have been chosen with careful consideration to avoid any confounding factors. Further, these controls are also

[127] As some more recent research points out, even the inclusion of some truly omitted variables (i.e. good controls) might not necessarily reduce the magnitude of the OVB if we move from a short regression to a long regression and where both are subject to misspecification, that is, if the longer estimated model continues to omit some relevant variables (De Luca et al., 2018). Here, good controls are covariates that are arguably pre-determined (fixed) or exogenous to the 'treatment' i.e. to anticipating discrimination.

likely to be fixed at the time anticipated discrimination is observed (and determined), which also limits potential issue of reverse causality.[128]

Exploring the stability of parameters of interest by varying the inclusion of control variables in different specifications of the reduced-form model is a common approach used in the literature to address the OVB problem. If the estimated coefficients remain stable across specifications after the inclusion of observed controls, this is considered a sign of a limited OVB. However, it is important to acknowledge that it is impossible to guarantee a complete elimination of the omitted variable bias arising from unobservables in the absence of a randomised experiment, and there is a limit to establishing causality in this manner. As emphasised by Oster (2019), the usual practice of discussing coefficient movements after the inclusion of observed controls fails to take into account how much of the variance in the outcome is explained by the added controls (i.e. R-squared movements), and rests on the assumption that the bias arising from observed controls is informative of the bias arising from omitted unobservables. Taking this issue into account, I explore remaining biases and examine the robustness of my main results to a possible selection on omitted unobserved variables using the method proposed by Oster (2019). This method examines the extent of the bias that unobserved omitted variables would cause, under certain assumptions. I provide a more detailed description of the method and how I apply it to the data alongside the results in subsection 4.5.3.

[128] While I cannot rule out this possibility completely, it is unlikely that this would be an issue when focusing on GCSE scores. Since the expectations of future discrimination were present and measured before sitting these exams, this temporal difference makes it improbable that the GCSE results themselves would have affected the formation of these expectations.

4.5 Results

4.5.1 Anticipated discrimination and GCSE results

I first explore the relationship between anticipated discrimination and educational performance for the pooled sample of ethnic minority pupils. Table 4.5 displays my summarised findings. Here, I report results from three specifications estimated using Linear Probability Model (LPM) when the dependent variable is dichotomous and Ordinary Least Squares (OLS) for continuous dependent variables. Model 0 presents the raw gaps by anticipated discrimination with no controls. Model 1 reports the results controlling for the standard set of controls, which includes several individual socio-demographic characteristics, family and parental background and school fixed effects. Model 2 includes the extended set of controls, which includes the standard controls plus thirteen additional variables related to parental expectations and aspirations, pupil future orientation and career perspectives.

The results from Model 1 suggest that anticipating discrimination in the labour market generally increases GCSE performance. For example, the probability of achieving 5 or more A*-C GCSEs including English and Maths, the gold standard, increases by 6.2 percentage points (pp) when pupils anticipate discrimination, while this effect is slightly higher at 6.4 pp for A*-C in Maths only and somewhat lower at 4.7 pp for A*-C in English only. Anticipated discrimination is also associated with a higher Science score by 0.11 standard deviations. The one exception is achievement at the top of the performance distribution i.e. achieving A*-A grades. This suggests that anticipating labour market discrimination generally improves GCSE outcomes of ethnic minority pupils, but perhaps this effect is not so important at the very top of the distribution. Adding the standard set of controls increases the educational gap related to expecting labour market discrimination. Ethnicity explains a large share of the observed anticipated discrimination gap in attainment because there are considerable differences in the proportion of pupils who anticipate discrimination across ethnic groups. For example, pupils of Black African and Black Caribbean origin are much more likely to expect discrimination, but they have, in general, lower educational attainment. While ethnicity is likely to be one of the variables predominantly influencing this gap, other background characteristics also play a role and could explain part of the difference between raw gaps and adjusted gaps. One example is gender, girls are on average less likely to anticipate discrimination and have higher achievement.

Results from the more complete specification, Model 2, also suggest that anticipating discrimination is associated with higher GCSE performance, and controlling for the extended set of controls leads to slightly larger coefficients. For example, the probability achieving the gold standard is 6.4 pp higher among pupils who anticipate discrimination. For the pooled sample of all ethnic minorities, adding the additional covariates slightly increased the educational gap related to anticipated discrimination further. This would be expected if the added characteristics such as future orientation are on average more observed (i.e. higher proportion) for pupils who do not anticipate discrimination. This could be the case if pupils who think about the future frequently (future-orientated) are less likely to attribute any expected difficulties to discrimination. Another example is bullying - if the absence of bullying experiences in the past is related to higher educational outcomes and also to pupils being less likely to anticipate discrimination, the inclusion of this variable in the model is expected to increase the gap.

The coefficient of anticipated discrimination is only slightly changed by the inclusion of the school fixed effects, as shown by the results displayed in Table C.11, suggesting that the relationship between anticipated discrimination and educational performance is not primarily due unobserved characteristics at the school-level, such as school-level policies, or neighbourhood effects. Including school fixed effects compares outcomes across students with different expectations that attend the same school and also suggests that those who anticipate discrimination have higher attainment.

Table 4.5 Anticipated discrimination and GCSE achievement - summarised results (OLS and LPM)

	Educational Outcomes						
	Total GCSE score	5+ A*-C	5+ A*-C Eng,Math	A*-C Eng	A*-C Math	Science score	Any A*-A
Model 0: Raw Gaps							
Anticipate Disc	0.027	0.033*	0.044**	0.045**	0.038**	0.058	0.006
	(0.037)	(0.018)	(0.017)	(0.018)	(0.017)	(0.037)	(0.018)
Model 1: Baseline Model							
Anticipate Disc	0.078**	0.050**	0.062***	0.047**	0.064***	0.110***	0.032
	(0.038)	(0.020)	(0.019)	(0.019)	(0.018)	(0.039)	(0.020)
Standard controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Model 2: Extended Model							
Anticipate Disc	0.084**	0.054***	0.064***	0.050***	0.066***	0.116***	0.030
	(0.036)	(0.019)	(0.018)	(0.018)	(0.017)	(0.037)	(0.019)
Standard controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Extended controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	3315	3315	3315	3315	3315	3315	3315

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and controls are described in the text in section 4.4.1. Pooled sample of ethnic minority students. Source: Next Steps.

Complete results for all control variables included in Models 1 and 2 are reported in Tables C.12 and C.13 respectively. These results provide evidence in line with other research on the determinants of educational achievement, now with a special focus on the performance of ethnic minority pupils. Being a girl is positively associated with higher attainment across most outcomes, except in A*-C in Maths. This role of gender has been described previously using UK data by Strand et al. (2006) and Demie (2001), who also further suggest that it has a uniform effect across different ethnic groups. Pupils of Pakistani and Black African origin obtain considerably worse GCSE outcomes across all measures, and those of Black Caribbean heritage have significantly worse attainment in some outcomes. For pupils from the Bangladeshi and Indian groups, the differences relative to Mixed/Other are smaller and not significant. Not having any Special Education Needs is also significantly associated with better educational attainment across all measures, as expected. Finally, speaking English as a first language does not appear as a significant factor, after all else is controlled for.

A number of indicators of socio-economic status and parental background also play an important role, as expected. My results are in line with previous findings by other studies on the importance of social class on educational achievement of ethnic minorities at the age of 16 (see for example Strand, 2014 and Strand et al., 2015). Having more siblings is associated with worse outcomes across all measures. On the other hand, being in a two-parent household, and owning a home, higher levels of parental education, parents who work in professional/managerial occupations and higher household income also predict higher GCSE attainment.

Among the additional variables in Model 2, parental aspirations that the pupil will continue in full-time education (FTE) post 16 and parental expectations that the child will attend university are highly associated with better GCSE outcomes. Conversely, parental involvement does not seem very important. As expected, the absence of past bullying experiences is associated with better outcomes and future orientation has a positive coefficient for most outcomes. Adding these new variables provided a modest boost to the explanatory power compared to Model 1, raising the percentage of the variance explained by at least 5pp for all outcomes. Moreover, including these factors slightly reduces the overall influence of gender, ethnicity, SEN, and of most indicators of socio-economic status in the standard set of controls.

The key finding here, is that educational performance at the GCSE level is positively and strongly linked to anticipated discrimination for ethnic minority pupils. The association between anticipating discrimination and higher attainment at the GCSE level was positive and significant for most educational outcomes analysed after controlling for a large number of pupil characteristics. These characteristics included socio-demographics, family and parental background and school fixed effects. This positive result suggests that, ceteris paribus, ethnic minorities who anticipate discrimination are incentivised to attain better results in these high stake exams. Henceforth, it is likely that they invest more in acquiring skills and abilities to counteract the expected ethnic penalty in the labour market.

This is a striking empirical finding that can provide new insights into how discrimination may influence individual's decisions as part of a pre-labour market process. Theoretical models of discrimination consider the acquisition of human capital an endogenous choice, which means that labour market discrimination can influence individual's decisions of investing in human capital even prior to entry in the market. These models predict that the effects of discrimination could vary in direction. In earlier models, when investments in human capital are assumed to be not directly observable, employers must rely on a noisy productivity signal or on their prior beliefs to assess a potential worker's productivity (Coate and Loury, 1993*b*; Lundberg and Startz, 1983). These models predict that in the presence of labour market discrimination, minorities will invest less in human capital and become the low-wage/low-skilled group. More recent models, on the other hand, have described that when at least some aspect of human capital investment is directly observable by the market, minority workers will have an incentive to signal their ability using education (Arcidiacono et al., 2010; Lang and Manove, 2011).

Due to the paucity of data on expectations of discrimination, little was previously known about how these expectations may influence individuals' productive characteristics. Linking my results to this theoretical literature on discrimination, I put forth evidence more in line with the predictions made by Lang and Manove (2011) and Arcidiacono et al. (2010). When investments in education are directly observable, as in the case of the cognitive skills demonstrated by GCSE results, individuals from minority groups who expect discrimination might have an incentive to over-invest in education as a way of signalling their ability. I identify that minority pupils who anticipate discrimination do achieve on average higher GCSE results, even comparing between pupils from the same ethnicity and in the same schools. This indicates that they recognise the importance of these exams to future outcomes and that achieving high results could be part of a signalling strategy. This signalling could manifest as either signalling their ability to future employers in the labour market or improving their chances to continue into higher education, given that a degree could be an even more valuable signal in the graduate labour market.

Interestingly, my findings suggest that discrimination plays a role at an earlier stage than previously thought, already acting during compulsory schooling and before important decisions related to continuation in post-compulsory education are made. This finding, that investments in education are made by ethnic minorities as part of a compensatory strategy supports the findings of previous research. In the US, ethnic minority individuals invest more in human capital through additional years of education (Carneiro et al., 2005; Lang and Manove, 2011) and by obtaining a college degree (Arcidiacono et al., 2010), conditional on ability. In the UK, this pattern is reflected in increased participation in postgraduate education after first degree (Lessard-Phillips et al., 2018) and educational qualifications of minorities (Blackaby et al., 2002; Li, 2018).

4.5.2 Ethnic differences

This section examines whether there are differences in the effect of anticipating discrimination across ethnic groups. Considering that the main focus of this analysis is the expectation of being discriminated against in the labour market based on ethnicity and skin colour, one important aspect of this study was to examine a possible heterogeneity by ethnicity. From the literature discussed in section 4.3, there is extensive evidence on differential labour market outcomes and educational attainment between different groups. In addition, from the descriptive statistics, it is notable that the share of pupils who anticipate discrimination varies across ethnic groups. To shed further light on how anticipating discrimination matters, by examining potential heterogeneities in this effect, I estimate separate regressions for each ethnic group, allowing the effect of all control variables to vary by group.

The respective results by ethnicity are displayed in Table 4.6 for Model 1, including the standard set of controls, and in Table 4.7 including extended set of controls. The different columns report the estimates for pupils of each ethnic group separately. These results show that the relationship between anticipated discrimination and educational attainment is more pronounced for pupils from certain ethnic groups. For example, the results from Model 1 (Table 4.6) suggest that for pupils of Pakistani origin, anticipating discrimination is associated with a 16.1 pp increase in the probability of achieving five or more A*-C including English and Maths, and a 15.1 pp increase in the likelihood of achieving an A*-C in Maths. When considering the results from Model 2, with the extended set of controls, the results are similar, at 15.2 pp and 14.9 pp, respectively. For this group, anticipating discrimination appears to matter across the board, with increases in the total GCSE point score and the probability of achieving the highest grades. It is also evident that pupils from other South Asian backgrounds, such as those of Bangladeshi and Indian origin perform better in the GCSEs when expecting discrimination in the labour market.

For other ethnic groups, such as Black Caribbean and Black African, this association between anticipated discrimination and attainment is much weaker in terms of magnitude and statistical precision. Interestingly, however, with the exception of the coefficient for the probability of achieving an A*-C in English, all the estimated coefficients for the Black African group indicate a negative relationship between anticipated discrimination and educational attainment (although statistically insignificant). Unfortunately, a lack of statistical precision in some of these smaller groups precludes drawing firm conclusions on differences between ethnic groups.

Table 4.6 Anticipated discrimination and GCSE achievement, by ethnic group – Model 1 (OLS and LPM)

	Ethnic Group					
Dependent Variable	Indian	Pakistani	Bangladeshi	Black Caribbean	Black African	Mixed/Other
Total GCSE score						
Anticipate Disc	0.043	0.254**	0.250*	0.042	-0.065	-0.003
	(0.096)	(0.105)	(0.141)	(0.187)	(0.208)	(0.124)
5+ A-C*						
Anticipate Disc	0.074	0.134**	0.083	-0.030	0.059	0.046
	(0.048)	(0.053)	(0.063)	(0.098)	(0.110)	(0.068)
5+ A-C incl. Eng, Mat*						
Anticipate Disc	0.069	0.161***	0.157**	-0.044	0.012	0.074
	(0.053)	(0.052)	(0.068)	(0.089)	(0.114)	(0.069)
A-C Eng*						
Anticipate Disc	0.097*	0.106**	0.088	0.135	0.015	0.003
	(0.051)	(0.050)	(0.060)	(0.101)	(0.112)	(0.061)
A-C Math*						
Anticipate Disc	0.059	0.151***	0.126**	-0.023	0.038	0.050
	(0.047)	(0.053)	(0.060)	(0.094)	(0.105)	(0.067)
Science score						
Anticipate Disc	0.073	0.274**	0.355**	-0.011	0.040	0.001
	(0.091)	(0.111)	(0.152)	(0.181)	(0.220)	(0.127)
Any A-A*						
Anticipate Disc	0.019	0.108**	0.121*	-0.005	0.026	0.016
	(0.060)	(0.062)	(0.047)	(0.069)	(0.092)	(0.109)
N	712	676	477	366	289	795

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. Model 1 includes the standard set of controls. Source: Next Steps

Table 4.7 Anticipated discrimination and GCSE achievement, by ethnic group – Model 2 (OLS and LPM)

	Ethnic Group					
Dependent Variable	Indian	Pakistani	Bangladeshi	Black Caribbean	Black African	Mixed/Other
Total GCSE score						
Anticipate Disc	0.065	0.244**	0.223*	-0.003	-0.026	0.042
	(0.100)	(0.096)	(0.133)	(0.186)	(0.201)	(0.121)
5+ A-C*						
Anticipate Disc	0.091*	0.129**	0.064	-0.056	0.057	0.067
	(0.048)	(0.050)	(0.065)	(0.0101)	(0.106)	(0.069)
5+ A-C incl. Eng, Mat*						
Anticipate Disc	0.081	0.152***	0.143**	-0.038	-0.001	0.104
	(0.050)	(0.051)	(0.064)	(0.089)	(0.125)	(0.071)
A-C Eng*						
Anticipate Disc	0.115**	0.104**	0.068	0.138	0.014	0.025
	(0.050)	(0.048)	(0.062)	(0.115)	(0.103)	(0.062)
A-C Math*						
Anticipate Disc	0.069	0.149***	0.110*	-0.027	0.032	0.073
	(0.048)	(0.051)	(0.061)	(0.094)	(0.114)	(0.070)
Science score						
Anticipate Disc	0.102	0.288***	0.340**	-0.078	0.069	0.041
	(0.095)	(0.099)	(0.147)	(0.170)	(0.226)	(0.127)
Any A-A*						
Anticipate Disc	0.010	0.089**	0.103	0.010	0.030	0.048
	(0.060)	(0.045)	(0.072)	(0.096)	(0.117)	(0.062)
N	712	676	477	366	289	795

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. Model 2 includes the standard and the extended sets of controls. Source: Next Steps

Overall, I find that for most ethnic groups, this relationship between anticipated discrimination and educational attainment is positive and significant. This is common among students of South Asian origin, particularly those from the Pakistani, Indian and Bangladeshi groups. However, for other ethnicities, the observed association was much weaker, such as for the Black African group, and a negative relationship (although not significant) is observed for the Black Caribbean group. Surprisingly, the ethnic groups exhibiting weaker, and sometimes negative relationships between anticipated discrimination and educational attainment are proportionally more likely to anticipate discrimination (Black African and Black Caribbean).[129] These ethnic differences are interesting, and examining the determinants of educational achievement for each group could provide further insights on the origins of these ethnic differences. There is some variation in the key characteristics that explain educational outcomes for the different groups, as demonstrated by the inclusion of the controls from the extended set (parental aspirations, future orientation and career perspective and bullying), which did not result in unidirectional coefficient movement across all ethnic groups. While many previous studies consider ethnic minorities as a single group, here I present important findings of heterogeneous experiences among ethnic minorities.

In addition to the observed differences in the share of people that anticipate discrimination across ethnic groups, it is possible that the heterogeneous effects of anticipating discrimination by ethnicity relate to the type and/or severity of experienced and perceived discrimination. Ethnicity may influence both the extent to which the anticipated ethnic penalty is perceived a constraint and the perceived potential of the formal education system as a means to achieving better outcomes in the market. If pupils from certain groups perceive discrimination in a way that they think it is impossible to overcome, they may adopt a more pessimistic view of their potential in the labour market. An expectation of no pay-off from additional education due to discrimination based on the reputation attracted to a specific group i.e. a negative stereotype might reduce individual incentives to invest in education. On the other hand, if the expected

[129]They are also proportionally more likely to have responded ‘yes’ rather than ‘don’t know’ to the discrimination question.

discrimination is perceived as surmountable, individuals might adopt a compensatory strategy by over-investing in education. This interpretation suggests that the effect of anticipated discrimination is linked to the existence of a stereotype threat effect and to whether pupils believe they are able to overcome an ethnic penalty in the market or not with additional education.

Although the exact reasons for the between-group differences are unknown, one hypothesis is that they could relate to the distinct migration histories and cultures of different ethnic groups. The presence of non-white ethnic groups in the UK is relatively recent and closely linked to immigration flows that took place after the Second World War (Dustmann and Theodoropoulos, 2010). Initially, migrants predominantly arrived from the Caribbean, followed by large inflows from India, Pakistan and Bangladesh through the 1950s and 60s. Their reasons were varied: to fill shortages in the British labour market, to escape civil war, to seek better economic opportunities and to join family members already settled in the UK. It is important to acknowledge these differences in the timing and reasons for migration, in the economic and social capital brought with them, as well as keep open to the possibility that their experiences in Britain could be different too (Berthoud, 2000; Dale et al., 2002; Dustmann and Theodoropoulos, 2010). It is possible that this shapes people's self-perceptions and their understanding of social context and opportunities. Therefore, the past and present experiences of minorities may be affected by different endowments of economic, social and cultural capital and by the way that people and their ethnicities are constructed within society (Platt, 2019; Shah et al., 2010). Whilst I was not able to distinguish between pupils who were born in the UK or elsewhere, nor had information on the parents' country of birth, an investigation of differences between immigrants (even second-generation ones) and UK-born minorities would be a useful way to examine this hypothesis. This remains a limitation of the current study but it is left as a suggestion for future research.

A further confounding factor, linked to these historical and cultural differences, relates to the way different ethnic minority groups perceive formal education as the means to overcoming discrimination. For some groups, such as the British South Asian and Chinese, it has been suggested that their drive to overcoming disadvantages and achieving higher educational success lies in the families and communities getting their children to internalise high educational aspirations and to enforce appropriate behaviour (Modood, 2004). Interviews with young people of Pakistani and Bangladeshi origin reveal that their view of education is very positively associated with obtaining qualifications, as they are seen as the only route to a good job, reflecting "the extent to which discrimination in the labour market is perceived as a constraint and the recognition that, if you were Asian, you would only succeed by being better qualified than your competitors" (Dale et al., 2002, p.950). This is further supported by the findings by Shah et al. (2010) that provide evidence on the role of shared norms and values i.e. 'ethnic capital' among British Pakistani families in promoting educational achievement. On the other hand, the presence of self-fulfilling prophecies about achievement in education and in the labour market is a hypothesis commonly discussed in relation to the educational outcomes of pupils from Black Caribbean origin, particularly of boys (Berthoud, 2000). Further research is needed to understand the reasons behind variations in the formation of expectations of discrimination and in the reactions to those expectations.

4.5.3 Selection on unobservables

As explained in section 4.4.3, concerns about possible omitted variable bias (OVB) are present in this study, as in most non-experimental studies in economics. In an attempt to reduce this bias, the previous models included a rich set of controls that capture observable characteristics of pupils, their families and school fixed effects. While using both standard and extended sets of controls alleviates endogeneity concerns, remaining biases may still exist since some unobserved characteristics that affect educational outcomes might also be related to anticipated discrimination. In this section, I examine the robustness of my main results by considering the possibility of selection on unobserved variables.

The method developed by Oster (2019) is useful to assess the extent of a potential bias arising from omitting unobserved variables, by considering simultaneously the coefficient stability and the R-squared movements after the addition of observed controls. To do this, the method uses information on the correlation between the observed variables and the 'treatment' (to use the language of causal inference) i.e. anticipated discrimination. The Oster method requires an 'controlled' regression, which includes all observable factors and an 'uncontrolled' regression, which includes only covariates that are not informative of selection on unobservables. Here, the 'uncontrolled' regression includes only anticipated discrimination as an independent variable and the 'controlled' regression (specified as in equation 4.2) includes anticipated discrimination and all the observed variables from the extended set of controls.[130] In addition to this, the method also uses information on R-squared movements once controls are added. As Oster (2019) explains, the omitted variable bias is proportional to coefficient movements (i.e. less coefficient movement is a sign of more limited bias) but only if these movements are scaled by how much of the variance in the outcome is explained by the inclusion of controls. Thus, the 'controlled' and 'uncontrolled' regressions are estimated to extract the main coefficient of interest (β), for anticipated discrimination, and the R-squared for each model; $\mathring{\beta}, \mathring{R}$ (uncontrolled model) and $\tilde{\beta}$, $\tilde{R}$ (controlled model).

In order to be able to calculate the bias-adjusted estimates for anticipated discrimination, two additional key parameters are needed to specify: (i) the relationship between observables and unovservables, and (ii) the maximum amount of variation that can be explained by the model. The first parameter, δ, represents the relative degree of selection on observed and unobserved variables. There are no standard values for δ, but Oster (2019) suggests that $\delta = 1$ is an appropriate upper bound because the observed variables included in the model are usually chosen based on the fact that they are the most important controls. When $\delta = 1$, there is equal selection on observables and unobservables i.e. both are equally important and affect β in the same direction. When $0 < \delta < 1$, unobservables are less important than observed factors, and the opposite is true when $\delta > 1$. $\delta = -1$ is the analogous version of equal selection when the selection on observables and unobservables occurs in opposite directions

[130] Note that, by definition, the set of unobserved variables potentially correlated with both anticipated discrimination and educational outcomes (which would cause the OVB) is necessarily omitted from the controlled regression.

(i.e. one is positive and the other one is negative), and possible in theory. The second key parameter is the R-squared from a hypothetical regression of the outcome on the main variable of interest on the complete set of observed and unobserved controls. This is called R_{max}. Following the recommendation in Oster (2019), I assume $R_{max} = 1.3\tilde{R}$. This value of R_{max} reflects how much of the variation in GCSE scores (for each respective outcome) could be explained if I had full controls for all the relevant determinants of educational attainment.

This approach allows the estimation of the upper and lower bounds for the main coefficient of interest under certain assumptions. Given R_{max}, it is possible to compute a bounding set with bounds on the value of β. The results are presented in Table 4.8. The first column repeats the estimates from the controlled regression ($\tilde{\beta}$), which assumes no selection on unobservables ($\delta = 0$). Then, the second column presents the bias-adjusted Oster estimates assuming $\delta = -1$ (lower bound) and the third column presents the bias-adjusted Oster estimates under the assumption that $\delta = 1$ (upper bound). It is notable that the estimates from the original controlled regression lie between the bounded estimates obtained by the Oster method. The bias-adjusted effects of anticipated discrimination on GCSE scores have the same sign, and are close in magnitude to the estimated coefficients from the controlled regressions, with overlapping 95% confidence intervals.[131] Overall, this indicates that my main results are most likely not threatened by an omitted variable bias from unobserved variables.

Although possible in theory, it is difficult to conceive a situation where $\delta = -1$ would be a plausible assumption with respect to anticipated discrimination, as this would mean that the selection on unobservables works in the opposite direction from the selection on observables. Therefore, the original results (with $\delta = 0$) are more likely to be a more realistic lower bound, with the bias-adjusted Oster estimates β^* ($\delta = 1$) as the upper bound.

[131] Although some bias is exhibited, this result should be interpreted with caution because the bounds are very close to the original point estimates.

Table 4.8 Bounding statements about β

	$\delta = 0(\tilde{\beta})$ (OLS/LPM)	$\delta = -1$ $R_{max} = 1.3\tilde{R}$	$\delta = 1$ (β^*) $R_{max} = 1.3\tilde{R}$
Total GCSE score	0.067	0.052	0.084
5A*-C	0.049	0.042	0.057
5+ A*-C Eng,Math	0.057	0.046	0.069
A*-C Eng	0.046	0.045	0.046
A*-C Math	0.060	0.045	0.076
Science score	0.101	0.078	0.126
Any A*-A	0.019	0.007	0.033

Notes: All outcomes and controls are described in the text in section 4.4.1. $\tilde{\beta}$ and $\tilde{R}$ are obtained from the controlled regression based on Model 2, which includes the extended set of controls. Regressions run for the pooled sample of ethnic minority pupils (N=3315). Source: Next Steps

To complement this analysis, Table 4.9 shows values of δ that would produce $\beta = 0$ under the assumed R_{max} for each educational outcome. These numbers indicate how important the unobservables would have to be to bring the coefficients on anticipated discrimination to zero. For example, looking at the achievement of five or more A*-C including English and Maths, unobservables would need to be 6.71 times more important than observables (and act in the opposite direction) to eliminate the observed effect. For all outcomes analysed, the values of δ are quite large, indicating that unobservables would need to be considerably more important than observables to bring $\beta = 0$. This is a highly unlikely scenario considering that the observable characteristics come from a rich set of controls that have been established as key determinants of educational outcomes by an extensive literature. Therefore, these findings are reassuring that the estimated relationship between anticipated discrimination and educational achievement is not seriously biased by some selection on unobserved characteristic(s) unaccounted for.

Table 4.9 Statements about δ (δ for $\beta = 0$)

	$R_{max} = 1.3\tilde{R}$
Total GCSE score	-5.72
5A*-C	-10.93
5+ A*-C Eng,Math	-6.71
A*-C Eng	16.49
A*-C Math	-5.12
Science score	-5.68
Any A*-A	-1.68

Notes: All outcomes and controls are described in the text in section 4.4.1. $\tilde{R}$ is obtained from the controlled regression based on Model 2, which includes the standard and extended sets of controls. Regressions run for the pooled sample of ethnic minority pupils (N=3315). Source: Next Steps.

4.5.4 Considering prior achievement: Value-added models

In this section, I explore whether the observed association between anticipated discrimination and educational attainment at GCSE level is driven by differences in pupils' prior achievement. Particularly, I analyse whether the observed relationship likely reflects a selection effect whereby high achievers are more likely to anticipate discrimination.

After considering all the variables in the standard and extended set of controls, a remaining challenge for this analysis concerns a very important variable in the education production function, namely pupils' unobserved initial ability endowment. Clearly, students with higher ability will perform better, on average, in national exams. Accounting for innate ability is necessary to be able to distinguish between the effects of effort, non-cognitive skills and natural innate ability on educational performance. The omission of this variable in the reduced-form equation could produce biased coefficients of anticipated discrimination due to OVB, if anticipating discrimination is correlated with innate ability. Lang and Manove (2011) describe how the incentive to acquire more human capital varies across the ability distribution. This could be the case, for example, if pupils with higher innate ability, who are more likely to achieve better GCSE results, are also more likely to anticipate discrimination due to an increased awareness of discriminatory behaviour and its potential consequences in the labour market. Ideally, one would need to control for some measure of pupil's innate ability, such as IQ scores to address this concern. Unfortunately, such measures are unavailable in Next Steps.

Other studies examining the determinants of educational achievement and educational gaps have alternatively used value-added models of educational achievement (see for example Wilson et al., 2011). The idea behind this type of model is that prior educational achievement can be used as a rough approximation for the history of a series of inputs, including as an indicator of initial ability endowment. In Next Steps, I have access to students' total point scores for the tests taken at the end of Key Stage 2, when pupils are aged 10-11. A well-specified value-added model would provide estimates of important the determinants of educational progress between KS2 and KS4.[132] Here, I explore the implications of adding KS2 scores as an additional control to the results from my baseline and extended models. Of course, estimating such value-added models is empirically challenging in this setting and it is important to note that it does not come without limitations, an issue I return to below.

I begin by examining the distribution of prior achievement scores in my pooled sample. Figure 4.4 shows the density plot for the distribution of KS2 scores by anticipated discrimination. This figure reveals students who anticipate discrimination have achieved, in general, higher KS2 results in comparison to those who do not anticipate discrimination. This is also the pattern observed when looking at the distribution of KS2 scores separately for each ethnic group (Figure 4.5). For all groups analysed, the distribution of KS2 scores is shifted to the right for pupils who anticipate discrimination. Furthermore, examining variations in the role of anticipating discrimination along the distribution of KS2 scores (in Table C.20), I observe that the positive coefficients for the pooled sample seem to be driven by pupils who achieve

[132]Ideally, when using longitudinal data to examine educational progress (value-added), it is important to control for other time-varying characteristics. In this data set, however, most available control variables are time invariant (.e.g gender, ethnic group), or time-varying but measured only once (e.g. anticipated discrimination).

in the middle (second and third quartiles) of the KS2 distribution, and not so much by those at the very top or the very bottom of this distribution.

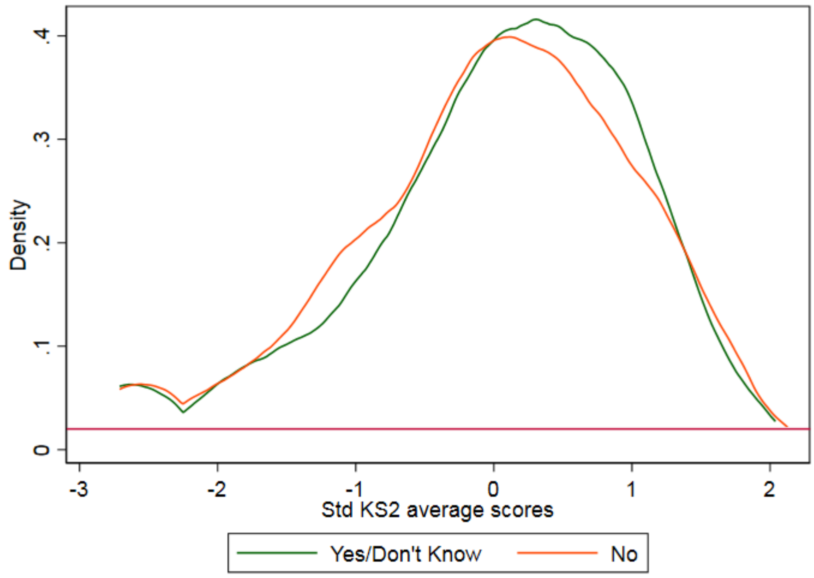


Figure 4.4 Kernel density distribution of KS2 scores by anticipation of discrimination

Notes: KS2 scores are the standardised average points scores with mean zero and standard deviation one. The red horizontal line delimits the threshold for which the sample size is considered non-disclosive and compliant with the secure data service requirements. Pooled sample of ethnic minority pupils. Source: Next Steps.

Table 4.10 reports the summarised results of the value-added models for the pooled sample of all ethnic minority pupils, based on Models 1 and 2. Overall, I find a significant positive association between KS2 scores and GCSE attainment across all measures, as expected. As shown by many other studies, prior achievement is usually a significant predictor of test results. However, controlling for KS2 scores results in smaller, mostly non-significant coefficients for anticipated discrimination for the pooled sample of all ethnic minority pupils. The coefficient for achieving the gold standard is around 2.3 times smaller when estimated using the value-added specifications. The coefficients for achieving an A*-C in Maths, still significant at 10% and 5%, also drop by more than half. It is important to note that the interpretation of these coefficients from value-added models differs from those presented previously, as now they reflect how anticipated discrimination influences the progress made between KS2 and KS4 or, in other words, the achievement in GCSEs conditional on KS2 results.

Table 4.11 presents the same results by ethnic groups, considering the value-added specification of Model 2. Most of positive associations are lost when including prior achievement, but not for all ethnic groups and outcomes. Even after including KS2 scores, the association remains positive for pupils of South Asian heritage and some of the coefficients remain significantly different from zero. Anticipating discrimination predicts that pupils of Indian origin are 9.8 pp more likely to achieve an A*-C in English, and that pupils from the Pakistani group are 8.5 pp more likely to achieve five or more A*-C GCSEs including English and Maths.

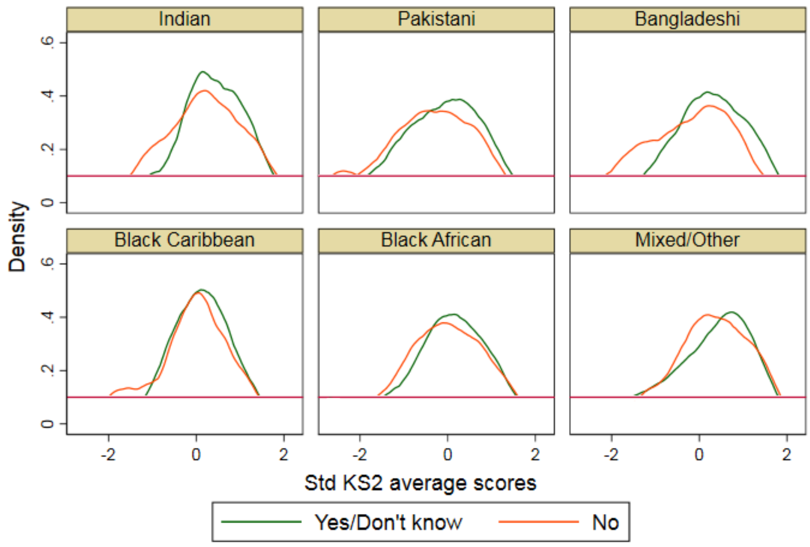


Figure 4.5 Kernel density distribution of KS2 scores by anticipated of discrimination - separately by ethnic group

Notes: KS2 scores are the standardised average points scores with mean zero and standard deviation one. The red horizontal line delimits the threshold for which the sample size is considered non-disclosive and compliant with the secure data service requirements. Source: Next Steps.

While the relationship between anticipated discrimination and achievement becomes weaker and not significant for most outcomes, I find it to still be positive for some outcomes and ethnic groups, which suggests that selection on prior achievement cannot fully explain the differences in GCSE achievement by anticipated discrimination. While this confounding factor is not ruled out here, it seems unlikely that a selection on prior achievement explains the whole story. It is plausible that anticipated discrimination influences prior achievement before the age of 16, for example the KS2 scores. A study by Carneiro et al. (2005) for the US finds that ethnic minority children as young as age 10 already had pessimistic expectations about educational achievement. If this is also the case here, the role of anticipated discrimination after controlling for prior achievement would be underestimated. I concede that further research is required to understand the age at which expectations of discrimination are formed and when it may manifest itself in academic achievement to fully disentangle these findings.

Table 4.10 Anticipated discrimination and GCSE achievement - summarised results value-added models VA-M1 and VA-M2 (OLS and LPM)

	Educational Outcomes						
	Total GCSE score	5+ A*-C	5+ A*-C Eng,Math	A*-C Eng	A*-C Math	Science score	Any A*-A
Panel A: Model VA-M1							
Anticipate Disc	0.002	0.017	0.026	0.013	0.027*	0.032	0.004
	(0.029)	(0.017)	(0.017)	(0.016)	(0.015)	(0.030)	(0.018)
KS2 score	0.583***	0.254***	0.283***	0.264***	0.282***	0.596***	0.209***
	(0.020)	(0.009)	(0.009)	(0.009)	(0.009)	(0.021)	(0.010)
R2	0.645	0.501	0.529	0.516	0.534	0.607	0.447
Panel B: Model VA-M2							
Anticipate Disc	0.011	0.022	0.029*	0.017	0.031**	0.041	0.004
	(0.029)	(0.017)	(0.016)	(0.016)	(0.015)	(0.030)	(0.018)
KS2 score	0.535***	0.229***	0.257***	0.239***	0.259***	0.547***	0.188***
	(0.020)	(0.010)	(0.009)	(0.009)	(0.010)	(0.022)	(0.010)
R2	0.660	0.520	0.548	0.535	0.550	0.623	0.464
N	3315	3315	3315	3315	3315	3315	3315

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score, Science score, and KS2 score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. VA-M1 includes the standard set of controls plus prior achievement in KS2. VA-M2 includes the extended set of controls plus prior achievement in KS2. Pooled sample of ethnic minority pupils. Source: Next Steps.

Table 4.11 Anticipated discrimination and GCSE achievement - Summarised results value-added model VA-M2 (OLS and LPM); by ethnic group

	Ethnic Group					
Dependent variable	Indian	Pakistani	Bangladeshi	Black Caribbean	Black African	Mixed/Other
Total GCSE score						
Anticipate Disc	0.023	0.080	0.056	0.024	-0.074	-0.030
	(0.075)	(0.078)	(0.115)	(0.179)	(0.158)	(0.108)
KS2 score	0.564***	0.601***	0.479***	0.403***	0.402***	0.622***
	(0.044)	(0.043)	(0.063)	(0.115)	(0.075)	(0.071)
5+ A-C*						
Anticipate Disc	0.074*	0.052	-0.010	-0.041	0.033	0.038
	(0.040)	(0.048)	(0.055)	(0.097)	(0.082)	(0.067)
KS2 score	0.219***	0.281***	0.213***	0.216***	0.205***	0.250***
	(0.027)	(0.024)	(0.024)	(0.051)	(0.049)	(0.034)
5+ A-C incl. Eng, Mat*						
Anticipate Disc	0.061	0.085*	0.052	-0.020	-0.027	0.071
	(0.039)	(0.046)	(0.055)	(0.082)	(0.104)	(0.067)
KS2 score	0.258***	0.246***	0.263***	0.266***	0.217***	0.286***
	(0.028)	(0.024)	(0.026)	(0.055)	(0.058)	(0.035)
A-C Eng*						
Anticipate Disc	0.098**	0.027	-0.021	0.153	-0.007	-0.006
	(0.042)	(0.043)	(0.052)	(0.109)	(0.087)	(0.062)
KS2 score	0.237***	0.283***	0.258***	0.224***	0.178***	0.263***
	(0.027)	(0.025)	(0.026)	(0.067)	(0.055)	(0.033)
A-C Math*						
Anticipate Disc	0.051	0.077	0.018	0.018	-0.009	-0.002
	(0.039)	(0.049)	(0.050)	(0.085)	(0.082)	(0.062)
KS2 score	0.233***	0.263***	0.264***	0.264***	0.283***	0.308***
	(0.026)	(0.027)	(0.021)	(0.052)	(0.048)	(0.037)
Science score						
Anticipate Disc	0.061	0.123	0.164	-0.045	0.022	-0.035
	(0.078)	(0.080)	(0.126)	(0.158)	(0.195)	(0.108)
KS2 score	0.547***	0.606***	0.509***	0.487***	0.393***	0.659***
	(0.037)	(0.053)	(0.068)	(0.102)	(0.087)	(0.069)
Any A-A*						
Anticipate Disc	-0.005	0.037	0.042	0.018	0.003	0.027
	(0.058)	(0.041)	(0.061)	(0.095)	(0.095)	(0.061)
KS2 score	0.210***	0.191***	0.177***	0.119**	0.219***	0.182***
	(0.021)	(0.024)	(0.042)	(0.060)	(0.053)	(0.035)
N	712	676	477	366	289	795

Notes: Robust standard errors clustered at the secondary school level in parentheses: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Total GCSE score, Science score and KS2 score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. VA-M2 includes the extended set of controls plus prior achievement in KS2. Source: Next Steps

4.5.5 Robustness checks

Separating 'Yes' and 'Don't Know' categories

As described in section 4.4.2, for the main analysis in this chapter, I combine the 'yes' and 'don't know' response categories for the question on anticipated discrimination. It is important, however, to analyse the implications of combining these two response categories for my main results.

In order to examine this, I estimate the main models again, considering the three response categories of anticipated discrimination separately. The proportion of pupils in each of these categories by ethnicity is presented in Table C.1 in the Appendix. The regression results for the pooled sample are shown in Table C.21. I also provide the p value of the t-test of equality of the 'yes' and 'don't know' coefficients. Overall, these results suggest that combining the two response categories does not change my main results qualitatively. It can be seen from Table C.21 that the point estimates for 'yes' are larger than those for 'don't know' for both Models 1 and 2, with the exception of the coefficients for Science score which are only marginally smaller. These estimates suggest that the influence of anticipated discrimination is on average slightly stronger for pupils who respond 'yes', that is, who believe that their ethnic origin might make it more difficult to get a job in the future. Nonetheless, the positive (and mostly statistically significant) coefficients for 'don't know' also suggest that it is likely that the pupils who gave this answer anticipate this type of discrimination. Nevertheless, the t-test comparing 'yes' and 'don't know' estimated coefficients reveals that these small differences observed are generally not statistically significant, thus providing further support to combining these two response categories.[133]

An examination of the results obtained from this alternative specification that considers the 'yes' and 'don't know' answers separately by ethnic group supports these findings. As

[133] The few exceptions where some significant differences are observed are the larger coefficient for 'yes' for achieving an A*-C in English, which is significantly different from that of 'don't know' at 5% in Model 1 and 1% in Model 2, and for the estimates for Any A*-A results in both models, are also larger for 'yes'.

shown in Table C.22, these two coefficients are not significantly different from each other for most outcomes and ethnic groups. Moreover, the sign and magnitude of ‘yes’ coefficients for pupils in the Indian, Pakistani and Bangladeshi groups support the findings from the main analysis, and provide additional suggestive evidence of differences across ethnic groups. It is interesting to note that among pupils of Black Caribbean and Black African descent, the ‘yes’ estimates are usually negative and smaller than the ‘don’t know’ (except for A*-C in English). Although it is not possible to observe statistical differences between these coefficients, likely due to smaller samples, this is an important topic for further investigation as it might hold key insights into the type of discrimination expected by different ethnic groups and its resulting implications for human capital investment.

Differences by gender

I also explore the existence of differences in the role of anticipated discrimination between boys and girls. I focus on alternative specifications of Model 1 and 2, which also includes an interaction term of anticipated discrimination and gender of the pupil. Respective results are reported in Table C.23. These results reveal that there is little evidence of heterogeneity of effects by gender.

Differences by interview period (Post July/2005)

In a previous study, also using wave 2 of Next Steps, Hole and Ratcliffe (2019) find a rise in expectations of facing future labour market discrimination among Muslim students interviewed in the aftermath of the 2005 Islamic terror attacks in London. As this may have temporarily raised expectations of facing future labour market discrimination, especially among certain ethnic minority groups, it is therefore important to examine whether there is a differential effect of anticipated discrimination between pupils interviewed before and after July 2005. The results, presented in Table C.24, reveal that there is no significant differential effect by the period of interview.

Weighted regressions

Results considering the survey weights provided in Next Steps are reported in Tables C.25 to C.28 in Appendix C. The main regression results are shown to be robust to using weights.

4.6 Conclusions

This chapter has investigated whether expecting future discrimination in the labour market influences the educational achievement of ethnic minority pupils. It contributes to the broad literature on education, and to two specific strands of this literature, on the role of discrimination for human capital formation and on the determinants of educational achievement for ethnic minorities. This study uses data from the first waves of Next Steps, which includes a unique question on adolescents' expectations of facing future discrimination in the labour market. In addition to this, Next Steps contains a wealth of information on pupils' socio-demographic and economic characteristics for a large cohort of English students, as well as data on educational achievement at key stages of schooling.

The main analysis investigates the influence of anticipating labour market discrimination on performance in GCSE exams, which represent the first high-stake exams that pupils encounter as part of their education. These exams determine progression to further education and form part of the requirements in university applications. In the primary analysis, which pools all ethnic minority pupils, I find evidence of a positive relationship between anticipating discrimination and higher achievement on a number of educational outcomes, while holding important background factors constant. Specifically, I control for a number of socio-economic, individual and family characteristics, as well as school fixed effects. The preferred specification indicates that, ceteris paribus, anticipating labour market discrimination increases the probability of obtaining five GCSEs A*-C including English and Maths, i.e. achieving the gold standard, by around 6.4 percentage points. In addition, pupils anticipating discrimination are around 5.0 pp more likely to obtain an A*-C in English and 6.6 pp more likely to obtain an

A*-C in Maths, relative to similar individuals who do not expect this type of discrimination. Such evidence suggests that the expectations of facing future discrimination can play an important role in the future of young people, since they may influence their decisions related to investment in human capital.

Overall, these findings support the theoretical predictions of Lang and Manove (2011) and the empirical evidence from other work (Arcidiacono et al., 2010; Lang and Manove, 2011; Leslie and Drinkwater, 1999; Lessard-Phillips et al., 2018) that suggests that ethnic minority individuals increase their investment in human capital to offset a possible ethnic penalty they expect to face in the market. In contrast, this evidence is less well aligned with the predictions of some earlier theoretical work on discrimination (such as Coate and Loury (1993*b*) and Lundberg and Startz (1983)), that predicts an under-investment in skill accumulation by minorities.

This analysis also provides key evidence on the heterogeneous effects of anticipating discrimination by ethnic group. In particular, pupils of South Asian origin who expect discrimination have much better GCSE results than those who do not. In contrast, for other groups, such as for Black Caribbean and Black African pupils, the effect of anticipating discrimination is much smaller, and sometimes even negative, but never statistically different from zero. This suggests that the type and or severity of (expected) discrimination could vary across ethnic groups - for pupils of Black African and Black Caribbean origin (the largest groups anticipating discrimination), discrimination might be perceived as so bad that nothing can be done to overcome it. Meanwhile, for pupils from other ethnic groups, who also anticipate discrimination, discrimination might be perceived as surmountable and so they adopt a compensatory strategy by investing more in education. Although the exact reasons for these differences are unknown, one hypothesis is that it relates to historical migration patterns of different ethnic groups, the timing and reasons for migrating to the country as well as the different endowments of social, cultural and economic capital. Future research should seek to understand variations in the type and severity of discrimination experienced and expected by individuals from different ethnic groups and its origins, to better understand these findings.

I employ a series of techniques to address endogeneity concerns and selection issues. Considering omitted variable bias concerns, while it is not impossible that there is some selection based on unobserved characteristics driving this association, the bounded estimates obtained using Oster's method indicate that my main results are most likely not threatened by an omitted variable bias from unobservables. Furthermore, the observed relationship may reflect a causal effect of anticipated discrimination on educational achievement. Alternatively, it could reflect the selection of high achievers into being more likely to anticipate discrimination. The selection story is a possibility - in fact, very little is known about the formation of these expectations of facing future discrimination. I have considered this possibility by examining whether differences in prior achievement drive the relationship. My results reveal that conditioning on prior achievement considerably reduces the educational gap from anticipating discrimination, although not uniformly for all outcomes. However, it is possible that prior achievement, measured by test scores at ages 10-11 is already influenced by expectations of discrimination and therefore captures most of this effect.

Despite my key findings, it is important to acknowledge the shortcomings of this study that preclude me from making definitive assertions on the causal effect of anticipated discrimination on educational attainment. While this analysis provides evidence of a significant, positive association between anticipated discrimination and educational achievement, it is not possible to establish full causality. The discussion and examination of endogeneity concerns and selection issues have indicated that it could be important to control for a measure of innate ability, such as IQ. This would also afford the opportunity to examine whether the role of expectations varies along the ability distribution. Finally, in the analysis of the differential ethnicity effects, findings of statistically insignificant coefficients may be due to data restrictions and sample sizes due to disaggregation of the main sample in smaller ethnic groups. A larger sample size and dataset considering ethnic minority individuals may therefore have benefited this part of the analysis.

While there are no directly comparable studies, the effect of anticipated discrimination is more modest than what is usually reported for other key predictors of achievement, such as gender, poverty status or English as first language (Dustmann et al., 2010; Strand, 2014; Wilson et al., 2011). This was to be expected. However, even after controlling for these (and various other) characteristics, anticipated discrimination still significantly explains some of the differences in educational achievement. This unveils the presence of a deeper layer of factors that influence attainment of minorities and, thus, may influence the dynamics of social mobility in British society.

There has been little previous research regarding how the anticipation of discrimination in the labour market may influence educational decisions that happen prior to entry in the market. The results from this chapter indicate that perceptions of unequal opportunities and expectations of facing discriminatory treatment could influence real decisions related to investment in human capital. There are important implications to be drawn from these findings. First, given the difficulty of identifying a priori those pupils who perceive opportunities as unequal, their reasons and when expectations of discrimination are formed, these findings emphasise the importance of investigating the formative process of such expectations. Second, these results could have interesting implications for policies aiming to address ethnic gaps in the labour market. The finding that some groups attain better results in response to expected discrimination could, in a vacuum, perhaps be seen as a positive outcome. However, this over-investment in human capital acquisition to counteract the expected discrimination could represent a loss of efficiency with broader implications. By over-investing in education, time and resources are used towards this end and important investments in other areas, for example in non-cognitive skills, may be crowded out. Future research is needed to investigate whether this occurs. In addition, an issue for policy is that the additional investments in education and better qualifications are not translating into better future outcomes for minority ethnic groups in the UK. As previous evidence suggests, UK ethnic minorities obtain relatively lower labour market returns (Blackaby et al., 2002; Heath and Cheung, 2006; Metcalf, 2009).

Education is considered a main driver of social mobility and this channel is particularly crucial for ethnic minorities who strive to move up the social ladder. One conclusion one may reach from this analysis is that if discrimination represents a constraint for ethnic minorities, then this challenges the notion that a more equal society can be achieved through promoting equality of opportunity through education. A lack of equality of opportunity can add to the effect of socio-economic background, conspiring to prevent ethnic minorities from receiving the potential benefits from the qualifications they attain (Platt, 2007). The very existence of expectations discrimination highlights the need of continuing the promotion of anti-discrimination laws and for schools, universities and employers to provide targeted support to ensure these minority groups are able to achieve their career ambitions and progress in the labour market.

Appendix C

C.1 Variable definitions

Standard set of controls

- Female: takes value 1 if sex of pupil is female and 0 if male.
- Ethnicity: Mixed/Other (reference cat.), Indian, Pakistani, Bangladeshi, Black Caribbean, Black African.
- Region of residence: London (reference cat.), North East, North West, Yorkshire and the Humber, East Midlands, West Midlands, East of England, South East, South West.
- EFL: takes value 1 if pupil speaks English as first/main language or is bilingual, 0 otherwise.
- No SEN: takes value 1 if pupil has no special education needs, 0 otherwise.
- Number of siblings: number of siblings in the household, continuous variable.
- Two-parent family: takes value 1 if pupil lives with both parents in the household, 0 otherwise.
- Homeowner: takes value 1 if family home is owned or being bought, 0 otherwise.
- No FSM: takes value 1 if pupil is not eligible for free school meals, 0 otherwise.
- Parent(s) highest education levels: low/no qualifications (reference cat.), GCSE, A-levels/HE below degree, degree.
- Parent(s) employment status: takes value 1 if any parent is at work, 0 otherwise.
- Parent(s) in professional occupations: takes value 1 if any parent works in professional or managerial occupations, 0 otherwise.

- Household income over £21k: takes value 1 if household income is above the median for the sample (£21k), 0 otherwise.
- Parent(s) aged < 40: takes value 1 if any parent is younger than 40 (median for the sample) in wave 1, 0 otherwise.
- Parent(s) very good health: takes value 1 if any parent reports to be in very good health, 0 otherwise.

Extended set of controls (all controls from the standard set plus):

- Parent would like child to continue in FTE: takes value 1 if parent reports it would like child to continue in full time education after the age of 16, 0 otherwise.
- Higher education very likely: takes value 1 if parent reports it to be very likely pupil will go to university, 0 otherwise.
- Very involved in school life: takes value 1 if parent reports to be very involved in their child's school life, 0 otherwise.
- Future-orientated: takes value 1 if pupil reports thinking about what they will be doing in a few years time, 0 otherwise.
- Raising family important: takes value 1 if pupil considers raising a family important, 0 otherwise.
- Importance of job characteristics: Separate indicators for 'Job helping other important', 'Well paying job important','Being own boss important','Interesting job important', 'Job with promotion important', 'Job with regular hours important'. These variables take value 1 if pupil thinks that the respective job characteristic is important, 0 otherwise.
- Has ideas for career: Takes value 1 if pupil has any ideas for career, 0 otherwise.
- Bullying: Takes a value of 1 if pupil has not experienced any type of bullying in the 12 months prior to wave 1, 0 otherwise.

C.2 Supplementary results

Table C.1 Anticipation of labour market discrimination by ethnic minority group (%)

	Ethnic Minority Group						
Anticipate discrimination (%)	Indian	Pakistani	Bangladeshi	Black Caribbean	Black African	Mixed/ Other	Total
No	70.5	61.4	62.3	49.1	47.6	67.5	62.6
Yes	11.4	12.0	13.7	27.2	30.1	11.8	15.4
Don't Know	18.1	26.6	24.0	23.8	22.3	20.6	22.0
Total	100	100	100	100	100	100	100

Notes: Survey weights are applied. Source: Next Steps.

Table C.2 Anticipation of labour market discrimination by ethnic minority group (%), unweighted

	Ethnic Minority Group						
Anticipate discrimination (%)	Indian	Pakistani	Bangladeshi	Black Caribbean	Black African	Mixed/ Other	Total
No	69.7	61.4	61.6	46.4	48.4	65.0	61.3
Yes	12.1	13.0	13.0	28.4	29.4	14.2	16.2
Don't Know	18.3	25.6	25.4	25.1	22.1	20.7	22.5
Total	100	100	100	100	100	100	100

Notes: Unweighted statistics. Source: Next Steps.

Table C.3 Raw gaps in GCSE performance by anticipated labour market discrimination (unweighted)

		Anticipate Discrimination	
Educational Outcomes	All	Yes/Don't Know	No
Total GCSE points score	390.084	392.585	388.505
Total GCSE score (Std)	0.000	0.017	-0.011
Achieved 5 or more GCSEs (proportion)	0.629	0.649	0.617*
Achieved 5 or more GCSEs incl. Eng, Maths (proportion)	0.482	0.509	0.465**
Achieved A*-C in GCSE English (proportion)	0.622	0.650	0.605**
Achieved A*-C in GCSE Maths (proportion)	0.559	0.582	0.544**
Science score (Std)	0.000	0.035	-0.022
Achieved any A*-A (proportion)	0.376	0.380	0.374
N	3315	1283	2032

Notes: Unweighted statistics. For the t-test of differences in means between 'Yes/Don't Know' and 'No':*$p < 0.10$,**$p < 0.05$, ***$p < 0.01$. Source: Next Steps.

Table C.4 Differences in sample means by anticipated discrimination: pooled sample

Variable	All	Anticipate discrimination Yes/ Don't Know	No
Individual Characteristics			
Female	0.52	0.51	0.52
Indian	0.19	0.15	0.21***
Pakistani	0.18	0.19	0.18
Bangladeshi	0.07	0.07	0.07
Black Caribbean	0.11	0.16	0.09***
Black African	0.10	0.14	0.07***
Mixed/Other	0.35	0.30	0.37***
English first/main language	0.84	0.84	0.84
No special education needs	0.86	0.86	0.86
Region of Residence			
North East	0.02	0.02	0.02
North West	0.10	0.10	0.10
Yorkshire and The Humber	0.10	0.10	0.09
East Midlands	0.06	0.05	0.06**
West Midlands	0.14	0.15	0.14
East of England	0.06	0.06	0.06
London	0.43	0.44	0.42
South East	0.08	0.07	0.08
South West	0.02	0.02	0.03
Family Characteristics			
Number of siblings	1.98	1.94	2.01
Two-parent family	0.70	0.68	0.72
Homeowner	0.61	0.60	0.62
Not eligible to FSM	0.73	0.73	0.73
Parental Characteristics			
Parent(s) degree	0.07	0.08	0.07
Parent(s) A-levels/HE below degree	0.17	0.18	0.16
Parent(s) GCSE	0.17	0.16	0.17
Parent(s) low/no qualification	0.59	0.57	0.60
Parent(s) work	0.68	0.67	0.69
Parent(s) in professional occupations	0.29	0.29	0.30
Parent(s) SOC missing	0.11	0.11	0.11
Household income over £21k	0.36	0.36	0.37
Household income missing	0.30	0.32	0.29
Parent(s) aged <40	0.41	0.39	0.43
Parent(s) in very good health	0.53	0.51	0.55
Parental aspirations and involvement			
Parent would like YP to continue in FTE	0.94	0.94	0.94
Higher education very likely (Parent)	0.56	0.57	0.55
Very involved in school life (Parent)	0.33	0.34	0.33
Personality and Future Orientation			
Future-oriented	0.70	0.71	0.69
Raising family important	0.90	0.90	0.90
Job helping other important	0.45	0.47	0.44
Well paying job important	0.76	0.77	0.75
Being own boss important	0.35	0.36	0.34
Interesting job important	0.69	0.70	0.68
Job with promotion important	0.71	0.71	0.71
Job with regular hours important	0.48	0.47	0.49
Has ideas for career	0.76	0.75	0.77
Bullying			
Not bullied	0.60	0.57	0.62**
Bullying missing	0.02	0.03	0.02
Prior achievement			
KS2 score (Std)	-0.03	0.01	-0.05
Unweighted count	3315	1283	2032

Notes: Survey weights applied. For the t-test of differences in means:*$p < 0.10$,**$p < 0.05$, ***$p < 0.01$. The unweighted sample size for household income over 21k is 2258, for not bullied is 3236 and for parent(s) in professional occupation is 2894. Source: Next Steps.

Table C.5 Differences in sample means by anticipated discrimination: Indian

Variable	All	Anticipate discrimination Yes/ Don't Know	No
Outcomes			
Total GCSE point score	427.07	437.45	422.71
Total GCSE score (Std)	0.25	0.32	0.22
Achieved 5 or more GCSEs	0.70	0.76	0.68*
Achieved 5 or more GCSEs incl. Eng, Maths	0.60	0.66	0.57*
Achieved A*-C in GCSE English	0.70	0.76	0.68*
Achieved A*-C in GCSE Maths	0.70	0.76	0.67**
Science score (Std)	0.27	0.33	0.24
Achieved any A*-A	0.46	0.46	0.46
Individual Characteristics			
Female	0.50	0.45	0.51
English first/main language	0.86	0.87	0.85
No special education needs	0.89	0.89	0.89
Region of Residence			
North East	R	R	R
North West	0.08	R	R
Yorkshire and The Humber	0.06	0.07	0.05
East Midlands	0.12	0.11	0.12
West Midlands	0.22	0.26	0.21
East of England	0.04	R	R
London	0.38	0.36	0.39
South East	0.08	0.07	0.09
South West	R	R	R
Family Characteristics			
Number of siblings	1.76	1.64	1.80**
Two-parent family	0.86	0.85	0.87
Homeowner	0.88	0.88	0.88
Not eligible to FSM	0.90	0.88	0.90
Parental Characteristics			
Parent(s) degree	0.05	0.05	0.05
Parent(s) A-levels/HE below degree	0.14	0.15	0.14
Parent(s) GCSE	0.19	0.18	0.20
Parent(s) low/no qualification	0.61	0.62	0.61
Parent(s) work	0.86	0.87	0.86
Parent(s) in professional occupations	0.31	0.28	0.33
Parent(s) SOC missing	0.04	0.03	0.05
Household income over £21k	0.46	0.44	0.47
Household income missing	0.33	0.31	0.34
Parent(s) aged <40	0.40	0.37	0.41
Parent(s) in very good health	0.56	0.50	0.58*
Parental aspirations and involvement			
Parent would like YP to continue in FTE	R	R	R
Higher education very likely (Parent)	0.62	R	R
Very involved in school life (Parent)	0.28	0.28	0.28
Personality and Future Orientation			
Future-oriented	0.68	0.70	0.67
Raising family important	0.94	R	R
Job helping other important	0.40	0.42	0.39
Well paying job important	0.76	0.73	0.77
Being own boss important	R	R	R
Interesting job important	0.67	0.72	0.65*
Job with promotion important	0.73	0.73	0.73
Job with regular hours important	0.49	0.49	0.49
Has ideas for career	0.74	0.74	0.75
Bullying			
Not bullied	0.64	0.57	0.67**
Bullying missing	0.02	R	R
Prior achievement			
KS2 score (Std)	0.17	0.27	0.14
Unweighted count	712	216	496

Notes: Survey weights applied. For the t-test of differences in means:*$p < 0.10$,**$p < 0.05$, ***$p < 0.01$. 'R' represents values that have been suppressed to comply with the secure data service non-disclosure restrictions. Source: Next Steps.

Table C.6 Differences in sample means by anticipated discrimination: Pakistani

Variable	All	Anticipate discrimination Yes/ Don't Know	No
Outcomes			
Total GCSE and equivalents new style point score	341.99	368.26	325.45***
Std GCSE points score	-0.32	-0.15	-0.43***
Achieved 5 or more GCSEs	0.49	0.56	0.44**
Achieved 5 or more GCSEs incl. Eng, Maths	0.35	0.45	0.29***
Achieved A*-C in GCSE English	0.50	0.56	0.45***
Achieved A*-C in GCSE Maths	0.42	0.51	0.37***
Higher Science std score	-0.31	-0.11	-0.43***
Achieved any A*-A	0.27	0.32	0.24**
Individual Characteristics			
Female	0.49	0.50	0.49
English first/main language	0.77	0.79	0.76
No special education needs	0.89	0.92	0.87
Region of Residence			
North East	R	R	R
North West	0.21	0.19	0.21
Yorkshire and The Humber	0.25	0.25	0.25
East Midlands	0.02	R	R
West Midlands	0.17	0.14	0.18
East of England	0.07	0.08	0.06
London	0.19	0.19	0.19
South East	0.07	R	R
South West	R	R	R
Family Characteristics			
Number of siblings	2.71	2.65	2.76
Two-parent family	0.84	0.83	0.84
Homeowner	0.76	0.77	0.76
Not eligible to FSM	0.62	0.63	0.61
Parental Characteristics			
Parent(s) degree	0.03	R	R
Parent(s) A-levels/HE below degree	0.06	R	R
Parent(s) GCSE	0.07	0.08	0.07
Parent(s) low/no qualification	0.84	0.85	0.82
Parent(s) work	0.56	0.58	0.55
Parent(s) in professional occupations	0.23	0.26	0.21
Parent(s) SOC missing	0.19	0.20	0.19
Household income over £21k	0.22	0.24	0.20
Household income missing	0.39	0.38	0.40
Parent(s) aged <40	0.46	0.44	0.47
Parent(s) in very good health	0.44	0.45	0.43
Parental aspirations and involvement			
Parent would like YP to continue in FTE	0.95	R	R
Higher education very likely (Parent)	0.50	0.52	0.48
Very involved in school life (Parent)	0.28	0.25	0.30
Personality and Future Orientation			
Future-oriented	0.64	0.68	0.62
Raising family important	0.90	0.90	0.90
Job helping other important	0.54	0.49	0.56**
Well paying job important	0.81	0.83	0.80
Being own boss important	0.42	0.38	0.45
Interesting job important	0.69	0.63	0.72**
Job with promotion important	0.76	0.73	0.78
Job with regular hours important	0.50	0.50	0.49
Has ideas for career	0.72	0.68	0.74
Bullying			
Not bullied	0.60	0.60	0.60
Bullying missing	0.03	R	R
Prior achievement			
Std KS2 average scores	-0.37	-0.19	-0.48***
Unweighted count	676	261	415

Notes: Survey weights applied. For the t-test of differences in means:*$p < 0.10$,**$p < 0.05$, ***$p < 0.01$. 'R' represents values that have been suppressed to comply with the secure data service non-disclosure restrictions. Source: Next Steps.

Table C.7 Differences in sample means by anticipated discrimination: Bangladeshi

Variable	All	Anticipate discrimination: Yes/ Don't Know	Anticipate discrimination: No
Outcomes			
Total GCSE and equivalents new style point score	370.39	396.14	354.81**
Std GCSE points score	-0.13	0.04	-0.24**
Achieved 5 or more GCSEs	0.60	0.67	0.55**
Achieved 5 or more GCSEs incl. Eng, Maths	0.42	0.55	0.35***
Achieved A*-C in GCSE English	0.56	0.64	0.51**
Achieved A*-C in GCSE Maths	0.50	0.61	0.43***
Higher Science std score	-0.22	0.02	-0.37***
Achieved any A*-A	0.36	0.42	0.32**
Individual Characteristics			
Female	0.58	0.54	0.61
English first/main language	0.59	0.64	0.56
No special education needs	0.93	R	R
Region of Residence			
North East	0.03	R	R
North West	0.09	0.10	0.09
Yorkshire and The Humber	0.07	R	R
East Midlands	R	R	R
West Midlands	0.11	0.08	0.12**
East of England	0.10	0.13	0.08
London	0.55	0.55	0.55
South East	0.03	R	R
South West	R	R	R
Family Characteristics			
Number of siblings	3.18	3.18	3.19
Two-parent family	0.86	0.86	0.86
Homeowner	0.46	0.49	0.45
Not eligible to FSM	0.41	0.41	0.42
Parental Characteristics			
Parent(s) degree	R	R	R
Parent(s) A-levels/HE below degree	R	R	R
Parent(s) GCSE	0.05	R	R
Parent(s) low/no qualification	0.91	0.89	0.92
Parent(s) work	0.39	0.39	0.39
Parent(s) in professional occupations	0.16	0.17	0.15
Parent(s) SOC missing	0.28	0.28	0.28
Household income over £21k	0.13	R	R
Household income missing	0.41	0.41	0.40
Parent(s) aged <40	0.53	0.50	0.54
Parent(s) in very good health	0.43	0.44	0.43
Parental aspirations and involvement			
Parent would like YP to continue in FTE	0.95	R	R
Higher education very likely (Parent)	0.55	0.58	0.53
Very involved in school life (Parent)	0.37	0.39	0.35
Personality and Future Orientation			
Future-oriented	0.61	0.62	0.60
Raising family important	0.86	0.87	0.86
Job helping other important	0.53	0.56	0.52
Well paying job important	0.73	0.75	0.72
Being own boss important	0.36	0.35	0.37
Interesting job important	0.64	0.63	0.65
Job with promotion important	0.70	0.69	0.71
Job with regular hours important	0.49	0.48	0.50
Has ideas for career	0.62	0.65	0.60
Bullying			
Not bullied	0.67	0.68	0.66
Bullying missing	0.05	R	R
Prior achievement			
Std KS2 average scores	-0.14	0.13	-0.31***
Unweighted count	477	183	294

Notes: Survey weights applied. For the t-test of differences in means:*$p < 0.10$,**$p < 0.05$, ***$p < 0.01$. 'R' represents values that have been suppressed to comply with the secure data service non-disclosure restrictions. Source: Next Steps.

Table C.8 Differences in sample means by anticipated discrimination: Black Caribbean

Variable	All	Anticipate discrimination Yes/ Don't Know	No
Outcomes			
Total GCSE and equivalents new style point score	335.52	344.84	325.84
Std GCSE points score	-0.37	-0.30	-0.43
Achieved 5 or more GCSEs	0.46	0.45	0.48
Achieved 5 or more GCSEs incl. Eng, Maths	0.31	0.27	0.34
Achieved A*-C in GCSE English	0.54	0.61	0.47**
Achieved A*-C in GCSE Maths	0.36	0.33	0.39
Higher Science std score	-0.39	-0.33	-0.45
Achieved any A*-A	0.22	0.20	0.24
Individual Characteristics			
Female	0.55	0.58	0.53
English first/main language	R	R	R
No special education needs	0.78	0.82	0.74
Region of Residence			
North East	R	R	R
North West	R	R	R
Yorkshire and The Humber	0.03	R	R
East Midlands	0.05	R	R
West Midlands	0.19	0.21	0.16
East of England	R	R	R
London	0.60	0.62	0.58
South East	R	R	R
South West	R	R	R
Family Characteristics			
Number of siblings	1.46	1.33	1.60**
Two-parent family	0.40	0.40	0.39
Homeowner	0.45	0.49	0.41
Not eligible to FSM	0.81	0.85	0.76*
Parental Characteristics			
Parent(s) degree	0.08	R	R
Parent(s) A-levels/HE below degree	0.33	0.37	0.29
Parent(s) GCSE	0.29	0.26	0.31
Parent(s) low/no qualification	0.30	0.27	0.33
Parent(s) work	0.77	0.81	0.73
Parent(s) in professional occupations	0.22	0.23	0.22
Parent(s) SOC missing	0.06	R	R
Household income over £21k	0.43	0.46	0.40
Household income missing	0.24	0.32	0.16***
Parent(s) aged <40	0.38	0.34	0.42
Parent(s) in very good health	0.48	0.48	0.48
Parental aspirations and involvement			
Parent would like YP to continue in FTE	0.92	0.93	0.91
Higher education very likely (Parent)	0.44	0.40	0.48
Very involved in school life (Parent)	0.43	0.44	0.41
Personality and Future Orientation			
Future-oriented	0.75	0.79	0.72
Raising family important	0.85	0.85	0.84
Job helping other important	0.52	0.51	0.52
Well paying job important	0.80	0.83	0.77
Being own boss important	0.35	0.33	0.37
Interesting job important	0.67	0.73	0.59**
Job with promotion important	0.71	0.73	0.70
Job with regular hours important	0.48	0.46	0.49
Has ideas for career	0.82	0.83	0.82
Bullying			
Not bullied	0.57	0.54	0.60
Bullying missing	R	R	R
Prior achievement			
Std KS2 average scores	-0.11	-0.04	-0.19
Unweighted count	366	196	170

Notes: Survey weights applied. For the t-test of differences in means:*$p < 0.10$,**$p < 0.05$, ***$p < 0.01$. 'R' represents values that have been suppressed to comply with the secure data service non-disclosure restrictions. Source: Next Steps.

Table C.9 Differences in sample means by anticipated discrimination: Black African

Variable	All	Anticipate discrimination Yes/ Don't Know	No
Outcomes			
Total GCSE and equivalents new style point score	384.91	381.47	388.69
Std GCSE points score	-0.04	-0.06	-0.01
Achieved 5 or more GCSEs	0.61	0.65	0.57
Achieved 5 or more GCSEs incl. Eng, Maths	0.45	0.48	0.43
Achieved A*-C in GCSE English	0.62	0.66	0.57
Achieved A*-C in GCSE Maths	0.54	0.55	0.52
Higher Science std score	0.01	-0.02	0.03
Achieved any A*-A	0.35	0.37	0.34
Individual Characteristics			
Female	0.53	0.56	0.50
English first/main language	0.77	0.73	0.81
No special education needs	0.90	R	R
Region of Residence			
North East	R	R	R
North West	R	R	R
Yorkshire and The Humber	R	R	R
East Midlands	R	R	R
West Midlands	R	R	R
East of England	R	R	R
London	0.77	0.77	0.77
South East	0.06	R	R
South West	R	R	R
Family Characteristics			
Number of siblings	2.21	2.17	2.25
Two-parent family	0.58	0.62	0.54
Homeowner	0.32	0.34	0.31
Not eligible to FSM	0.63	0.62	0.64
Parental Characteristics			
Parent(s) degree	0.16	0.16	0.15
Parent(s) A-levels/HE below degree	0.23	0.24	0.21
Parent(s) GCSE	0.13	0.12	0.14
Parent(s) low/no qualification	0.47	0.47	0.48
Parent(s) work	0.60	0.61	0.58
Parent(s) in professional occupations	0.32	0.33	0.31
Parent(s) SOC missing	0.18	0.17	0.19
Household income over £21k	0.31	0.34	0.27
Household income missing	0.23	0.23	0.24
Parent(s) aged <40	0.37	0.33	0.424*
Parent(s) in very good health	0.64	0.64	0.63
Parental aspirations and involvement			
Parent would like YP to continue in FTE	R	R	R
Higher education very likely (Parent)	0.78	0.78	0.78
Very involved in school life (Parent)	0.53	0.51	0.55
Personality and Future Orientation			
Future-oriented	0.75	0.75	0.75
Raising family important	0.90	R	R
Job helping other important	0.46	0.46	0.46
Well paying job important	0.77	0.79	0.74
Being own boss important	0.46	0.50	0.41
Interesting job important	0.70	0.67	0.72
Job with promotion important	0.76	0.75	0.77
Job with regular hours important	0.55	0.58	0.52
Has ideas for career	0.78	0.74	0.82
Bullying			
Not bullied	0.61	0.59	0.62
Bullying missing	R	R	R
Prior achievement			
Std KS2 average scores	-0.18	-0.18	-0.19
Unweighted count	289	149	140

Notes: Survey weights applied. For the t-test of differences in means:*$p < 0.10$,**$p < 0.05$, ***$p < 0.01$. 'R' represents values that have been suppressed to comply with the secure data service non-disclosure restrictions. Source: Next Steps.

Table C.10 Differences in sample means by anticipated discrimination: Mixed/Other

Variable	All	Anticipate discrimination Yes/ Don't Know	No
Outcomes			
GCSE total point score	391.94	382.37	396.54
Std GCSE total score	0.01	-0.05	0.04
Achieved 5 or more GCSEs	0.62	0.62	0.62
Achieved 5 or more GCSEs incl. Eng, Maths	0.52	0.50	0.52
Achieved A*-C in GCSE English	0.66	0.64	0.67
Achieved A*-C in GCSE Maths	0.59	0.59	0.60
Higher Science std score	0.07	0.02	0.10
Achieved any A*-A	0.43	0.40	0.44
Individual Characteristics			
Female	0.51	0.49	0.51
English first/main language	0.88	0.87	0.89
No special education needs	0.83	0.78	0.86*
Region of Residence			
North East	0.03	R	R
North West	0.09	0.10	0.09
Yorkshire and The Humber	0.08	0.09	0.07
East Midlands	0.07	0.07	0.07
West Midlands	0.11	0.14	0.10
East of England	0.08	0.08	0.07
London	0.40	0.36	0.42
South East	0.10	0.10	0.09
South West	0.05	R	R
Family Characteristics			
Number of siblings	1.59	1.54	1.61
Two-parent family	0.65	0.63	0.66
Homeowner	0.56	0.55	0.56
Not eligible to FSM	0.77	0.78	0.77
Parental Characteristics			
Parent(s) degree	0.09	0.10	0.08
Parent(s) A-levels/HE below degree	0.20	0.18	0.21
Parent(s) GCSE	0.19	0.20	0.19
Parent(s) low/no qualification	0.51	0.50	0.51
Parent(s) work	0.71	0.66	0.73
Parent(s) in professional occupations	0.35	0.34	0.36
Parent(s) SOC missing	0.07	0.06	0.07
Household income over £21k	0.41	0.39	0.42
Household income missing	0.25	0.29	0.24
Parent(s) aged <40	0.40	0.40	0.40
Parent(s) in very good health	0.58	0.52	0.61*
Parental aspirations and involvement			
Parent would like YP to continue in FTE	0.90	0.89	0.91
Higher education very likely (Parent)	0.54	0.56	0.53
Very involved in school life (Parent)	0.29	0.27	0.30
Personality and Future Orientation			
Future-oriented	0.71	0.71	0.72
Raising family important	0.90	0.92	0.89
Job helping other important	0.40	0.44	0.37
Well paying job important	0.71	0.70	0.71
Being own boss important	0.29	0.35	0.26**
Interesting job important	0.71	0.73	0.70
Job with promotion important	0.66	0.66	0.66
Job with regular hours important	0.45	0.41	0.47
Has ideas for career	0.80	0.80	0.81
Bullying			
Not bullied	0.57	0.52	0.59
Bullying missing	R	R	R
Prior achievement			
Std KS2 average scores	0.14	0.08	0.17
Unweighted count	795	278	517

Notes: Survey weights applied. For the t-test of differences in means:*$p < 0.10$,**$p < 0.05$, ***$p < 0.01$. 'R' represents values that have been suppressed to comply with the secure data service non-disclosure restrictions. Source: Next Steps.

Table C.11 Anticipated discrimination and GCSE achievement - summarised results, models with and without school fixed effects (OLS and LPM)

	Educational Outcomes						
	Total GCSE score	**5+ A*-C**	**5+ A*-C Eng,Math**	**A*-C Eng**	**A*-C Math**	**Science Score**	**Any A*-A**
Model 1							
Anticipate Disc	0.078**	0.050**	0.062***	0.047**	0.064***	0.110***	0.032
	(0.038)	(0.020)	(0.019)	(0.019)	(0.018)	(0.039)	(0.020)
Standard controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Model 1: Without SFE							
Anticipate Disc	0.067**	0.048***	0.067***	0.058**	0.061***	0.101***	0.027
	(0.033)	(0.017)	(0.016)	(0.016)	(0.016)	(0.034)	(0.017)
Standard controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	No	No	No	No	No	No	No
Model 2							
Anticipate Disc	0.084**	0.054***	0.064***	0.050***	0.066***	0.116***	0.030
	(0.036)	(0.019)	(0.018)	(0.018)	(0.017)	(0.037)	(0.019)
Standard controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Extended controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Model 2: Without SFE							
Anticipate Disc	0.069**	0.050***	0.067***	0.058***	0.060***	0.102***	0.023
	(0.031)	(0.016)	(0.016)	(0.015)	(0.015)	(0.032)	(0.016)
Standard controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Extended controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	No	No	No	No	No	No	No
N	3315	3315	3315	3315	3315	3315	3315

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and controls are described in the text in section 4.4.1. Pooled sample of ethnic minority students. Source: Next Steps.

Table C.12 Anticipated discrimination and GCSE achievement - complete results Model 1 (OLS and LPM)

	Educational Outcomes						
	Total GCSE score	**5+ A*-C**	**5+ A*-C Eng,Math**	**A*-C Eng**	**A*-C Math**	**Science score**	**Any A*-A**
Anticipate Disc	0.078**	0.050**	0.062***	0.047**	0.064***	0.110***	0.032
	(0.038)	(0.020)	(0.019)	(0.019)	(0.018)	(0.039)	(0.020)
Female	0.279***	0.137***	0.071***	0.167***	-0.001	0.110**	0.140***
	(0.042)	(0.022)	(0.025)	(0.023)	(0.024)	(0.044)	(0.023)
Indian	0.140**	0.031	0.055	0.031	0.063*	0.075	0.027
	(0.060)	(0.031)	(0.037)	(0.031)	(0.033)	(0.058)	(0.033)
Pakistani	-0.120*	-0.060*	-0.053	-0.073**	-0.067*	-0.120*	-0.065*
	(0.062)	(0.032)	(0.036)	(0.034)	(0.034)	(0.068)	(0.036)
Bangladeshi	0.069	0.048	0.050	0.046	0.029	0.038	0.009
	(0.079)	(0.038)	(0.041)	(0.040)	(0.041)	(0.078)	(0.043)
Black Caribbean	-0.258***	-0.127***	-0.174***	-0.104***	-0.170***	-0.282***	-0.152***
	(0.062)	(0.034)	(0.034)	(0.034)	(0.034)	(0.066)	(0.034)
Black African	-0.010	-0.016	-0.047	-0.015	-0.045	-0.050	-0.059
	(0.079)	(0.042)	(0.042)	(0.041)	(0.041)	(0.079)	(0.040)
EFL	0.127**	0.048*	0.038	0.062**	0.029	0.146**	0.035
	(0.054)	(0.028)	(0.026)	(0.025)	(0.025)	(0.057)	(0.027)
No SEN	0.721***	0.334***	0.309***	0.338***	0.320***	0.659***	0.219***
	(0.061)	(0.029)	(0.027)	(0.030)	(0.029)	(0.064)	(0.026)
Number of siblings	-0.041***	-0.014*	-0.018**	-0.019***	-0.017**	-0.033**	-0.008
	(0.014)	(0.007)	(0.007)	(0.007)	(0.007)	(0.015)	(0.007)
Two-parent family	0.223***	0.089***	0.077***	0.065**	0.101***	0.194***	0.049*
	(0.057)	(0.027)	(0.028)	(0.028)	(0.030)	(0.055)	(0.027)
Homeowner	0.121**	0.070***	0.016	0.026	0.002	0.107**	0.038
	(0.050)	(0.026)	(0.025)	(0.025)	(0.026)	(0.052)	(0.026)
No FSM	0.049	0.010	0.020	-0.008	0.020	0.028	0.013
	(0.051)	(0.029)	(0.027)	(0.029)	(0.030)	(0.054)	(0.025)
Parent(s) degree	0.295***	0.139***	0.125***	0.091**	0.123***	0.316***	0.133***
	(0.078)	(0.038)	(0.042)	(0.037)	(0.041)	(0.078)	(0.044)
Parent(s) A-levels	0.124**	0.062**	0.081***	0.085***	0.089***	0.161***	0.054*
	(0.054)	(0.029)	(0.030)	(0.030)	(0.030)	(0.054)	(0.029)
Parent(s) GCSE	0.042	0.035	0.051**	0.029	0.059**	0.032	0.019
	(0.044)	(0.024)	(0.026)	(0.025)	(0.025)	(0.050)	(0.025)
Parent(s) work	0.069	0.033	0.025	0.059*	0.024	0.066	0.026
	(0.055)	(0.030)	(0.031)	(0.030)	(0.032)	(0.060)	(0.026)
Parent(s) in professional occupations	0.115***	0.016	0.048*	0.020	0.045*	0.084*	0.041
	(0.043)	(0.024)	(0.026)	(0.024)	(0.027)	(0.046)	(0.026)
Parent(s) SOC missing	-0.042	-0.019	-0.008	-0.052	-0.028	-0.047	0.017
	(0.069)	(0.034)	(0.031)	(0.036)	(0.032)	(0.078)	(0.035)
Parent(s) aged <40	-0.048	-0.043**	-0.050**	-0.032*	-0.036*	-0.043	-0.059***
	(0.036)	(0.019)	(0.020)	(0.019)	(0.020)	(0.037)	(0.021)
Parent(s) in very good health	-0.005	-0.000	0.017	-0.003	0.022	0.028	0.017
	(0.038)	(0.020)	(0.021)	(0.019)	(0.021)	(0.042)	(0.022)
Household income over £21k	0.053	0.054**	0.076***	0.066**	0.044	0.047	0.049*
	(0.051)	(0.027)	(0.029)	(0.026)	(0.028)	(0.051)	(0.027)
Household income missing	-0.087*	-0.018	0.005	-0.008	-0.009	-0.077	-0.001
	(0.048)	(0.024)	(0.026)	(0.024)	(0.023)	(0.049)	(0.023)
Region of residence	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	3315	3315	3315	3315	3315	3315	3315
R2	0.448	0.341	0.343	0.343	0.347	0.400	0.338

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score and science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. Model 1 includes the standard set of controls. Constant included in all regressions but is not reported to comply with the secure data service non-disclosure guidelines. Omitted groups: *Gender:* Male; *Ethnicity:* Mixed/Other; *EFL*: no EFL; *No SEN:* has SEN; *Two-parent family*: single parent family; *Homeowner*: not homeowner; *No FSM: eligible for FSM*;*Parent(s) education:* low/no qualifications; *Parent(s) work: not working*; *Parent(s) occupation: not in managerial/professional occupations*; *Parent(s) age:* $>=40$; *Parent(s) health: not very good health*; *Household income: lower than £21k*; *Region of residence:* London. Source: Next Steps.

Table C.13 Anticipated discrimination and GCSE achievement - complete results Model 2 (OLS and LPM)

	Educational Outcomes						
	Total GCSE score	**5+ A*-C**	**5+ A*-C Eng,Math**	**A*-C Eng**	**A*-C Math**	**Science score**	**Any A*-A**
Anticipate Disc	0.084**	0.054***	0.064***	0.050***	0.066***	0.116***	0.030
	(0.036)	(0.019)	(0.018)	(0.018)	(0.017)	(0.037)	(0.019)
Female	0.210***	0.109***	0.034	0.132***	-0.038*	0.035	0.108***
	(0.039)	(0.022)	(0.024)	(0.023)	(0.023)	(0.041)	(0.023)
Indian	0.109**	0.014	0.038	0.011	0.047	0.042	0.015
	(0.054)	(0.028)	(0.033)	(0.029)	(0.030)	(0.054)	(0.031)
Pakistani	-0.109*	-0.055*	-0.048	-0.073**	-0.062*	-0.111*	-0.059*
	(0.061)	(0.031)	(0.034)	(0.032)	(0.033)	(0.065)	(0.036)
Bangladeshi	0.061	0.044	0.044	0.036	0.025	0.031	0.006
	(0.075)	(0.035)	(0.038)	(0.038)	(0.038)	(0.073)	(0.041)
Black Caribbean	-0.233***	-0.115***	-0.157***	-0.097***	-0.155***	-0.254***	-0.136***
	(0.059)	(0.034)	(0.033)	(0.034)	(0.033)	(0.063)	(0.034)
Black African	-0.056	-0.041	-0.073*	-0.049	-0.069*	-0.099	-0.078*
	(0.076)	(0.042)	(0.041)	(0.039)	(0.040)	(0.077)	(0.040)
EFL	0.102**	0.037	0.027	0.054**	0.018	0.120**	0.025
	(0.051)	(0.026)	(0.026)	(0.024)	(0.024)	(0.054)	(0.027)
No SEN	0.549***	0.251***	0.218***	0.252***	0.240***	0.484***	0.147***
	(0.057)	(0.028)	(0.026)	(0.029)	(0.028)	(0.061)	(0.026)
Number of siblings	-0.037***	-0.011	-0.017**	-0.017**	-0.015**	-0.030**	-0.007
	(0.013)	(0.007)	(0.007)	(0.007)	(0.007)	(0.015)	(0.007)
Two-parent family	0.196***	0.076***	0.062**	0.050*	0.088***	0.167***	0.036
	(0.055)	(0.026)	(0.027)	(0.027)	(0.028)	(0.052)	(0.026)
Homeowner	0.103**	0.062**	0.008	0.018	-0.006	0.089*	0.032
	(0.047)	(0.024)	(0.024)	(0.023)	(0.024)	(0.048)	(0.025)
No FSM	0.058	0.013	0.022	-0.002	0.024	0.036	0.013
	(0.048)	(0.027)	(0.026)	(0.027)	(0.029)	(0.050)	(0.024)
Parent(s) degree	0.215***	0.100***	0.084**	0.053	0.084**	0.234***	0.099**
	(0.076)	(0.036)	(0.041)	(0.036)	(0.039)	(0.076)	(0.043)
Parent(s) A-levels	0.095*	0.048*	0.066**	0.070***	0.077***	0.131**	0.042
	(0.050)	(0.027)	(0.027)	(0.027)	(0.028)	(0.051)	(0.028)
Parent(s) GCSE	0.049	0.037	0.053**	0.031	0.061**	0.036	0.020
	(0.044)	(0.024)	(0.026)	(0.025)	(0.025)	(0.049)	(0.025)
Parent(s) work	0.054	0.026	0.018	0.052*	0.018	0.053	0.022
	(0.049)	(0.028)	(0.028)	(0.029)	(0.030)	(0.055)	(0.025)
Parent(s) in professional occupations	0.099**	0.007	0.039	0.012	0.039	0.066	0.034
	(0.041)	(0.023)	(0.026)	(0.023)	(0.026)	(0.044)	(0.025)
Parent(s) SOC missing	-0.037	-0.016	-0.003	-0.049	-0.025	-0.039	0.022
	(0.063)	(0.031)	(0.029)	(0.034)	(0.029)	(0.071)	(0.033)
Parent(s) aged <40	-0.032	-0.036**	-0.042**	-0.025	-0.028	-0.026	-0.051**
	(0.034)	(0.018)	(0.019)	(0.018)	(0.018)	(0.034)	(0.020)
Parent(s) in very good health	-0.041	-0.017	-0.003	-0.022	0.003	-0.012	0.001
	(0.037)	(0.019)	(0.021)	(0.019)	(0.021)	(0.041)	(0.022)
Household income over £21k	0.037	0.049*	0.067**	0.061**	0.036	0.029	0.041
	(0.049)	(0.025)	(0.028)	(0.025)	(0.027)	(0.048)	(0.025)
Household income missing	-0.075*	-0.010	0.010	-0.001	-0.005	-0.067	0.002
	(0.045)	(0.022)	(0.024)	(0.023)	(0.022)	(0.045)	(0.023)
Parent would like YP to continue in FTE	0.280***	0.126***	0.121***	0.173***	0.123***	0.208***	0.064
	(0.072)	(0.042)	(0.040)	(0.041)	(0.041)	(0.071)	(0.039)
Higher education very likely (Parent)	0.345***	0.167***	0.203***	0.175***	0.185***	0.390***	0.170***
	(0.037)	(0.021)	(0.020)	(0.020)	(0.019)	(0.039)	(0.020)
Very involved in school life (Parent)	-0.008	0.009	-0.010	0.017	-0.011	-0.019	0.000
	(0.037)	(0.018)	(0.020)	(0.019)	(0.019)	(0.037)	(0.019)
Not bullied	0.123***	0.070***	0.055***	0.068***	0.042**	0.119***	0.016
	(0.034)	(0.018)	(0.019)	(0.018)	(0.020)	(0.037)	(0.017)
Bullying missing	-0.110	-0.039	-0.031	-0.035	-0.065	-0.204	0.025
	(0.127)	(0.055)	(0.065)	(0.057)	(0.062)	(0.141)	(0.067)
Future-oriented	0.228***	0.096***	0.101***	0.115***	0.084***	0.204***	0.088***
	(0.038)	(0.021)	(0.020)	(0.020)	(0.020)	(0.041)	(0.020)
Raising family important	0.066	0.064**	0.036	0.033	0.030	0.040	0.027
	(0.058)	(0.029)	(0.029)	(0.029)	(0.030)	(0.060)	(0.028)
Job helping other important	-0.076**	-0.053***	-0.042**	-0.023	-0.040**	-0.068*	-0.032*
	(0.034)	(0.018)	(0.018)	(0.017)	(0.018)	(0.035)	(0.017)
Well paying job important	-0.073*	-0.045**	-0.044**	-0.028	-0.051**	-0.068	-0.027
	(0.042)	(0.020)	(0.022)	(0.019)	(0.021)	(0.044)	(0.023)
Being own boss important	-0.233***	-0.101***	-0.123***	-0.101***	-0.119***	-0.244***	-0.119***
	(0.033)	(0.019)	(0.020)	(0.018)	(0.020)	(0.035)	(0.018)
Interesting job important	0.058*	0.032*	0.037**	0.019	0.011	0.059	0.039**
	(0.033)	(0.019)	(0.017)	(0.019)	(0.018)	(0.036)	(0.019)
Job with promotion important	0.008	0.010	-0.010	-0.002	-0.008	0.006	0.002
	(0.038)	(0.021)	(0.022)	(0.021)	(0.022)	(0.040)	(0.020)
Job with regular hours important	0.007	0.010	0.020	0.024	0.023	0.042	-0.002
	(0.032)	(0.016)	(0.019)	(0.018)	(0.019)	(0.038)	(0.018)
Has ideas for career	0.107***	0.065***	0.043**	0.033	0.059***	0.114***	0.011
	(0.038)	(0.020)	(0.021)	(0.022)	(0.021)	(0.041)	(0.021)
Region of residence	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	3315	3315	3315	3315	3315	3315	3315
R2	0.508	0.402	0.409	0.408	0.405	0.465	0.384

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. Model 2 includes the extended set of controls. Constant included in all regressions but is not reported to comply with the secure data service non-disclosure guidelines. Omitted groups: *Gender:* Male; *Ethnicity:* Mixed/Other; *EFL*: no EFL; *No SEN:* has SEN; *Two-parent family*: single parent family; *Homeowner*: not homeowner; *No FSM:* eligible for FSM;*Parent(s) education:* low/no qualifications; *Parent(s) work:* not working; *Parent(s) occupation:* not in managerial/professional occupations; *Parent(s) age:* ≥ 40; *Parent(s) health:* not very good; *Household income:* lower than £21k; *Parent(s) aspirations:* would not like child to continue in FTE ; *HE:* HE not very likely; *Parent(s) involvement:* not very involved in school life; *Bullying :* bullied in past 12 months; *Future-oriented:* not future oriented;*Raising family*: not very important ; *Job characteristics:* respective characteristic not important; *Career ambition:* no career ideas;*Region of residence:* London. Source: Next Steps.

Table C.14 Anticipated discrimination and GCSE achievement - complete Results Model 2 (OLS and LPM); Indian group only

	Educational Outcomes						
	Total GCSE score	**5+ A*-C**	**5+ A*-C Eng,Math**	**A*-C Eng**	**A*-C Math**	**Science score**	**Any A*-A**
Anticipate Disc	0.065	0.091*	0.081	0.115**	0.069	0.102	0.010
	(0.100)	(0.048)	(0.050)	(0.050)	(0.048)	(0.095)	(0.060)
Female	0.192**	0.106**	0.040	0.128**	-0.031	-0.011	0.101*
	(0.091)	(0.051)	(0.061)	(0.054)	(0.056)	(0.090)	(0.059)
EFL	0.236	0.104	0.101	0.117	0.089	0.225	0.099
	(0.157)	(0.077)	(0.077)	(0.084)	(0.067)	(0.165)	(0.089)
No SEN	0.672***	0.270***	0.260***	0.268***	0.281***	0.444***	0.183**
	(0.158)	(0.083)	(0.085)	(0.092)	(0.086)	(0.142)	(0.077)
Number of siblings	-0.063	-0.002	-0.022	0.003	-0.016	-0.063	-0.029
	(0.044)	(0.022)	(0.023)	(0.023)	(0.023)	(0.043)	(0.024)
Two-parent family	0.276**	0.032	0.094	0.040	0.162**	0.203**	0.131*
	(0.123)	(0.078)	(0.073)	(0.092)	(0.075)	(0.098)	(0.074)
Homeowner	-0.082	0.101	-0.033	-0.028	-0.059	-0.093	-0.058
	(0.156)	(0.079)	(0.077)	(0.100)	(0.067)	(0.132)	(0.076)
No FSM	0.055	0.063	0.071	0.063	0.045	0.080	0.022
	(0.143)	(0.087)	(0.096)	(0.086)	(0.107)	(0.139)	(0.086)
Parent(s) degree	0.262	-0.009	0.020	-0.027	0.023	0.317*	0.147*
	(0.171)	(0.068)	(0.070)	(0.074)	(0.074)	(0.181)	(0.082)
Parent(s) A-levels	0.075	0.042	0.079	0.083	0.108	0.133	0.055
	(0.127)	(0.071)	(0.069)	(0.058)	(0.074)	(0.124)	(0.074)
Parent(s) GCSE	0.041	0.036	0.091	-0.007	0.100*	0.047	0.012
	(0.104)	(0.062)	(0.065)	(0.062)	(0.055)	(0.093)	(0.061)
Parent(s) work	0.088	-0.007	-0.006	0.043	0.079	0.051	0.017
	(0.155)	(0.087)	(0.089)	(0.084)	(0.089)	(0.134)	(0.104)
Parent(s) in professional occupations	0.136	0.039	0.050	0.055	0.008	0.096	0.012
	(0.090)	(0.050)	(0.060)	(0.053)	(0.051)	(0.087)	(0.059)
Parent(s) SOC missing	0.172	0.078	0.123	-0.031	0.143	-0.004	0.138
	(0.211)	(0.123)	(0.130)	(0.134)	(0.114)	(0.226)	(0.135)
Parent(s) aged <40	-0.090	-0.027	-0.032	-0.040	-0.039	-0.028	-0.071
	(0.083)	(0.052)	(0.050)	(0.050)	(0.052)	(0.078)	(0.054)
Parent(s) in very good health	-0.029	0.008	-0.018	-0.011	-0.035	-0.024	0.006
	(0.091)	(0.047)	(0.054)	(0.045)	(0.047)	(0.092)	(0.062)
Household income over £21k	0.293***	0.189***	0.199***	0.163**	0.119**	0.242**	0.090
	(0.108)	(0.059)	(0.057)	(0.065)	(0.054)	(0.103)	(0.059)
Household income missing	0.111	0.086*	0.101**	0.115*	0.053	0.144	0.090
	(0.099)	(0.052)	(0.049)	(0.059)	(0.054)	(0.092)	(0.057)
Parent would like YP to continue in FTE	0.192	0.128	0.193	0.259**	0.067	0.071	0.188**
	(0.170)	(0.114)	(0.133)	(0.112)	(0.150)	(0.181)	(0.095)
Higher education very likely (Parent)	0.277***	0.150**	0.171***	0.146**	0.165***	0.328***	0.172***
	(0.104)	(0.066)	(0.063)	(0.057)	(0.056)	(0.093)	(0.055)
Very involved in school life (Parent)	0.098	0.052	0.030	0.047	0.047	0.102	0.022
	(0.084)	(0.050)	(0.057)	(0.049)	(0.056)	(0.080)	(0.055)
Not bullied	0.181**	0.079*	0.042	0.050	0.067	0.140*	0.005
	(0.081)	(0.045)	(0.048)	(0.038)	(0.051)	(0.084)	(0.046)
Bullying missing	-0.101	-0.106	-0.141	-0.096	-0.107	-0.354	0.121
	(0.362)	(0.184)	(0.179)	(0.178)	(0.128)	(0.366)	(0.235)
Future-oriented	0.041	0.058	0.015	0.051	0.010	0.032	0.037
	(0.089)	(0.045)	(0.041)	(0.042)	(0.047)	(0.087)	(0.049)
Raising family important	0.015	0.097	0.077	0.051	0.002	0.071	0.053
	(0.179)	(0.108)	(0.102)	(0.081)	(0.101)	(0.200)	(0.095)
Job helping other important	0.029	-0.044	-0.001	-0.011	-0.013	-0.041	0.010
	(0.078)	(0.050)	(0.045)	(0.045)	(0.049)	(0.076)	(0.049)
Well paying job important	0.041	-0.047	-0.011	0.014	-0.072	0.062	-0.003
	(0.087)	(0.052)	(0.053)	(0.044)	(0.047)	(0.084)	(0.055)
Being own boss important	-0.327***	-0.101**	-0.072	-0.033	-0.126***	-0.275***	-0.143***
	(0.081)	(0.046)	(0.051)	(0.044)	(0.040)	(0.079)	(0.050)
Interesting job important	0.136	0.046	0.060	0.019	0.021	0.091	0.057
	(0.091)	(0.052)	(0.048)	(0.044)	(0.049)	(0.089)	(0.056)
Job with promotion important	0.084	-0.014	-0.030	-0.039	0.046	-0.022	0.044
	(0.083)	(0.048)	(0.052)	(0.041)	(0.047)	(0.082)	(0.045)
Job with regular hours important	-0.227***	-0.051	-0.078	-0.109**	-0.029	-0.136*	-0.068
	(0.081)	(0.041)	(0.049)	(0.050)	(0.049)	(0.080)	(0.047)
Has ideas for career	0.055	-0.007	0.016	0.019	0.042	0.009	0.025
	(0.081)	(0.048)	(0.051)	(0.049)	(0.047)	(0.082)	(0.051)
Region of residence	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	712	712	712	712	712	712	712
R2	0.660	0.546	0.564	0.551	0.540	0.611	0.548

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. Model 2 includes the extended set of controls. Constant included in all regressions but is not reported to comply with the secure data service non-disclosure guidelines. Omitted groups: *Gender:* Male; *Ethnicity:* Mixed/Other; *EFL*: no EFL; *No SEN:* has SEN; *Two-parent family*: single parent family; *Homeowner*: not homeowner; *No FSM:* eligible for FSM;*Parent(s) education:* low/no qualifications; *Parent(s) work:* not working; *Parent(s) occupation:* not in managerial/professional occupations; *Parent(s) age:* ≥ 40; *Parent(s) health:* not very good; *Household income:* lower than £21k; *Parent(s) aspirations:* would not like child to continue in FTE ; *HE:* HE not very likely; *Parent(s) involvement:* not very involved in school life; *Bullying:* bullied in past 12 months; *Future-oriented:* not future oriented;*Raising family*: not very important ; *Job characteristics:* respective characteristic not important; *Career ambition:* no career ideas;*Region of residence:* London. Source: Next Steps.

Table C.15 Anticipated discrimination and GCSE achievement - complete Results Model 2 (OLS and LPM); Pakistani group only

	Educational Outcomes						
	Total GCSE score	**5+ A*-C**	**5+ A*-C Eng,Math**	**A*-C Eng**	**A*-C Math**	**Science score**	**Any A*-A**
Anticipate Disc	0.244**	0.129**	0.152***	0.104**	0.149***	0.288***	0.089**
	(0.096)	(0.050)	(0.051)	(0.048)	(0.051)	(0.099)	(0.045)
Female	0.251**	0.186***	0.049	0.134*	-0.041	0.076	0.122**
	(0.111)	(0.065)	(0.068)	(0.070)	(0.065)	(0.137)	(0.058)
EFL	0.178	0.099	0.034	0.055	0.038	0.200	0.055
	(0.121)	(0.068)	(0.053)	(0.058)	(0.052)	(0.145)	(0.062)
No SEN	0.595***	0.221***	0.208***	0.299***	0.253***	0.533***	0.179***
	(0.166)	(0.083)	(0.055)	(0.081)	(0.077)	(0.182)	(0.060)
Number of siblings	-0.047	-0.007	-0.023	-0.036**	-0.027	-0.029	-0.009
	(0.029)	(0.017)	(0.016)	(0.018)	(0.017)	(0.039)	(0.014)
Two-parent family	0.175	0.008	0.121*	0.074	0.105	0.123	-0.023
	(0.167)	(0.074)	(0.070)	(0.078)	(0.075)	(0.190)	(0.077)
Homeowner	0.044	0.066	0.027	0.011	-0.021	-0.045	-0.021
	(0.133)	(0.068)	(0.061)	(0.060)	(0.069)	(0.130)	(0.066)
No FSM	0.024	0.013	-0.007	-0.032	0.002	0.044	0.033
	(0.112)	(0.061)	(0.059)	(0.063)	(0.065)	(0.121)	(0.061)
Parent(s) degree	0.535**	0.249	0.158	0.193	0.238	0.343	0.222
	(0.255)	(0.153)	(0.166)	(0.167)	(0.145)	(0.213)	(0.141)
Parent(s) A-levels	0.235	0.002	0.049	0.107	0.022	0.278	0.121
	(0.245)	(0.119)	(0.096)	(0.125)	(0.098)	(0.225)	(0.113)
Parent(s) GCSE	0.199	0.159	0.192**	0.037	0.214**	0.194	0.094
	(0.189)	(0.097)	(0.089)	(0.111)	(0.093)	(0.212)	(0.087)
Parent(s) work	0.087	0.005	0.042	0.096	0.042	0.053	0.000
	(0.115)	(0.077)	(0.070)	(0.078)	(0.072)	(0.137)	(0.066)
Parent(s) in professional occupations	0.040	0.023	0.063	-0.038	0.074	-0.018	0.071
	(0.146)	(0.083)	(0.081)	(0.075)	(0.091)	(0.158)	(0.087)
Parent(s) SOC missing	-0.089	-0.053	0.022	-0.037	0.010	-0.070	0.060
	(0.147)	(0.073)	(0.076)	(0.073)	(0.076)	(0.180)	(0.074)
Parent(s) aged <40	0.078	-0.024	-0.050	0.000	-0.054	0.044	-0.004
	(0.095)	(0.048)	(0.048)	(0.051)	(0.048)	(0.100)	(0.046)
Parent(s) in very good health	-0.140	-0.072	-0.044	-0.079	-0.017	-0.127	0.016
	(0.111)	(0.054)	(0.049)	(0.054)	(0.051)	(0.115)	(0.048)
Household income over £21k	-0.029	-0.022	-0.016	0.003	-0.055	-0.090	0.056
	(0.150)	(0.080)	(0.074)	(0.082)	(0.089)	(0.164)	(0.086)
Household income missing	-0.106	-0.063	-0.047	-0.007	-0.038	-0.096	0.004
	(0.112)	(0.061)	(0.050)	(0.053)	(0.051)	(0.128)	(0.055)
Parent would like YP to continue in FTE	0.233	0.124	0.108	0.207**	0.138	0.215	-0.044
	(0.166)	(0.106)	(0.083)	(0.094)	(0.104)	(0.180)	(0.092)
Higher education very likely (Parent)	0.463***	0.227***	0.206***	0.192***	0.190***	0.505***	0.177***
	(0.094)	(0.056)	(0.052)	(0.054)	(0.059)	(0.105)	(0.045)
Very involved in school life (Parent)	-0.111	-0.064	-0.094*	-0.085	-0.072	-0.184	-0.066
	(0.126)	(0.061)	(0.053)	(0.057)	(0.056)	(0.112)	(0.047)
Not bullied	0.084	0.037	0.012	0.057	-0.020	0.063	-0.005
	(0.096)	(0.061)	(0.051)	(0.056)	(0.053)	(0.105)	(0.043)
Bullying missing	-0.476	-0.177	-0.173	-0.274	-0.264**	-0.811**	-0.019
	(0.382)	(0.153)	(0.165)	(0.173)	(0.120)	(0.337)	(0.164)
Future-oriented	0.231**	0.084	0.125**	0.132**	0.112**	0.194*	0.130***
	(0.092)	(0.056)	(0.050)	(0.053)	(0.049)	(0.116)	(0.046)
Raising family important	0.204	0.088	0.124*	0.055	0.151*	0.197	0.067
	(0.168)	(0.101)	(0.069)	(0.095)	(0.080)	(0.179)	(0.079)
Job helping other important	-0.108	-0.063	-0.025	-0.000	-0.026	-0.058	-0.003
	(0.093)	(0.059)	(0.048)	(0.054)	(0.051)	(0.084)	(0.040)
Well paying job important	-0.182*	-0.062	-0.062	-0.063	-0.124**	-0.188	0.005
	(0.105)	(0.057)	(0.056)	(0.059)	(0.055)	(0.119)	(0.068)
Being own boss important	-0.111	-0.057	-0.077	-0.117**	-0.108	-0.153	-0.088**
	(0.095)	(0.053)	(0.061)	(0.059)	(0.066)	(0.119)	(0.044)
Interesting job important	0.122	0.063	0.095*	0.079	0.076	0.206**	0.074
	(0.094)	(0.055)	(0.052)	(0.054)	(0.052)	(0.104)	(0.050)
Job with promotion important	-0.013	0.008	-0.070	-0.033	-0.064	-0.033	-0.058
	(0.109)	(0.055)	(0.051)	(0.060)	(0.056)	(0.117)	(0.052)
Job with regular hours important	0.051	0.017	0.012	0.050	-0.015	-0.014	-0.011
	(0.091)	(0.049)	(0.047)	(0.048)	(0.047)	(0.117)	(0.047)
Has ideas for career	0.153*	0.105**	0.061	0.091	0.047	0.246**	0.036
	(0.088)	(0.052)	(0.048)	(0.061)	(0.051)	(0.097)	(0.044)
Region of residence	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	676	676	676	676	676	676	676
R2	0.556	0.490	0.542	0.506	0.517	0.524	0.499

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. Model 2 includes the extended set of controls. Constant included in all regressions but is not reported to comply with the secure data service non-disclosure guidelines. Omitted groups: *Gender:* Male; *Ethnicity:* Mixed/Other; *EFL*: no EFL; *No SEN:* has SEN; *Two-parent family*: single parent family; *Homeowner*: not homeowner; *No FSM:* eligible for FSM;*Parent(s) education:* low/no qualifications; *Parent(s) work:* not working; *Parent(s) occupation:* not in managerial/professional occupations; *Parent(s) age:* ≥ 40; *Parent(s) health:* not very good; *Household income:* lower than £21k; *Parent(s) aspirations:* would not like child to continue in FTE ; *HE:* HE not very likely; *Parent(s) involvement:* not very involved in school life; *Bullying:* bullied in past 12 months; *Future-oriented:* not future oriented;*Raising family*: not very important ; *Job characteristics:* respective characteristic not important; *Career ambition:* no career ideas;*Region of residence:* London. Source: Next Steps.

Table C.16 Anticipated discrimination and GCSE achievement - complete Results Model 2 (OLS and LPM); Bangladeshi group only

	Educational Outcomes						
	Total GCSE score	**5+ A*-C**	**5+ A*-C Eng,Math**	**A*-C Eng**	**A*-C Math**	**Science score**	**Any A*-A**
Anticipate Disc	0.223*	0.064	0.143**	0.068	0.110*	0.340**	0.103
	(0.133)	(0.065)	(0.064)	(0.062)	(0.061)	(0.147)	(0.072)
Female	0.341***	0.204***	-0.050	0.114	-0.062	-0.050	0.061
	(0.128)	(0.066)	(0.100)	(0.094)	(0.107)	(0.220)	(0.094)
EFL	0.384***	0.172**	0.136**	0.183***	0.145**	0.429***	0.120**
	(0.142)	(0.072)	(0.066)	(0.057)	(0.061)	(0.111)	(0.058)
No SEN	0.498***	0.085	0.155	0.059	0.298***	0.409	0.010
	(0.177)	(0.130)	(0.096)	(0.119)	(0.097)	(0.288)	(0.096)
Number of siblings	0.012	0.008	-0.001	0.007	-0.004	0.028	-0.015
	(0.040)	(0.020)	(0.023)	(0.019)	(0.023)	(0.053)	(0.028)
Two-parent family	0.269	0.208*	0.132	0.120	0.157	0.229	0.147
	(0.233)	(0.114)	(0.123)	(0.114)	(0.132)	(0.246)	(0.130)
Homeowner	0.133	0.070	-0.080	0.012	-0.025	0.180	-0.009
	(0.148)	(0.078)	(0.079)	(0.062)	(0.077)	(0.166)	(0.059)
No FSM	0.227*	0.045	0.038	-0.018	0.120	0.182	0.050
	(0.136)	(0.094)	(0.083)	(0.077)	(0.104)	(0.128)	(0.080)
Parent(s) degree	1.137**	-0.350	-0.659***	-0.493*	-0.331	1.165**	0.755***
	(0.452)	(0.219)	(0.208)	(0.280)	(0.204)	(0.571)	(0.200)
Parent(s) A-levels	0.341	0.029	-0.022	0.053	0.121	0.357	0.009
	(0.220)	(0.142)	(0.142)	(0.173)	(0.160)	(0.362)	(0.182)
Parent(s) GCSE	0.118	-0.062	-0.025	-0.147	-0.055	-0.069	0.061
	(0.362)	(0.158)	(0.170)	(0.158)	(0.159)	(0.316)	(0.161)
Parent(s) work	-0.214	-0.107	-0.094	-0.035	-0.185	-0.307	-0.066
	(0.197)	(0.099)	(0.112)	(0.123)	(0.116)	(0.212)	(0.088)
Parent(s) in professional occupations	0.072	0.021	0.055	0.038	0.054	0.019	0.083
	(0.179)	(0.107)	(0.119)	(0.125)	(0.133)	(0.245)	(0.105)
Parent(s) SOC missing	-0.034	0.088	0.001	0.029	-0.011	-0.113	0.052
	(0.151)	(0.079)	(0.079)	(0.080)	(0.092)	(0.159)	(0.093)
Parent(s) aged <40	0.006	0.027	-0.052	0.015	-0.000	0.078	-0.011
	(0.134)	(0.063)	(0.068)	(0.070)	(0.065)	(0.175)	(0.071)
Parent(s) in very good health	-0.003	-0.006	0.025	0.010	0.019	0.043	0.049
	(0.152)	(0.074)	(0.101)	(0.081)	(0.085)	(0.186)	(0.090)
Household income over £21k	-0.010	0.068	0.019	0.107	-0.058	-0.071	0.095
	(0.330)	(0.113)	(0.126)	(0.147)	(0.116)	(0.287)	(0.168)
Household income missing	-0.320*	-0.065	-0.092	-0.132	-0.073	-0.377*	-0.059
	(0.168)	(0.081)	(0.081)	(0.094)	(0.072)	(0.217)	(0.086)
Parent would like YP to continue in FTE	0.400	0.216	0.119	0.370***	0.212*	0.570**	0.102
	(0.274)	(0.168)	(0.121)	(0.108)	(0.112)	(0.287)	(0.133)
Higher education very likely (Parent)	0.171	0.172**	0.213***	0.168***	0.194***	0.213	0.130*
	(0.137)	(0.079)	(0.073)	(0.063)	(0.067)	(0.137)	(0.073)
Very involved in school life (Parent)	-0.040	-0.055	-0.136**	-0.066	-0.076	-0.013	-0.004
	(0.141)	(0.057)	(0.067)	(0.065)	(0.074)	(0.145)	(0.074)
Not bullied	0.186	0.091	0.069	0.080	0.028	0.119	0.027
	(0.123)	(0.068)	(0.068)	(0.060)	(0.078)	(0.139)	(0.060)
Bullying missing	-0.051	-0.043	-0.052	0.023	0.008	-0.245	0.118
	(0.273)	(0.140)	(0.153)	(0.225)	(0.119)	(0.319)	(0.158)
Future-oriented	0.370***	0.089	0.175**	0.205***	0.109	0.382**	0.080
	(0.110)	(0.063)	(0.081)	(0.076)	(0.073)	(0.153)	(0.069)
Raising family important	0.019	0.041	-0.028	0.018	-0.050	-0.069	0.023
	(0.164)	(0.069)	(0.082)	(0.083)	(0.075)	(0.200)	(0.065)
Job helping other important	0.014	-0.006	0.003	0.007	0.023	0.134	0.029
	(0.122)	(0.065)	(0.071)	(0.064)	(0.066)	(0.139)	(0.070)
Well paying job important	-0.024	-0.021	-0.096	-0.034	-0.045	-0.088	0.057
	(0.138)	(0.071)	(0.064)	(0.072)	(0.070)	(0.159)	(0.070)
Being own boss important	-0.452***	-0.167***	-0.242***	-0.222***	-0.209***	-0.431***	-0.152**
	(0.113)	(0.057)	(0.052)	(0.054)	(0.061)	(0.141)	(0.074)
Interesting job important	-0.005	0.039	0.092*	0.100	0.038	-0.007	0.041
	(0.135)	(0.059)	(0.055)	(0.068)	(0.061)	(0.108)	(0.076)
Job with promotion important	0.053	0.069	0.029	0.031	0.046	0.106	0.006
	(0.136)	(0.070)	(0.079)	(0.087)	(0.078)	(0.172)	(0.078)
Job with regular hours important	0.221*	0.070	0.086	0.053	0.107	0.315*	0.106*
	(0.122)	(0.066)	(0.072)	(0.067)	(0.076)	(0.177)	(0.062)
Has ideas for career	0.175	0.107	0.080	0.056	0.119	0.248*	0.041
	(0.148)	(0.082)	(0.078)	(0.075)	(0.078)	(0.146)	(0.090)
Region of residence	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	477	477	477	477	477	477	477
R2	0.596	0.527	0.550	0.554	0.532	0.590	0.504

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. Model 2 includes the extended set of controls. Constant included in all regressions but is not reported to comply with the secure data service non-disclosure guidelines. Omitted groups: *Gender:* Male; *Ethnicity:* Mixed/Other; *EFL*: no EFL; *No SEN:* has SEN; *Two-parent family*: single parent family; *Homeowner*: not homeowner; *No FSM:* eligible for FSM;*Parent(s) education:* low/no qualifications; *Parent(s) work:* not working; *Parent(s) occupation:* not in managerial/professional occupations; *Parent(s) age:* $\geq$ 40; *Parent(s) health:* not very good; *Household income:* lower than £21k; *Parent(s) aspirations:* would not like child to continue in FTE ; *HE:* HE not very likely; *Parent(s) involvement:* not very involved in school life; *Bullying:* bullied in past 12 months; *Future-oriented:* not future oriented;*Raising family*: not very important ; *Job characteristics:* respective characteristic not important; *Career ambition:* no career ideas;*Region of residence:* London. Source: Next Steps.

Table C.17 Anticipated discrimination and GCSE achievement - complete Results Model 2 (OLS and LPM); Black Caribbean group only

	Educational Outcomes						
	Total GCSE score	**5+ A*-C**	**5+ A*-C Eng,Math**	**A*-C Eng**	**A*-C Math**	**Science score**	**Any A*-A**
Anticipate Disc	-0.003	-0.056	-0.038	0.138	-0.027	-0.078	0.010
	(0.186)	(0.101)	(0.089)	(0.115)	(0.094)	(0.170)	(0.096)
Female	0.360*	0.177*	0.018	0.080	-0.031	0.227	0.192*
	(0.209)	(0.096)	(0.099)	(0.111)	(0.108)	(0.202)	(0.099)
EFL	-0.832	0.090	-0.108	0.415	0.121	-0.400	0.140
	(0.672)	(0.348)	(0.388)	(0.394)	(0.395)	(0.616)	(0.361)
No SEN	0.620***	0.451***	0.298***	0.284**	0.339***	0.580***	0.170*
	(0.222)	(0.109)	(0.096)	(0.125)	(0.108)	(0.203)	(0.099)
Number of siblings	0.002	-0.007	-0.021	-0.048	0.020	-0.032	-0.028
	(0.078)	(0.038)	(0.041)	(0.047)	(0.045)	(0.084)	(0.033)
Two-parent family	0.190	0.078	-0.030	0.038	0.018	0.073	0.171
	(0.214)	(0.114)	(0.124)	(0.114)	(0.124)	(0.216)	(0.104)
Homeowner	0.111	-0.046	0.043	-0.027	-0.027	0.078	0.007
	(0.227)	(0.102)	(0.106)	(0.123)	(0.130)	(0.216)	(0.094)
No FSM	0.143	0.093	0.030	-0.086	0.045	0.367	-0.012
	(0.292)	(0.136)	(0.106)	(0.182)	(0.135)	(0.323)	(0.126)
Parent(s) degree	0.718**	0.364**	0.216	0.023	0.134	0.642*	0.203
	(0.301)	(0.167)	(0.181)	(0.199)	(0.225)	(0.336)	(0.184)
Parent(s) A-levels	0.013	-0.086	-0.023	-0.046	-0.025	0.013	-0.056
	(0.222)	(0.122)	(0.105)	(0.121)	(0.118)	(0.237)	(0.113)
Parent(s) GCSE	0.273	-0.014	0.037	0.062	-0.011	0.054	-0.009
	(0.270)	(0.134)	(0.125)	(0.134)	(0.134)	(0.255)	(0.112)
Parent(s) work	-0.168	0.073	0.190*	0.047	0.178	-0.142	0.078
	(0.208)	(0.113)	(0.105)	(0.168)	(0.128)	(0.255)	(0.112)
Parent(s) in professional occupations	-0.013	-0.218*	-0.076	-0.054	-0.008	-0.051	0.025
	(0.237)	(0.119)	(0.138)	(0.138)	(0.145)	(0.259)	(0.108)
Parent(s) SOC missing	-0.567	-0.123	0.042	-0.137	0.021	-0.512	0.099
	(0.534)	(0.273)	(0.177)	(0.267)	(0.165)	(0.513)	(0.134)
Parent(s) aged <40	0.034	0.020	0.073	-0.002	0.045	0.106	0.092
	(0.175)	(0.101)	(0.093)	(0.117)	(0.095)	(0.187)	(0.076)
Parent(s) in very good health	0.055	0.034	0.089	-0.009	0.064	0.187	-0.080
	(0.204)	(0.115)	(0.117)	(0.141)	(0.107)	(0.228)	(0.117)
Household income over £21k	-0.331	-0.066	0.004	0.066	-0.073	-0.290	0.004
	(0.283)	(0.152)	(0.148)	(0.131)	(0.144)	(0.260)	(0.102)
Household income missing	-0.156	-0.002	0.034	-0.068	-0.026	-0.024	-0.046
	(0.252)	(0.123)	(0.138)	(0.133)	(0.138)	(0.208)	(0.094)
Parent would like YP to continue in FTE	0.008	-0.054	-0.005	-0.122	0.008	-0.100	0.042
	(0.275)	(0.161)	(0.138)	(0.160)	(0.162)	(0.304)	(0.192)
Higher education very likely (Parent)	0.058	0.106	0.125	0.130	0.184	0.140	0.076
	(0.172)	(0.106)	(0.096)	(0.104)	(0.111)	(0.159)	(0.094)
Very involved in school life (Parent)	-0.214	-0.060	-0.087	0.030	-0.069	-0.253	0.004
	(0.184)	(0.105)	(0.120)	(0.109)	(0.114)	(0.208)	(0.090)
Not bullied=1	0.107	0.093	0.108	0.121	0.076	-0.067	0.058
	(0.198)	(0.105)	(0.109)	(0.119)	(0.111)	(0.195)	(0.093)
Bullying missing=1	-0.117	0.096	0.103	0.104	0.020	-0.152	0.088
	(0.751)	(0.242)	(0.309)	(0.281)	(0.350)	(0.542)	(0.259)
Future-oriented	0.269	0.050	0.069	0.033	0.132	0.197	0.053
	(0.241)	(0.122)	(0.139)	(0.117)	(0.144)	(0.237)	(0.141)
Raising family important	-0.254	0.047	-0.032	-0.033	-0.081	-0.121	-0.025
	(0.259)	(0.142)	(0.133)	(0.145)	(0.146)	(0.242)	(0.104)
Job helping other important	-0.130	-0.105	0.026	-0.082	-0.052	-0.220	-0.018
	(0.186)	(0.120)	(0.092)	(0.107)	(0.095)	(0.188)	(0.084)
Well paying job important	-0.014	-0.034	0.023	0.010	0.015	-0.064	0.036
	(0.250)	(0.151)	(0.124)	(0.152)	(0.136)	(0.208)	(0.097)
Being own boss important	-0.235	-0.233**	-0.222**	-0.072	-0.209**	-0.334*	-0.198*
	(0.184)	(0.104)	(0.093)	(0.112)	(0.104)	(0.185)	(0.105)
Interesting job important	0.103	0.094	0.106	0.038	0.101	0.098	0.106
	(0.204)	(0.101)	(0.109)	(0.123)	(0.123)	(0.194)	(0.093)
Job with promotion important	-0.058	-0.016	0.006	0.022	0.024	0.017	0.010
	(0.183)	(0.137)	(0.104)	(0.140)	(0.118)	(0.188)	(0.084)
Job with regular hours important	0.205	-0.028	0.027	0.079	0.007	0.348**	-0.066
	(0.177)	(0.099)	(0.094)	(0.104)	(0.109)	(0.168)	(0.077)
Has ideas for career	0.273	0.201	-0.000	0.058	0.092	0.047	-0.076
	(0.318)	(0.167)	(0.126)	(0.175)	(0.129)	(0.279)	(0.128)
Region of residence	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	366	366	366	366	366	366	366
R2	0.737	0.709	0.692	0.676	0.692	0.744	0.726

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. Model 2 includes the extended set of controls. Constant included in all regressions but is not reported to comply with the secure data service non-disclosure guidelines. Omitted groups: *Gender:* Male; *Ethnicity:* Mixed/Other; *EFL*: no EFL; *No SEN:* has SEN; *Two-parent family*: single parent family; *Homeowner*: not homeowner; *No FSM:* eligible for FSM;*Parent(s) education:* low/no qualifications; *Parent(s) work:* not working; *Parent(s) occupation:* not in managerial/professional occupations; *Parent(s) age:* ≥ 40; *Parent(s) health:* not very good; *Household income:* lower than £21k; *Parent(s) aspirations:* would not like child to continue in FTE ; *HE:* HE not very likely; *Parent(s) involvement:* not very involved in school life; *Bullying:* bullied in past 12 months; *Future-oriented:* not future oriented;*Raising family*: not very important ; *Job characteristics:* respective characteristic not important; *Career ambition:* no career ideas;*Region of residence:* London. Source: Next Steps.

Table C.18 Anticipated discrimination and GCSE achievement - complete Results Model 2 (OLS and LPM); Black African group only

	Educational Outcomes						
	Total GCSE score	**5+ A*-C**	**5+ A*-C Eng,Math**	**A*-C Eng**	**A*-C Math**	**Science score**	**Any A*-A**
Anticipate Disc	-0.026	0.057	-0.001	0.014	0.032	0.069	0.030
	(0.201)	(0.106)	(0.125)	(0.103)	(0.114)	(0.226)	(0.117)
Female	0.111	0.090	-0.122	-0.043	-0.062	-0.001	0.063
	(0.273)	(0.129)	(0.162)	(0.132)	(0.171)	(0.246)	(0.151)
EFL	-0.197	0.005	-0.157	0.027	-0.060	-0.087	-0.207
	(0.261)	(0.148)	(0.176)	(0.150)	(0.177)	(0.341)	(0.166)
No SEN	0.099	0.031	0.134	0.140	0.233	0.262	0.064
	(0.335)	(0.183)	(0.245)	(0.171)	(0.261)	(0.326)	(0.192)
Number of siblings	-0.081	-0.033	-0.035	-0.017	-0.023	-0.059	-0.030
	(0.052)	(0.031)	(0.029)	(0.032)	(0.036)	(0.057)	(0.029)
Two-parent family	0.403*	0.105	0.098	0.099	0.154	0.292	-0.024
	(0.225)	(0.122)	(0.129)	(0.099)	(0.149)	(0.249)	(0.137)
Homeowner	-0.070	0.130	0.081	0.106	0.160	-0.037	0.141
	(0.247)	(0.138)	(0.143)	(0.126)	(0.162)	(0.280)	(0.205)
No FSM	-0.171	-0.001	0.164	0.030	0.004	-0.163	0.022
	(0.283)	(0.140)	(0.170)	(0.137)	(0.191)	(0.263)	(0.155)
Parent(s) degree	-0.168	-0.150	-0.034	-0.144	-0.029	-0.190	-0.171
	(0.395)	(0.188)	(0.228)	(0.188)	(0.215)	(0.452)	(0.302)
Parent(s) A-levels	-0.087	-0.107	-0.089	-0.149	0.022	-0.101	-0.127
	(0.273)	(0.138)	(0.174)	(0.139)	(0.180)	(0.326)	(0.164)
Parent(s) GCSE	0.162	0.089	-0.045	-0.062	0.024	-0.029	0.189
	(0.283)	(0.121)	(0.169)	(0.164)	(0.167)	(0.359)	(0.196)
Parent(s) work	0.119	0.050	0.083	0.128	0.091	0.221	-0.028
	(0.296)	(0.165)	(0.166)	(0.142)	(0.183)	(0.343)	(0.200)
Parent(s) in professional occupations	0.129	0.082	0.176	0.088	0.129	0.125	0.118
	(0.210)	(0.135)	(0.125)	(0.098)	(0.143)	(0.271)	(0.152)
Parent(s) SOC missing	-0.199	-0.120	-0.048	-0.187	-0.088	-0.159	-0.024
	(0.347)	(0.172)	(0.168)	(0.137)	(0.217)	(0.365)	(0.210)
Parent(s) aged <40	-0.113	-0.005	-0.034	-0.074	0.032	-0.157	-0.023
	(0.189)	(0.104)	(0.124)	(0.109)	(0.121)	(0.211)	(0.117)
Parent(s) in very good health	-0.066	0.056	0.091	0.013	0.091	-0.090	0.019
	(0.226)	(0.142)	(0.116)	(0.115)	(0.134)	(0.272)	(0.133)
Household income over £21k	0.083	0.022	0.050	0.008	0.040	-0.029	0.039
	(0.239)	(0.148)	(0.155)	(0.127)	(0.151)	(0.279)	(0.146)
Household income missing	0.187	0.030	-0.052	0.034	-0.102	-0.018	0.040
	(0.262)	(0.121)	(0.143)	(0.099)	(0.162)	(0.291)	(0.139)
Parent would like YP to continue in FTE	2.437***	0.775**	-0.304	0.239	-0.169	2.556**	-0.340
	(0.738)	(0.362)	(0.484)	(0.429)	(0.587)	(1.060)	(0.593)
Higher education very likely (Parent)	0.305	-0.039	0.156	0.192*	0.024	0.463	0.173
	(0.219)	(0.116)	(0.128)	(0.114)	(0.143)	(0.286)	(0.153)
Very involved in school life (Parent)	-0.242	-0.043	-0.063	-0.041	0.009	-0.208	-0.036
	(0.175)	(0.091)	(0.110)	(0.097)	(0.109)	(0.207)	(0.121)
Not bullied	-0.108	0.065	0.070	0.102	-0.011	0.035	-0.185
	(0.214)	(0.123)	(0.121)	(0.100)	(0.110)	(0.265)	(0.138)
Bullying missing	0.041	0.136	-0.381	-0.260	-0.316	0.135	-0.459
	(0.468)	(0.200)	(0.383)	(0.358)	(0.479)	(0.845)	(0.462)
Future-oriented	0.374*	0.278**	0.294*	0.253*	0.182	0.363	0.094
	(0.203)	(0.140)	(0.151)	(0.147)	(0.152)	(0.263)	(0.171)
Raising family important	0.207	0.131	-0.041	0.158	0.085	0.323	-0.069
	(0.319)	(0.195)	(0.191)	(0.173)	(0.222)	(0.339)	(0.191)
Job helping other important	-0.117	-0.097	-0.006	0.039	-0.056	-0.071	-0.102
	(0.185)	(0.117)	(0.124)	(0.082)	(0.138)	(0.196)	(0.119)
Well paying job important	0.040	0.117	0.030	-0.009	0.083	0.060	-0.023
	(0.245)	(0.126)	(0.149)	(0.101)	(0.174)	(0.327)	(0.167)
Being own boss important	-0.073	-0.067	-0.178	-0.207*	-0.122	-0.277	-0.137
	(0.219)	(0.120)	(0.134)	(0.117)	(0.129)	(0.210)	(0.127)
Interesting job important	-0.023	0.047	0.040	-0.037	0.039	0.070	0.111
	(0.205)	(0.111)	(0.126)	(0.111)	(0.122)	(0.244)	(0.124)
Job with promotion important	0.080	0.058	0.047	0.028	0.084	0.006	-0.017
	(0.260)	(0.126)	(0.160)	(0.139)	(0.165)	(0.321)	(0.159)
Job with regular hours important	-0.026	0.031	0.007	0.071	0.018	0.055	0.080
	(0.176)	(0.103)	(0.116)	(0.111)	(0.112)	(0.233)	(0.117)
Has ideas for career	0.163	0.002	-0.161	-0.091	-0.041	-0.090	0.049
	(0.207)	(0.127)	(0.138)	(0.128)	(0.149)	(0.245)	(0.124)
Region of residence	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	289	289	289	289	289	289	289
R2	0.773	0.728	0.693	0.762	0.658	0.674	0.637

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. Model 2 includes the extended set of controls. Constant included in all regressions but is not reported to comply with the secure data service non-disclosure guidelines. Omitted groups: *Gender:* Male; *Ethnicity:* Mixed/Other; *EFL*: no EFL; *No SEN:* has SEN; *Two-parent family*: single parent family; *Homeowner*: not homeowner; *No FSM:* eligible for FSM;*Parent(s) education:* low/no qualifications; *Parent(s) work:* not working; *Parent(s) occupation:* not in managerial/professional occupations; *Parent(s) age:* ≥ 40; *Parent(s) health:* not very good; *Household income:* lower than £21k; *Parent(s) aspirations:* would not like child to continue in FTE ; *HE:* HE not very likely; *Parent(s) involvement:* not very involved in school life; *Bullying:* bullied in past 12 months; *Future-oriented:* not future oriented;*Raising family*: not very important ; *Job characteristics:* respective characteristic not important; *Career ambition:* no career ideas;*Region of residence:* London. Source: Next Steps.

Table C.19 Anticipated discrimination and GCSE achievement - complete Results Model 2 (OLS and LPM); Mixed/Other group only

	Educational Outcomes						
	Total GCSE score	5+ A*-C	5+ A*-C Eng,Math	A*-C Eng	A*-C Math	Science score	Any A*-A
Anticipate Disc	0.042	0.067	0.104	0.025	0.073	0.041	0.048
	(0.121)	(0.069)	(0.071)	(0.062)	(0.070)	(0.127)	(0.062)
Female	0.210	0.068	0.061	0.117	-0.000	0.104	0.105
	(0.143)	(0.076)	(0.079)	(0.073)	(0.075)	(0.139)	(0.074)
EFL	-0.180	-0.103	-0.109	-0.055	-0.215	-0.263	-0.217**
	(0.257)	(0.124)	(0.137)	(0.108)	(0.133)	(0.267)	(0.110)
No SEN	0.462**	0.239**	0.160*	0.232**	0.174*	0.415*	0.074
	(0.193)	(0.097)	(0.090)	(0.101)	(0.094)	(0.212)	(0.085)
Number of siblings	-0.115**	-0.033	-0.029	-0.027	-0.024	-0.082	-0.001
	(0.054)	(0.025)	(0.024)	(0.028)	(0.024)	(0.057)	(0.025)
Two-parent family	0.249	0.163*	0.141	0.092	0.099	0.305*	0.052
	(0.177)	(0.086)	(0.094)	(0.097)	(0.096)	(0.182)	(0.084)
Homeowner	0.398**	0.154*	0.133	0.117	0.127	0.240	0.098
	(0.158)	(0.083)	(0.093)	(0.081)	(0.089)	(0.154)	(0.077)
No FSM	-0.148	-0.119	-0.107	-0.038	-0.109	-0.200	-0.095
	(0.225)	(0.115)	(0.114)	(0.113)	(0.111)	(0.199)	(0.099)
Parent(s) degree	0.039	0.061	-0.018	0.064	0.075	0.171	-0.155
	(0.186)	(0.117)	(0.116)	(0.118)	(0.121)	(0.197)	(0.113)
Parent(s) A-levels	0.142	0.112	0.106	0.118	0.051	0.115	0.052
	(0.169)	(0.087)	(0.092)	(0.083)	(0.089)	(0.169)	(0.092)
Parent(s) GCSE	-0.052	0.033	0.028	0.056	-0.004	-0.039	-0.040
	(0.144)	(0.082)	(0.072)	(0.070)	(0.077)	(0.154)	(0.075)
Parent(s) work	0.043	0.043	-0.029	0.020	-0.022	0.169	-0.028
	(0.182)	(0.097)	(0.099)	(0.093)	(0.099)	(0.190)	(0.091)
Parent(s) in professional occupations	0.146	0.023	0.007	-0.010	-0.015	0.126	0.120
	(0.150)	(0.075)	(0.075)	(0.084)	(0.075)	(0.154)	(0.080)
Parent(s) SOC missing	0.275	0.049	0.112	0.006	-0.007	0.230	0.075
	(0.268)	(0.134)	(0.146)	(0.134)	(0.146)	(0.271)	(0.130)
Parent(s) aged <40	-0.084	-0.052	-0.024	0.001	-0.008	-0.047	-0.045
	(0.121)	(0.064)	(0.068)	(0.065)	(0.065)	(0.127)	(0.061)
Parent(s) in very good health	0.179	0.048	0.089	-0.012	0.084	0.136	0.055
	(0.125)	(0.074)	(0.070)	(0.059)	(0.071)	(0.131)	(0.063)
Household income over £21k	0.083	0.037	0.079	0.033	0.085	0.121	0.135
	(0.145)	(0.080)	(0.084)	(0.084)	(0.084)	(0.170)	(0.089)
Household income missing	0.006	-0.040	0.032	-0.004	0.084	0.032	0.114
	(0.163)	(0.088)	(0.089)	(0.085)	(0.088)	(0.160)	(0.089)
Parent would like YP to continue in FTE	0.243	0.227*	0.197*	0.232*	0.099	0.033	0.028
	(0.240)	(0.134)	(0.117)	(0.123)	(0.114)	(0.240)	(0.105)
Higher education very likely (Parent)	0.255	0.097	0.188***	0.166**	0.158**	0.297*	0.179**
	(0.159)	(0.076)	(0.066)	(0.074)	(0.069)	(0.158)	(0.070)
Very involved in school life (Parent)	0.207*	0.079	0.080	0.088	0.038	0.089	0.043
	(0.119)	(0.061)	(0.069)	(0.067)	(0.068)	(0.134)	(0.068)
Not bullied=1	0.154	0.019	-0.010	-0.001	0.030	0.166	0.050
	(0.123)	(0.067)	(0.068)	(0.067)	(0.068)	(0.122)	(0.067)
Bullying missing	0.187	0.105	0.033	-0.047	-0.070	0.269	0.344
	(0.739)	(0.224)	(0.303)	(0.311)	(0.360)	(0.728)	(0.298)
Future-oriented	0.234	0.060	0.037	0.054	-0.001	0.155	0.106
	(0.146)	(0.073)	(0.073)	(0.072)	(0.074)	(0.143)	(0.067)
Raising family important	0.221	0.063	0.037	0.049	0.011	0.142	-0.045
	(0.257)	(0.110)	(0.113)	(0.099)	(0.114)	(0.242)	(0.103)
Job helping other important	-0.211*	-0.046	-0.123*	-0.049	-0.107	-0.128	-0.127**
	(0.124)	(0.069)	(0.066)	(0.061)	(0.069)	(0.134)	(0.062)
Well paying job important	-0.096	-0.006	-0.062	-0.027	-0.029	-0.077	-0.088
	(0.141)	(0.072)	(0.078)	(0.069)	(0.076)	(0.148)	(0.070)
Being own boss important	-0.168	-0.119	-0.126*	-0.096	-0.133*	-0.147	-0.092
	(0.135)	(0.075)	(0.071)	(0.068)	(0.073)	(0.154)	(0.073)
Interesting job important	-0.008	-0.009	-0.061	-0.023	-0.037	-0.049	-0.032
	(0.127)	(0.060)	(0.068)	(0.062)	(0.065)	(0.128)	(0.068)
Job with promotion important	-0.063	-0.011	0.033	0.046	-0.052	0.019	-0.046
	(0.128)	(0.075)	(0.074)	(0.069)	(0.075)	(0.139)	(0.065)
Job with regular hours important	-0.144	-0.061	-0.043	-0.037	-0.018	-0.149	-0.007
	(0.129)	(0.062)	(0.068)	(0.063)	(0.068)	(0.142)	(0.064)
Has ideas for career	-0.051	-0.007	-0.010	-0.039	0.018	0.011	-0.020
	(0.160)	(0.078)	(0.085)	(0.078)	(0.089)	(0.168)	(0.079)
Region of residence	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	795	795	795	795	795	795	795
R2	0.740	0.667	0.680	0.677	0.671	0.714	0.701

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. Model 2 includes the extended set of controls. Constant included in all regressions but is not reported to comply with the secure data service non-disclosure guidelines. Omitted groups: *Gender:* Male; *Ethnicity:* Mixed/Other; *EFL*: no EFL; *No SEN:* has SEN; *Two-parent family*: single parent family; *Homeowner*: not homeowner; *No FSM:* eligible for FSM;*Parent(s) education:* low/no qualifications; *Parent(s) work:* not working; *Parent(s) occupation:* not in managerial/professional occupations; *Parent(s) age:* ≥ 40; *Parent(s) health:* not very good; *Household income:* lower than £21k; *Parent(s) aspirations:* would not like child to continue in FTE ; *HE:* HE not very likely; *Parent(s) involvement:* not very involved in school life; *Bullying:* bullied in past 12 months; *Future-oriented:* not future oriented;*Raising family*: not very important ; *Job characteristics:* respective characteristic not important; *Career ambition:* no career ideas;*Region of residence:* London. Source: Next Steps.

Table C.20 Differences along the distribution of KS2 scores - summarised results Model 2 (OLS and LPM)

	Educational Outcomes						
Quartiles	Total GCSE score	5+ A*-C	5+ A*-C Eng,Math	A*-C Eng	A*-C Math	Science score	Any A*-A
Q1 (Bottom)	0.058	0.020	0.026	-0.004	0.050	0.068	0.005
	(0.092)	(0.048)	(0.028)	(0.041)	(0.041)	(0.101)	(0.034)
N	829	829	829	829	829	829	829
Q2	0.109	0.113**	0.069	0.023	0.052	0.105	0.038
	(0.078)	(0.054)	(0.056)	(0.052)	(0.057)	(0.087)	(0.050)
N	831	831	831	831	831	831	831
Q3	0.015	0.045	0.014	0.021	0.025	0.046	-0.059
	(0.077)	(0.044)	(0.049)	(0.044)	(0.046)	(0.087)	(0.057)
N	834	834	834	834	834	834	834
Q4 (Top)	-0.006	-0.012	0.002	-0.014	-0.014	0.044	0.010
	(0.065)	(0.021)	(0.029)	(0.021)	(0.025)	(0.065)	(0.044)
N	821	821	821	821	821	821	821

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. Source: Next Steps.

Table C.21 Separating 'Yes' and 'Don't Know'(DK) categories - summarised results (OLS and LPM)

	Educational Outcomes						
	Total GCSE score	5+ A*-C	5+ A*-C Eng, Math	A*-C Eng	A*-C Math	Science score	Any A*-A
Panel A: Model 1							
Yes	0.104**	0.072***	0.077***	0.091***	0.065**	0.108**	0.072***
	(0.050)	(0.026)	(0.027)	(0.025)	(0.025)	(0.054)	(0.026)
DK	0.061	0.035	0.053**	0.018	0.064***	0.110**	0.004
	(0.044)	(0.023)	(0.023)	(0.022)	(0.021)	(0.046)	(0.024)
Yes=DK (p-value)	0.448	0.212	0.446	0.013**	0.970	0.973	0.029**
Panel B: Model 2							
Yes	0.108**	0.075***	0.078***	0.095***	0.066***	0.113**	0.070***
	(0.048)	(0.025)	(0.026)	(0.024)	(0.024)	(0.052)	(0.025)
DK	0.068*	0.039*	0.054**	0.020	0.066***	0.118***	0.003
	(0.041)	(0.022)	(0.022)	(0.022)	(0.020)	(0.043)	(0.024)
Yes=DK (p-value)	0.454	0.205	0.439	0.008***	0.997	0.923	0.023**
N	3315	3315	3315	3315	3315	3315	3315

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. Model 1 includes the standard set of controls. Model 2 includes the standard and extended sets of controls. Regressions run for the pooled sample of ethnic minority pupils (N=3315). Source: Next Steps.

Table C.22 Separating 'Yes' and 'Don't Know' categories - summarised results Model 2 (OLS and LPM); by ethnicity

	Ethnic Group					
	Indian	Pakistani	Bangladeshi	Black Caribbean	Black African	Mixed/Other
Total GCSE score						
Yes	0.057	0.294*	0.309*	-0.045	-0.042	0.272
	(0.119)	(0.177)	(0.181)	(0.233)	(0.205)	(0.191)
DK	0.070	0.222**	0.180	0.045	0.011	-0.090
	(0.133)	(0.095)	(0.151)	(0.224)	(0.303)	(0.136)
Yes=DK (p-value)	0.933	0.681	0.519	0.739	0.848	0.093*
5+ A-C*						
Yes	0.098*	0.149*	0.121	-0.088	0.019	0.184
	(0.058)	(0.080)	(0.083)	(0.131)	(0.118)	(0.117)
DK	0.086	0.120**	0.036	-0.020	0.147	-0.001
	(0.068)	(0.056)	(0.076)	(0.119)	(0.137)	(0.071)
Yes=DK (p-value)	0.888	0.737	0.373	0.652	0.356	0.118
5+ A-C incl. Eng, Mat*						
Yes	0.118*	0.150**	0.239***	-0.090	-0.053	0.202*
	(0.071)	(0.074)	(0.084)	(0.113)	(0.128)	(0.108)
DK	0.056	0.153***	0.096	0.022	0.123	0.047
	(0.070)	(0.058)	(0.076)	(0.113)	(0.183)	(0.077)
Yes=DK (p-value)	0.537	0.972	0.157	0.434	0.289	0.176
A-C Eng*						
Yes	0.145**	0.126*	0.239***	0.163	0.026	0.116
	(0.057)	(0.074)	(0.083)	(0.147)	(0.111)	(0.103)
DK	0.096	0.095*	-0.015	0.110	-0.014	-0.028
	(0.070)	(0.054)	(0.071)	(0.127)	(0.144)	(0.068)
Yes=DK (p-value)	0.564	0.701	0.009***	0.726	0.774	0.196
A-C Math*						
Yes	0.067	0.109	0.177*	-0.068	-0.041	0.156
	(0.066)	(0.080)	(0.093)	(0.111)	(0.116)	(0.107)
DK	0.070	0.167***	0.077	0.020	0.206	0.026
	(0.060)	(0.056)	(0.071)	(0.121)	(0.167)	(0.079)
Yes=DK (p-value)	0.973	0.503	0.352	0.513	0.090*	0.263
Science score						
Yes	0.098	0.400**	0.412*	-0.095	0.028	0.161
	(0.127)	(0.161)	(0.222)	(0.223)	(0.246)	(0.213)
DK	0.105	0.239**	0.305*	-0.058	0.165	-0.028
	(0.122)	(0.116)	(0.161)	(0.196)	(0.373)	(0.140)
Yes=DK (p-value)	0.968	0.382	0.650	0.883	0.725	0.417
Any A-A*						
Yes	0.065	0.102	0.192**	-0.077	0.013	0.216**
	(0.075)	(0.077)	(0.085)	(0.108)	(0.126)	(0.093)
DK	-0.026	0.083	0.060	0.110	0.069	-0.048
	(0.074)	(0.051)	(0.094)	(0.114)	(0.157)	(0.067)
Yes=DK (p-value)	0.309	0.827	0.278	0.113	0.706	0.006***
N	712	676	477	366	289	795

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. Model 2 includes the standard and extended sets of controls. Source: Next Steps.

Table C.23 Differences by gender - summarised results (OLS and LPM)

	Educational Outcomes						
	Total GCSE score	5+ A*-C	5+ A*-C Eng,Math	A*-C Eng	A*-C Math	Science score	Any A*-A
Panel A: Model 1							
Anticipate Disc	0.047	0.040	0.044	0.056**	0.049*	0.055	0.024
	(0.052)	(0.028)	(0.028)	(0.027)	(0.025)	(0.056)	(0.028)
Female	0.255***	0.129***	0.057**	0.173***	-0.012	0.070	0.134***
	(0.048)	(0.028)	(0.029)	(0.027)	(0.027)	(0.051)	(0.027)
Anticipate Disc*Female	0.062	0.020	0.036	-0.016	0.029	0.105	0.015
	(0.069)	(0.038)	(0.039)	(0.037)	(0.036)	(0.075)	(0.037)
Panel B: Model 2							
Anticipate Disc	0.070	0.052*	0.054**	0.067***	0.060**	0.080	0.027
	(0.049)	(0.026)	(0.026)	(0.025)	(0.024)	(0.052)	(0.027)
Female	0.199***	0.107***	0.027	0.144***	-0.042*	0.009	0.106***
	(0.045)	(0.027)	(0.028)	(0.027)	(0.025)	(0.049)	(0.027)
Anticipate Disc*Female	0.027	0.004	0.019	-0.032	0.011	0.070	0.005
	(0.064)	(0.036)	(0.036)	(0.035)	(0.033)	(0.070)	(0.035)
N	3315	3315	3315	3315	3315	3315	3315

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. Model 1 includes the standard set of controls. Model 2 includes the extended set of controls. Regressions run for the pooled sample of ethnic minority pupils (N=3315). Source: Next Steps.

Table C.24 Differences by interview month (post July/05) - summarised results (OLS and LPM)

	Educational Outcomes						
	Total GCSE scores	5+ A*-C	5+ A*-C Eng,Math	A*-C Eng	A*-C Math	Science score	Any A*-A
Panel A: Model 1							
Anticipate Disc	0.079*	0.061**	0.076***	0.054**	0.074***	0.104**	0.021
	(0.046)	(0.025)	(0.025)	(0.023)	(0.025)	(0.049)	(0.024)
Post July	-0.033	-0.025	-0.037	-0.011	-0.013	-0.049	-0.017
	(0.051)	(0.028)	(0.028)	(0.029)	(0.028)	(0.051)	(0.028)
Anticipate Disc*Post July	0.001	-0.021	-0.026	-0.013	-0.021	0.014	0.024
	(0.071)	(0.037)	(0.038)	(0.037)	(0.038)	(0.075)	(0.037)
Panel B: Model 2							
Anticipate Disc	0.092**	0.068***	0.080***	0.059***	0.080***	0.118**	0.021
	(0.044)	(0.023)	(0.024)	(0.022)	(0.024)	(0.047)	(0.023)
Post July	-0.031	-0.023	-0.036	-0.011	-0.010	-0.047	-0.019
	(0.048)	(0.026)	(0.027)	(0.028)	(0.027)	(0.048)	(0.027)
Anticipate Disc*Post July	-0.014	-0.028	-0.033	-0.019	-0.030	-0.001	0.021
	(0.067)	(0.035)	(0.036)	(0.035)	(0.036)	(0.069)	(0.035)
N	3315	3315	3315	3315	3315	3315	3315

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. Model 1 includes the standard set of controls. Model 2 includes the extended set of controls. Regressions run for the pooled sample of ethnic minority pupils (N=3315). Source: Next Steps.

Table C.25 Anticipated discrimination and GCSE achievement - summarised weighted results (OLS and LPM)

	Educational Outcomes						
	Total GCSE score	5+ A*-C	5+ A*-C Eng,Math	A*-C Eng	A*-C Math	Science Score	Any A*-A
Model 1: Baseline Model							
Anticipate Disc	0.055	0.043*	0.051**	0.040*	0.053***	0.089*	0.017
	(0.042)	(0.023)	(0.022)	(0.021)	(0.020)	(0.047)	(0.021)
Standard controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Model 2: Extended Model							
Anticipate Disc	0.067*	0.049**	0.057***	0.046**	0.060***	0.101**	0.019
	(0.039)	(0.021)	(0.020)	(0.020)	(0.019)	(0.044)	(0.020)
Standard controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Extended controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Unweighted count	3315	3315	3315	3315	3315	3315	3315

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Sample weights applied to the regressions. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. Regressions run for the pooled sample of ethnic minority pupils (N=3315). Source: Next Steps

Table C.26 Anticipated discrimination and GCSE achievement - summarised weighted results value-added models VA-M1 and VA-M2 (OLS and LPM)

	Educational Outcomes						
	Total GCSE score	5+ A*-C	5+ A*-C Eng,Math	A*-C Eng	A*-C Math	Science Score	Any A*-A
Panel A: Model VA-M1							
Anticipate Disc	-0.011	0.014	0.021	0.011	0.022	0.021	-0.007
	(0.033)	(0.019)	(0.019)	(0.019)	(0.018)	(0.039)	(0.020)
KS2 scores	0.586***	0.254***	0.269***	0.253***	0.272***	0.598***	0.211***
	(0.023)	(0.010)	(0.011)	(0.010)	(0.013)	(0.025)	(0.012)
Panel B: Model VA-M2							
Anticipate Disc	-0.000	0.021	0.027	0.017	0.028	0.032	-0.004
	(0.033)	(0.019)	(0.019)	(0.018)	(0.018)	(0.038)	(0.020)
KS2 scores	0.540***	0.229***	0.244***	0.229***	0.252***	0.554***	0.189***
	(0.023)	(0.010)	(0.011)	(0.010)	(0.013)	(0.027)	(0.012)
Unweighted count	3315	3315	3315	3315	3315	3315	3315

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Sample weights applied to the regressions. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. VA-M1 includes the standard set of controls plus prior achievement in KS2. VA-M2 includes extended set of controls plus prior achievement in KS2. Source: Next Steps.

Table C.27 Anticipated discrimination and GCSE achievement - summarised weighted results Model 2 (OLS and LPM) by ethnicity

	Ethnic Group					
	Indian	Pakistani	Bangladeshi	Black Caribbean	Black African	Mixed/Other
Total GCSE score						
Anticipate Disc	0.117	0.235**	0.153	-0.030	-0.027	0.065
	(0.108)	(0.102)	(0.135)	(0.180)	(0.199)	(0.135)
5+ A-C*						
Anticipate Disc	0.112**	0.122**	0.058	-0.082	0.084	0.096
	(0.051)	(0.052)	(0.065)	(0.103)	(0.104)	(0.069)
5+ A-C incl. Eng, Mat*						
Anticipate Disc	0.098*	0.148***	0.137**	-0.068	0.015	0.125*
	(0.054)	(0.049)	(0.058)	(0.086)	(0.122)	(0.069)
A-C Eng*						
Anticipate Disc	0.132**	0.105**	0.054	0.153	0.023	0.018
	(0.052)	(0.047)	(0.059)	(0.119)	(0.107)	(0.058)
A-C Math*						
Anticipate Disc	0.069	0.133**	0.110**	-0.050	0.047	0.102*
	(0.050)	(0.051)	(0.055)	(0.090)	(0.114)	(0.069)
Science score						
Anticipate Disc	0.147*	0.319***	0.285**	-0.082	0.081	0.074
	(0.101)	(0.117)	(0.142)	(0.173)	(0.226)	(0.152)
Any A-A*						
Anticipate Disc	0.017	0.076*	0.106	-0.006	0.017	0.044
	(0.061)	(0.043)	(0.065)	(0.091)	(0.117)	(0.055)
Unweighted count	712	676	477	366	289	795

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Sample weights applied to the regressions. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. Model 2 includes the extended set of controls. Source: Next Steps.

Table C.28 Anticipated discrimination and GCSE achievement - summarised weighted results value-added model VA-M2 (OLS and LPM), by ethnicity

	Ethnic Group					
	Indian	Pakistani	Bangladeshi	Black Caribbean	Black African	Mixed/Other
Total GCSE score						
Anticipate Disc	0.060	0.088	-0.013	-0.002	-0.073	-0.025
	(0.086)	(0.086)	(0.118)	(0.173)	(0.158)	(0.122)
KS2 scores	0.564***	0.592***	0.472***	0.424***	0.382***	0.626***
	(0.046)	(0.048)	(0.063)	(0.119)	(0.072)	(0.084)
5+ A-C*						
Anticipate Disc	0.090**	0.054	-0.017	-0.067	0.059	0.065
	(0.043)	(0.048)	(0.056)	(0.099)	(0.082)	(0.066)
KS2 scores	0.222***	0.276***	0.212***	0.226***	0.206***	0.221***
	(0.028)	(0.025)	(0.023)	(0.051)	(0.048)	(0.035)
5+ A-C incl. Eng, Mat*						
Anticipate Disc	0.072	0.089*	0.048	-0.051	-0.011	0.087*
	(0.044)	(0.046)	(0.055)	(0.081)	(0.104)	(0.066)
KS2 scores	0.259***	0.235***	0.253***	0.249***	0.210***	0.260***
	(0.030)	(0.024)	(0.027)	(0.054)	(0.054)	(0.037)
A-C Eng*						
Anticipate Disc	0.109**	0.037	-0.035	0.168	0.000	-0.015
	(0.046)	(0.042)	(0.054)	(0.112)	(0.089)	(0.061)
KS2 scores	0.228***	0.274***	0.251***	0.233***	0.188***	0.231***
	(0.027)	(0.026)	(0.027)	(0.072)	(0.055)	(0.035)
A-C Math*						
Anticipate Disc	0.046	0.067	0.019	-0.033	0.013	0.061
	(0.043)	(0.049)	(0.049)	(0.082)	(0.084)	(0.064)
KS2 scores	0.232***	0.268***	0.257***	0.254***	0.276***	0.285***
	(0.027)	(0.027)	(0.021)	(0.051)	(0.044)	(0.044)
Science score						
Anticipate Disc	0.091	0.175	0.113	-0.049	0.034	-0.024
	(0.084)	(0.107)	(0.120)	(0.167)	(0.198)	(0.128)
KS2 scores	0.553***	0.583***	0.488***	0.504***	0.378***	0.681***
	(0.037)	(0.073)	(0.066)	(0.111)	(0.085)	(0.090)
Any A-A*						
Anticipate Disc	-0.006	0.029	0.045	0.003	-0.009	0.020
	(0.058)	(0.038)	(0.056)	(0.089)	(0.094)	(0.056)
KS2 scores	0.224***	0.189***	0.172***	0.131**	0.214***	0.169***
	(0.022)	(0.023)	(0.041)	(0.055)	(0.049)	(0.037)
Unweighted count	712	676	477	366	289	795

Notes: Robust standard errors clustered at the secondary school level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Sample weights applied to the regressions. Total GCSE score and Science score are standardised points scores with mean zero and standard deviation one. All outcomes and main controls are described in the text in section 4.4.1. VA-M2 includes the extended set of controls plus prior achievement in KS2. Source: Next Steps.

Chapter 5

Conclusions

Motivated by the economic and political importance of social mobility in the context of increasing income inequality, this thesis has presented three studies on the topic of social mobility in the UK. Despite substantial previous research efforts, many aspects of how inequalities are perpetuated and socio-economic status is transmitted across generations in the UK remain elusive. To fill knowledge gaps in this literature, I have addressed key research questions that were previously unanswered or relatively unexplored. This thesis contributes to the existing literature from multiple angles. The first empirical chapter examines the degree of intergenerational income mobility at the national level for the UK. I use contemporaneous data from the harmonised BHPS and Understanding Society to examine the degree of intergenerational income mobility in the UK in the 21st century. The following empirical chapter expands on the analysis of income mobility at the national level to include an examination of income mobility along three key dimensions: (i) differences between sons and daughters, (ii) at different points of the income distribution, and (iii) across regions in the UK. The final empirical chapter shifts the focus from family background to individual factors that may influence socio-economic status in adulthood, through their influence on decisions on investment in human capital. In this chapter, I examine the role of discrimination for human capital formation. Specifically, I analyse the influence of individual expectations of facing future discrimination on educational attainment for ethnic minority pupils in England. Employing data from another panel survey, Next Steps, which allows me to directly observe expectations of facing future labour market discrimination, this chapter provides valuable insights into how discrimination can influence individual decisions prior to entry to the labour market.

5.1 Summary of Results

In Chapter 2, I tackled the question surrounding the transmission of socio-economic status across generations in the UK. In this chapter, I measure the degree of intergenerational income mobility at the national level using a new dataset comprised of the harmonised BHPS and Understanding Society data. This long panel provides information on income observed for two generations over 27 years of data, for both individuals and parents from multiple birth cohorts. For the first time, income mobility for the cohort of individuals born between 1973-1992, who are now aged 28-47, is analysed. This comprehensive dataset allows income mobility to be examined at the household level, focusing on the pooled resources of all members of the household. Moreover, this data allows me to overcome some limitations of previous work in this area and obtain income mobility estimates that are robust to several methodological choices and reduce measurement errors present in previous studies. I apply a novel method to this literature using a two-stage residual approach to obtain an improved measure of long-run income for both generations. This allows me to consider the role of life cycle effects and transitory shocks to income and generate a comparable measure of income for both generations. When considering household well-being, I show that there is a strong income persistence across generations in the UK. These results corroborate the previous estimates that indicate relatively low levels of social mobility in the UK, and substantiate the notion that the intergenerational transmission of income resources across generations is a major determinant of socio-economic status in adulthood. I find a similar degree of persistence when examining mobility with respect to alternative measures of income for the second generation. Finally, I show that these results are robust to changes in the model specifications, in the restrictions applied to the sample and to the treatment of outliers.

After examining income mobility at the national level, Chapter 3 expands this analysis to investigate heterogeneities in intergenerational income mobility across subgroups of British society. I start by examining differences in the degree of income mobility between sons and daughters. First, I find that there is no significant variation in intergenerational income mobility by gender when considering the transmission of resources at the household level. When examining income mobility at the individual level, however, some gender differences arise, which might be attributed to labour market participation effects. I then examine an important implicated mechanism of intergenerational transmission and find evidence of assortative mating on income. There are strong links between partner's income and in-laws income, especially for daughters, and this potentially reinforces the dynamics of persistence of inequalities. Further, I investigate differences in income mobility across points of the distribution of parents' and children's incomes. Transition probabilities reveal substantial differences across the parental income distribution, with children whose parents are located at the tails of the income distribution displaying lower levels of income mobility. An analysis of directional rank mobility reveals high small-scale rank mobility across the parental distribution. Examining upwards and downwards rank mobility, I also observe that the patterns of directional movements are likely to vary by gender. Meanwhile, when looking at differences across the children's income distribution, no significant variation was observed in the estimates obtained using quantile regressions. Delving deeper into the nuances of persistent inequalities in the UK, I examine the geography of income mobility across the country by region of upbringing. An interesting finding is revealed; that regions in the North of England display substantially lower levels of both relative and absolute income mobility than in the South. Examining these regions at a finer scale of disaggregation, I also observe additional regional variations that suggest a more nuanced picture of unequal opportunities for social mobility than purely a North/South divide. In general, regions with high relative mobility also display more opportunities for absolute upwards mobility.

The final empirical chapter, Chapter 4, examines the influence of expectations of future discrimination on the educational achievement of ethnic minority pupils in England. In particular, I examine whether anticipating labour market discrimination is associated with educational performance in GCSE exams. Here, I use data from Next Steps, which contains a wealth of information on pupils' characteristics and background, as well as unique data on pupils' expectations of facing future discrimination in the labour market. I find that higher attainment in GCSE exams is positively and significantly correlated with anticipated discrimination. This indicates that labour market discrimination might influence the amount of effort that pupils exert in the acquisition of cognitive skills during compulsory schooling, i.e. even prior to entry to the labour market. To avoid the generalisation of ethnic minorities as a whole, I split the sample by ethnic groups to unpick how this role of anticipated discrimination varies across ethnic groups. Here, I distinguish how pupils of different ethnicity might differ in their response to anticipated discrimination. My main finding, of a positive association between anticipated discrimination and higher attainment, is primarily driven by pupils of certain ethnic groups, notably of South Asian origin. For some ethnic groups, associations even exhibited the reverse pattern.

5.2 Policy Implications, Limitations and Future Research

Taken together, the results from this thesis shed new light on inequalities in British society. The results from Chapter 2 indicate that the UK exhibits considerably limited potential for social mobility in comparison to other developed countries. Given that family background plays an important role in determining the economic status of children in adulthood, the strong persistence of socio-economic status across generations contributes to the persistence of deep-rooted inequalities in British society. Children's opportunities are defined by a combination factors and circumstances related to family conditions, labour markets, local environment and current policy, and they determine the extent to which socio-economic status in adulthood is related to family background. In the context of high levels of income inequality, and without drastic changes in the availability of opportunities, my findings suggest that the levels of intergenerational mobility for the next generation are likely to be similar, if not worse. This is particularly worrying given the current Covid-19 pandemic which is exacerbating various forms of inequality and contributing to an even more polarised labour market (Blundell et al., 2020). Whilst immediate economic demands are currently at the forefront of the political discourse, it is important that policy-makers keep the social mobility agenda in mind. The extent to which socio-economic advantages and disadvantages are transmitted across generations will play a critical role the long-term economic development, and now the challenges to improve social mobility in the UK will be even greater. Future policy should focus on increasing the push towards more equality of opportunities to prevent a further widening of inequalities.

The findings from Chapter 3 offer important insights in this direction, by demonstrating the importance of looking at the statistics on social mobility beyond the national level. Several group specificities need to be considered when addressing social mobility, with special consideration of the most vulnerable groups. Previous work that examines mobility at the national level has often been lacking in the ability to inform targeted policy priorities due to the more aggregated nature of the research. Many of the observed heterogeneities in my research identify groups that experience less social mobility and this should be taken into consideration when designing policy that aims to equalise opportunities. Notably, my research points to the strong persistence for children from disadvantaged families, highlighting the need to support these children. Given the crucial role of education as a channel for social mobility, it is of key importance to foster further discussion into methods of reducing the inequalities in the UK education system. Further research and resources should be allocated towards reducing both within-school and between-school variation and reforming admission policies. In addition, my research shows that when considering the opportunities for social mobility of women, it is crucial to consider the implications of their patterns of labour market participation. Another important aspect of my research is that beyond the persistence of income inequalities across generations, there are important geographical differences, which further compound societal inequalities. I provide evidence that a UK postcode lottery does exist, and that children who grow up in certain regions have fewer opportunities available. Observing within-country regional variations can provide valuable insights into conditions that favour social mobility, and the causal mechanisms driving intergenerational mobility specific to the UK context. In this setting, policies focused at the local and regional level can prove to be a valuable strategy in tackling inequalities. Although the Social Mobility

Commission has repeatedly recommended various actions to mitigate current inequalities, not much progress has been made in the past decade (Social Mobility Commission, 2020), as such the social mobility agenda is often overlooked by the current government. My findings call for the importance of the intergenerational and regional dimensions of social mobility to be considered more seriously and as part of a coordinated strategy in order to identify the most effective policy actions to improve social mobility for the future.

The results from Chapter 4 indicate that expectations of receiving unequal treatment and fewer opportunities in the labour market can affect individual outcomes prior to entry in the market. By disentangling the role of expectations of discrimination from that of other factors that influence educational attainment, I show that these expectations influence the incentives to make observable human capital investments for ethnic minorities from some, but not all, groups. Interestingly, my findings suggest that ethnic minorities who anticipate discrimination have, on average, higher educational achievement at the end of secondary school. This suggests that individual beliefs and perceptions about discrimination can have a notable influence on actual outcomes related to human capital investment. Moreover, it suggests this process takes place at an earlier stage than perhaps would be expected, with evidence for this in teenage years. In addition, my research reveals that treating all ethnic minorities as a homogeneous group, as is frequently the case in other studies, can conceal important distinctions between these groups. Notably, pupils of South Asian origin who anticipate discrimination attain better outcomes. Although the exact reasons for these group differences are unknown, one possible explanation is that they could relate to the distinct migration histories and cultures of different ethnic groups. As education is a key channel of upward mobility, this under explored variation by ethnicity has further implications for the development and persistence of inequalities. It is important that specific characteristics of these historically disadvantaged groups are taken into account when developing policies that aim to equalise opportunities. This particularly calls for the need to critically reconsider the standard ‘colour-blind’ approach of simply comparing ‘whites’ and ‘non-whites’. The recent Black Lives Matter protests around the world have brought the issue of racial discrimination to the forefront of the political debate and are raising awareness of the various facets of discrimination and of the particular experiences of specific groups. To make these analyses possible in practice, this research highlights the need for surveys to oversample ethnic minorities to allow the disaggregation of the data. One conclusion from this study is that if discrimination represents a constraint for ethnic minorities, then this challenges the notion that a more equal society can be achieved through promoting equal opportunities via education. Therefore, it is essential to continue the promotion of anti-discrimination laws and practices and for schools, universities and employers to provide equal opportunities.

The analysis of intergenerational mobility in the UK, the topic of Chapters 2 and 3 requires further development which will only be possible with the availability of new, even more comprehensive datasets. Research in this field would greatly benefit from the use of administrative records on income and earnings for the whole British population. Research for other countries using this type of data has proven critical in generating robust measures of intergenerational mobility and gaining understanding nuances in the geography of social mobility, and most importantly, to improve our understanding of the causal mechanisms underlying the intergenerational transmission of resources. The regional analysis in Chapter 3 provides valuable insights into regional variations within the UK, but due to the limited sample size for some regions, there is uncertainty around the statistical precision of some of the estimates and strict conclusions cannot be inferred. It is of particular importance to obtain data at a more disaggregated regional level, for example for local authority districts, in order to gauge the levels of mobility and link this study to that of the local characteristics that could potentially foster social mobility. Another important avenue for future research is the examination of ethnic differences in the levels of intergenerational mobility. Although studies for the US have found evidence of ethnic differences, to my knowledge, this has not been examined for the UK. This issue is of particular relevance in the current times, given the disproportionate impact of the Covid-19 crisis on the incomes of individuals from certain ethnic groups (Blundell et al., 2020). In addition to administrative tax records, future research will also benefit from the release of the next waves of Understanding Society. This evolving dataset will enable the verification of some of my hypotheses and findings and to expand the study of mobility to include younger cohorts and more observations of income for individuals as adults. This is of critical importance to maintain an up-to-date understanding of developments in social mobility in order to guide appropriate policy actions. Moreover, such analyses will be important to examine the evolution of trends of income mobility over time. The econometric techniques used in this thesis, particularly the TSRA method proposed in Chapter 2, could be a useful tool in measuring intergenerational mobility.

The analysis of Chapter 4 could be further expanded by examining later educational and labour market outcomes. For example, future work could consider the impact of anticipating discrimination on A-level results and participation in higher education, or how expectations of future discrimination in the labour market match with actual labour market experiences. In addition, future research could examine more systematically the differences between ethnic groups, including differences between ethnic minority and majority groups. Future research should also investigate the process of formation of discriminatory expectations, particularly when they are formed and how they relate to one's own experiences of discrimination and that of similar individuals in the labour market. Observing anticipated discrimination at various periods for the same individuals over time could improve our understanding about the evolution of such expectations during childhood and adolescence. This also calls for a revision of data collection processes to incorporate more questions that may reveal individual expectations.

In summary, each chapter of this thesis advances the general topic of social mobility and contributes to various strands of this literature such as intergenerational income mobility, equality of opportunities, education and discrimination. As we enter a period of crisis and extreme uncertainty under the current socio-political climate, the findings of this thesis are particularly relevant for considering the implications of current policies (or lack thereof) for future inequalities. Throughout the thesis I explore different mechanisms that contribute to the perpetuation of inequalities in British society. Further, I demonstrate implications for the literature on regional development and on ethnic educational inequalities. In its entirety, this thesis builds upon and develops the body of academic research on economic inequality. Several future avenues of research have been highlighted to further develop our understanding of these topics. Given the motivations and outcomes of each chapter, the implications of this research should provide a valuable resource for policy makers and social scientists alike.

References

Acciari, P., Polo, A. and Violante, G. L. (2019), *'And Yet, it Moves': Intergenerational Mobility in Italy*, IZA Discussion Paper 12273, Institute for the Study of Labor (IZA).

Almlund, M., Duckworth, A. L., Heckman, J. J., Kautz, T., Hanushek, E. A., Machin, S. and Woessmann, L. (2011), Personality, psychology, and economics, *in* A. Hanushek, S. Machin and L. Woessmann, eds, 'Handbook of the Economics of Education', Vol. 4, Amsterdam: North Holland, chapter 1, pp. 1–181.

Altonji, J. G. and Dunn, T. A. (2000), 'An intergenerational model of wages, hours, and earnings', *Journal of Human Resources* **35**(2), 221–258.

Andrews, D. and Leigh, A. (2009), 'More inequality, less social mobility', *Applied Economics Letters* **16**(15), 1489–1492.

Angrist, J. D. and Pischke, J.-S. (2008), *Mostly harmless econometrics: An empiricist's companion*, Princeton: Princeton University Press.

Arcidiacono, P., Bayer, P. and Hizmo, A. (2010), 'Beyond signaling and human capital: Education and the revelation of ability', *American Economic Journal: Applied Economics* **2**(4), 76–104.

Arrow, K. (1973), 'The theory of discrimination', *Discrimination in Labor Markets* **3**(10), 3–33.

Atkinson, A. B. (1980), 'On intergenerational income mobility in Britain', *Journal of Post Keynesian Economics* **3**(2), 194–218.

Atkinson, A. B. (1983), 'The measurement of economic mobility', *Social Justice and Public Policy* **3**, 61–75.

Becker, B. (2011), 'Social disparities in children's vocabulary in early childhood. does pre-school education help to close the gap?', *The British Journal of Sociology* **62**(1), 69–88.

Becker, G. S. (1957), *The Economics of Discrimination*, Chicago: University of Chicago Press.

Becker, G. S. (2009), *Human capital: A theoretical and empirical analysis, with special reference to education*, Chicago: University of Chicago Press.

Becker, G. S. and Tomes, N. (1979), 'An equilibrium theory of the distribution of income and intergenerational mobility', *Journal of Political Economy* **87**(6), 1153–1189.

Becker, G. S. and Tomes, N. (1986), 'Human capital and the rise and fall of families', *Journal of Labor Economics* **4**(3, Part 2), 1–39.

Belfield, C., Blundell, R., Cribb, J., Hood, A. and Joyce, R. (2017), 'Two decades of income inequality in Britain: the role of wages, household earnings and redistribution', *Economica* **84**(334), 157–179.

Belfield, C., Crawford, C., Greaves, E., Gregg, P. and Macmillan, L. (2017), *Intergenerational income persistence within families*, IFS Working Paper W17/11, IFS.

Bell, B., Blundell, J. and Machin, S. (2018), *The changing geography of intergenerational mobility*, CEP Discussion Papers CEPDP1591, London: Centre for Economic Performance, LSE.

Berthoud, R. (2000), 'Ethnic employment penalties in Britain', *Journal of Ethnic and Migration Studies* **26**(3), 389–416.

Bhattacharya, D. and Mazumder, B. (2011), 'A nonparametric analysis of black–white differences in intergenerational income mobility in the United States', *Quantitative Economics* **2**(3), 335–379.

Björklund, A. (1993), 'A comparison between actual distributions of annual and lifetime income: Sweden 1951–89', *Review of Income and Wealth* **39**(4), 377–386.

Björklund, A. and Jäntti, M. (1997), 'Intergenerational income mobility in Sweden compared to the United States', *American Economic Review* **87**(5), 1009–1018.

Björklund, A. and Jäntti, M. (2011), Intergenerational income mobility and the role of family background, *in* W. Salverda, B. Nolan and T. Smeeding, eds, 'The Oxford Handbook of Economic Inequality', Oxford University Press, pp. 491–521.

Björklund, A. and Salvanes, K. G. (2010), Education and family background: Mechanisms and policies, *in* 'Handbook of the Economics of Education', Vol. 3, Amsterdam: North Holland, chapter 3, pp. 201–247.

Black, S. E. and Devereux, P. J. (2011), Recent developments in intergenerational mobility, *in* O. Ashenfelter and D. Card, eds, 'Handbook of Labor Economics', Vol. 4, Amsterdam: North Holland, pp. 1487–1541.

Blackaby, D. H., Leslie, D. G., Murphy, P. D. and O'Leary, N. C. (2002), 'White/ethnic minority earnings and employment differentials in Britain: evidence from the LFS', *Oxford Economic Papers* **54**(2), 270–297.

Blackaby, D., Leslie, D., Murphy, P. and O'Leary, N. (1999), 'Unemployment among Britain's ethnic minorities', *The Manchester School* **67**(1), 1–20.

Blanden, J. (2005*a*), 'Intergenerational mobility and assortative mating in the UK', *London: Centre for Economic Performance, LSE. Unpublished manuscript* .

Blanden, J. (2005*b*), 'Love and money: Intergenerational mobility and marital matching', *Analytical Studies Branch Research Paper Series, Family and Labour Studies Division, Statistics Canada* .

Blanden, J. (2009), *How much can we learn from international comparisons of intergenerational mobility?*, CEE Discussion Papers 0111, London: Centre for the Economics of Education, LSE.

Blanden, J., Goodman, A., Gregg, P. and Machin, S. (2004), Changes in intergenerational mobility in Britain, *in* M. Corak, ed., 'Generational income mobility in North America and Europe', Cambridge: Cambridge University Press, chapter 6, pp. 122–46.

Blanden, J., Gregg, P. and Macmillan, L. (2013), 'Intergenerational persistence in income and social class: the effect of within-group inequality', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **176**(2), 541–563.

Blau, P. M. and Duncan, O. D. (1967), *The American occupational structure*, New York: John Wiley and Sons.

Blundell, R., Dias, M. C., Joyce, R. and Xu, X. (2020), 'Covid-19 and inequalities', *Fiscal Studies (forthcoming)* .

Blundell, R. and Etheridge, B. (2010), 'Consumption, income and earnings inequality in britain', *Review of Economic Dynamics* **13**(1), 76–102.

Böhlmark, A. and Lindquist, M. J. (2006), 'Life-cycle variations in the association between current and lifetime income: Replication and extension for Sweden', *Journal of Labor Economics* **24**(4), 879–896.

Borah, B. J. and Basu, A. (2013), 'Highlighting differences between conditional and unconditional quantile regression approaches through an application to assess medication adherence', *Health Economics* **22**(9), 1052–1070.

Bratberg, E., Anti Nilsen, Ø. and Vaage, K. (2005), 'Intergenerational earnings mobility in Norway: Levels and trends', *Scandinavian Journal of Economics* **107**(3), 419–435.

Bratberg, E., Davis, J., Mazumder, B., Nybom, M., Schnitzlein, D. D. and Vaage, K. (2017), 'A comparison of intergenerational mobility curves in Germany, Norway, Sweden, and the US', *The Scandinavian Journal of Economics* **119**(1), 72–101.

Bratsberg, B., Røed, K., Raaum, O., Naylor, R., Jäntti, M., Eriksson, T. and Osterbacka, E. (2007), 'Nonlinearities in Intergenerational Earnings Mobility: Consequences for Cross-Country Comparisons', *The Economic Journal* **117**(519), 72–92.

Breen, R. and Jonsson, J. O. (2005), 'Inequality of opportunity in comparative perspective: Recent research on educational attainment and social mobility', *Annual Review of Sociology* **31**, 223–243.

Brewer, M. and Wren-Lewis, L. (2016), 'Accounting for changes in income inequality: decomposition analyses for the UK, 1978–2008', *Oxford Bulletin of Economics and Statistics* **78**(3), 289–322.

Brynin, M. and Güveli, A. (2012), 'Understanding the ethnic pay gap in Britain', *Work, Employment and Society* **26**(4), 574–587.

Burgess, S. M. and Heller-Sahlgren, G. (2018), 'Motivated to succeed? attitudes to education among native and immigrant pupils in England', *IZA Discussion Paper No. 11678* .

Cabinet Office (2017), *Race Disparity Audit – Summary Findings from the Ethnicity Facts and Figures Website*, https://www.ethnicity-facts-figures.service.gov.uk/static/race-disparity-audit-summary-findings.pdf. Accessed: 25/02/2020.

Carmichael, F. and Woods, R. (2000), 'Ethnic penalties in unemployment and occupational attainment: evidence for Britain', *International Review of Applied Economics* **14**(1), 71–98.

Carneiro, P., Heckman, J. J. and Masterov, D. V. (2005), 'Labor market discrimination and racial differences in premarket factors', *The Journal of Law and Economics* **48**(1), 1–39.

Chadwick, L. and Solon, G. (2002), 'Intergenerational income mobility among daughters', *American Economic Review* **92**(1), 335–344.

Chetty, R., Hendren, N., Kline, P. and Saez, E. (2014), 'Where is the land of opportunity? the geography of intergenerational mobility in the United States', *Quarterly Journal of Economics* **129**(4), 1553–1623.

Chowdry, H., Crawford, C. and Goodman, A. (2009), *Drivers and barriers to educational success-evidence from the Longitudinal Study of Young People in England*, Research Report DCSF-RR102, Department for Children, Schools and Families. Institute for Fiscal Studies.

Chowdry, H., Crawford, C. and Goodman, A. (2011), 'The role of attitudes and behaviours in explaining socio-economic differences in attainment at age 16', *Longitudinal and Life Course Studies* **2**(1), 59–76.

Coate, S. and Loury, G. (1993*a*), 'Antidiscrimination enforcement and the problem of patronization', *The American Economic Review* **83**(2), 92–98.

Coate, S. and Loury, G. C. (1993*b*), 'Will affirmative-action policies eliminate negative stereotypes?', *The American Economic Review* **83**(5), 1220–1240.

Connolly, M., Corak, M. and Haeck, C. (2019), 'Intergenerational mobility between and within Canada and the United States', *Journal of Labor Economics* **37**(S2), S595–S641.

Corak, M. (2004), Generational income mobility in North America and Europe: an introduction, *in* M. Corak, ed., 'Generational Income Mobility in North America and Europe', Cambridge: Cambridge University Press, chapter 1, pp. 1–37.

Corak, M. (2006), 'Do poor children become poor adults? Lessons from a cross-country comparison of generational earnings mobility', *Research on Economic Inequality* **13**(1), 143–188.

Corak, M. (2013), 'Income Inequality, Equality of Opportunity, and Intergenerational Mobility', *Journal of Economic Perspectives* **27**(3), 79–102.

Corak, M. and Heisz, A. (1999), 'The intergenerational earnings and income mobility of Canadian men: Evidence from longitudinal income tax data', *Journal of Human Resources* **34**, 504–533.

Corak, M., Lindquist, M. J. and Mazumder, B. (2014), 'A comparison of upward and downward intergenerational mobility in Canada, Sweden and the United States', *Labour Economics* **30**, 185–200.

Couch, K. A. and Lillard, D. R. (1998), 'Sample selection rules and the intergenerational correlation of earnings', *Labour Economics* **5**(3), 313–329.

Couch, K. A. and Lillard, D. R. (2004), Non-linear patterns of intergenerational mobility in Germany and the United States, *in* M. Corak, ed., 'Generational Income Mobility in North America and Europe', Cambridge: Cambridge University Press, chapter 8, pp. 190–206.

Crawford, C., Macmillan, L. and Vignoles, A. (2017), 'When and why do initially high-achieving poor children fall behind?', *Oxford Review of Education* **43**(1), 88–108.

Cunha, F., Heckman, J. J., Lochner, L. and Masterov, D. V. (2006), Interpreting the evidence on life cycle skill formation, *in* E. A. Hanushek and F. Welch, eds, 'Handbook of the Economics of Education', Vol. 1, Amsterdam: North Holland, chapter 12, pp. 697–812.

d'Addio, A. C. (2007), *Intergenerational transmission of disadvantage: mobility or immobility across generations? A review of the evidence for OECD countries*, OECD Social, Employment, and Migration Working Papers 52, Organisation for Economic Cooperation and Development (OECD).

Dahl, M. W. and DeLeire, T. (2008), *The association between children's earnings and fathers' lifetime earnings: estimates using administrative data*, Discussion Paper 1342-08, University of Wisconsin-Madison, Institute for Research on Poverty.

Dale, A., Shaheen, N., Kalra, V. and Fieldhouse, E. (2002), 'Routes into education and employment for young Pakistani and Bangladeshi women in the UK', *Ethnic and Racial Studies* **25**(6), 942–968.

Darity, W. A. and Mason, P. L. (1998), 'Evidence on discrimination in employment: Codes of color, codes of gender', *Journal of Economic Perspectives* **12**(2), 63–90.

Davidson, R. and MacKinnon, J. G. (2000), 'Bootstrap tests: How many bootstraps?', *Econometric Reviews* **19**(1), 55–68.

De Luca, G., Magnus, J. R. and Peracchi, F. (2018), 'Balanced variable addition in linear models', *Journal of Economic Surveys* **32**(4), 1183–1200.

Dearden, L., Machin, S. and Reed, H. (1997), 'Intergenerational mobility in Britain', *The Economic Journal* **107**(440), 47–66.

Demie, F. (2001), 'Ethnic and gender differences in educational achievement and implications for school improvement strategies', *Educational Research* **43**(1), 91–106.

Demie, F. and Strand, S. (2006), 'English language acquisition and educational attainment at the end of secondary school', *Educational Studies* **32**(2), 215–231.

Department for Education (2011), *LSYPE User Guide to the Datasets: Wave 1 to Wave 7*, Technical report, London: Department for Education.

Derenoncourt, E. (2019), 'Can you move to opportunity? evidence from the Great Migration', *Unpublished Manuscript* .

Deutscher, N. and Mazumder, B. (2019), 'Intergenerational mobility in australia: national and regional estimates using administrative data', *Life Course Centre Working Paper Series No.2019-02* .

DfES (2005), Ethnicity and education: The evidence on minority ethnic pupils aged 5-16, DfES Topic Paper RTP01-05, DfES.

Dustmann, C., Fabbri, F., Preston, I. and Wadsworth, J. (2003), Labour market performance of immigrants in the UK labour market, Home Office Online Report 05/03, Research Development and Statistics Directorate, Home Office.

Dustmann, C., Machin, S. and Schönberg, U. (2010), 'Ethnicity and educational achievement in compulsory schooling', *The Economic Journal* **120**(546), 272–297.

Dustmann, C. and Theodoropoulos, N. (2010), 'Ethnic minority immigrants and their children in Britain', *Oxford Economic Papers* **62**(2), 209–233.

Eide, E. R. and Showalter, M. H. (1999), 'Factors affecting the transmission of earnings across generations: A quantile regression approach', *Journal of Human Resources* **34**(2), 253–267.

Elliott, R. J. and Lindley, J. K. (2008), 'Immigrant wage differentials, ethnicity and occupational segregation', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **171**(3), 645–671.

Eriksen, J. and Munk, M. D. (2020), 'The geography of intergenerational mobility — Danish evidence', *Economics Letters* **189**(109024).

Eriksen, T. L. M., Nielsen, H. S. and Simonsen, M. (2014), 'Bullying in elementary school', *Journal of Human Resources* **49**(4), 839–871.

Erikson, R. and Goldthorpe, J. H. (1992), *The constant flux: A study of class mobility in industrial societies*, Oxford: Oxford University Press.

Erikson, R. and Goldthorpe, J. H. (2002), 'Intergenerational inequality: A sociological perspective', *Journal of Economic Perspectives* **16**(3), 31–44.

Erikson, R. and Goldthorpe, J. H. (2010), 'Has social mobility in Britain decreased? reconciling divergent findings on income and class mobility', *The British Journal of Sociology* **61**(2), 211–230.

Ermisch, J. and Francesconi, M. (2004), Intergenerational mobility in Britain: new evidence from the British Household Panel Survey, *in* M. Corak, ed., 'Generational Income Mobility in North America and Europe', Cambridge: Cambridge University Press, chapter 7, pp. 147–189.

Ermisch, J., Francesconi, M. and Siedler, T. (2006), 'Intergenerational mobility and marital sorting', *The Economic Journal* **116**(513), 659–679.

Ermisch, J., Jäntti, M. and Smeeding, T. (2012), Inequality from childhood to adulthood: A cross-national perspective on the transmission of advantage, *in* J. Ermisch, M. Jäntti and T. Smeeding, eds, 'From Parents to Children: The Intergenerational Transmission of Advantage', New York: Russell Sage Foundation.

Fairbrother, D. and Mahadevan, R. (2016), 'Do education and sex matter for intergenerational earnings mobility? some evidence from australia', *Australian Economic Papers* **55**(3), 212–226.

Fang, H. and Moro, A. (2011), Theories of statistical discrimination and affirmative action: A survey, *in* J. Benhabib, M. O. Jackson and A. Bisin, eds, 'Handbook of Social Economics', Vol. 1A, Amsterdam: North Holland, chapter 5, pp. 133–200.

Farber, H. S. and Gibbons, R. (1996), 'Learning and wage dynamics', *The Quarterly Journal of Economics* **111**(4), 1007–1047.

Farrington, D. P. (1993), 'Understanding and preventing bullying', *Crime and Justice* **17**, 381–458.

Fernández-Reino, M. (2016), 'Immigrant optimism or anticipated discrimination? explaining the first educational transition of ethnic minorities in England', *Research in Social Stratification and Mobility* **46**, 141–156.

Firpo, S., Fortin, N. M. and Lemieux, T. (2009), 'Unconditional quantile regressions', *Econometrica* **77**(3), 953–973.

Friedman, S. and Macmillan, L. (2017), 'Is London really the engine-room? migration, opportunity hoarding and regional social mobility in the UK', *National Institute Economic Review* **240**(1), R58–R72.

Goldthorpe, J. H. (2013), 'Understanding–and misunderstanding–social mobility in Britain: The entry of the economists, the confusion of politicians and the limits of educational policy', *Journal of Social Policy* **42**(3), 431–450.

Goodman, A., Gregg, P. and Washbrook, E. (2011), 'Children's educational attainment and the aspirations, attitudes and behaviours of parents and children through childhood', *Longitudinal and Life Course Studies* **2**(1), 1–18.

Gorman, E., Harmon, C., Mendolia, S., Staneva, A. and Walker, I. (2019), *The Causal Effects of Adolescent School Bullying Victimisation on Later Life Outcomes*, IZA Discussion Paper 12241, IZA.

Grawe, N. D. (2004), Intergenerational mobility for whom? the experience of high-and low-earning sons in international perspective, *in* M. Corak, ed., 'Generational Income Mobility in North America and Europe', Cambridge: Cambridge University Press, chapter 4, pp. 58–89.

Grawe, N. D. (2006), 'Lifecycle bias in estimates of intergenerational earnings persistence', *Labour Economics* **13**(5), 551–570.

Greene, B. A. and DeBacker, T. K. (2004), 'Gender and orientations toward the future: Links to motivation', *Educational Psychology Review* **16**(2), 91–120.

Gregg, P., Macmillan, L. and Vittori, C. (2017), 'Moving towards estimating sons' lifetime intergenerational economic mobility in the UK', *Oxford Bulletin of Economics and Statistics* **79**(1), 79–100.

Gregg, P., Macmillan, L. and Vittori, C. (2019), 'Intergenerational income mobility: access to top jobs, the low-pay no-pay cycle and the role of education in a common framework', *Journal of Population Economics* **32**(2), 501–528.

Gregg, P. and Washbrook, E. (2011), 'The role of attitudes and behaviours in explaining socio-economic differences in attainment at age 11', *Longitudinal and Life Course Studies* **2**(1), 41–58.

Güell, M., Pellizzari, M., Pica, G. and Rodríguez Mora, J. V. (2018), 'Correlating social mobility and economic outcomes', *The Economic Journal* **128**(612), F353–F403.

Güell, M., Rodríguez Mora, J. V. and Telmer, C. I. (2015), 'The informational content of surnames, the evolution of intergenerational mobility, and assortative mating', *The Review of Economic Studies* **82**(2), 693–735.

Guryan, J. and Charles, K. K. (2013), 'Taste-based or statistical discrimination: the economics of discrimination returns to its roots', *The Economic Journal* **123**(572), F417–F432.

Haider, S. and Solon, G. (2006), 'Life-cycle variation in the association between current and lifetime earnings', *American Economic Review* **96**(4), 1308–1320.

Haldane, A. (2020), 'Speech for the Industrial Strategy Council'.
URL: *https://industrialstrategycouncil.org/comprehensive-set-local-industrial-strategies-consistently-applied-key-levelling-uk*

Hanushek, E. (2010), Education production functions: Evidence from developed countries, *in* D. J. Brewer and P. J. McEwan, eds, 'Economics of Education', Oxford: Academic Press, pp. 132–135.

Hanushek, E. A. and Woessmann, L. (2010), The economics of international differences in educational achievement, *in* A. Hanushek, S. Machin and L. Woessmann, eds, 'Handbook of the Economics of Education', Vol. 3, Amsterdam: North Holland, chapter 2, pp. 89–200.

Heath, A. and Cheung, S. Y. (2006), *Ethnic penalties in the labour market: Employers and discrimination*, Research Report 341, Department for Work and Pensions.

Heath, A. F., Rothon, C. and Kilpi, E. (2008), 'The second generation in Western Europe: Education, unemployment, and occupational attainment', *Annual Review of Sociology* **34**, 211–235.

Heckman, J. J. and Mosso, S. (2014), 'The economics of human development and social mobility', *Annual Review of Economics* **6**(1), 689–733.

Heckman, J. J., Stixrud, J. and Urzua, S. (2006), 'The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior', *Journal of Labor Economics* **24**(3), 411–482.

Heidrich, S. (2017), 'Intergenerational mobility in Sweden: a regional perspective', *Journal of Population Economics* **30**(4), 1241–1280.

Hirvonen, L. H. (2008), 'Intergenerational earnings mobility among daughters and sons: Evidence from Sweden and a comparison with the United States', *American Journal of Economics and Sociology* **67**(5), 777–826.

Hole, A. R. and Ratcliffe, A. (2019), 'The impact of the London bombings on the well-being of adolescent Muslims', *The Scandinavian Journal of Economics* **000**(0), 1–34.

Holmlund, H. (2020), 'How much does marital sorting contribute to intergenerational socioeconomic persistence?', *Journal of Human Resources* **55**, 0519–10227R1.

Jäntti, M., Bratsberg, B., Røed, K., Raaum, O., Naylor, R., Österbacka, E., Björklund, A. and Eriksson, T. (2006), *American Exceptionalism in a New Light: A Comparison of Intergenerational Earnings Mobility in the Nordic Countries, the United Kingdom and the United States*, IZA Discussion Papers 1938, Institute for the Study of Labor (IZA).

Jäntti, M. and Jenkins, S. P. (2015), Income mobility, *in* A. Atkinson and F. Bourguignon, eds, 'Handbook of Income Distribution', Vol. 2A, Elsevier, chapter 10, pp. 807–935.

Jenkins, S. (1987), 'Snapshots versus movies:'lifecycle biases' and the estimation of intergenerational earnings inheritance', *European Economic Review* **31**(5), 1149–1158.

Jerrim, J. and Macmillan, L. (2015), 'Income inequality, intergenerational mobility, and the Great Gatsby Curve: Is education the key?', *Social Forces* **94**(2), 505–533.

Keeley, B. (2015), *Income inequality: The gap between rich and poor*, Paris: OECD Publishing.

Khattab, N. (2015), 'Students' aspirations, expectations and school achievement: What really matters?', *British Educational Research Journal* **41**(5), 731–748.

Khattab, N. (2018), 'Ethnicity and higher education: The role of aspirations, expectations and beliefs in overcoming disadvantage', *Ethnicities* **18**(4), 457–470.

Kingdon, G. and Cassen, R. (2010), 'Ethnicity and low achievement in English schools', *British Educational Research Journal* **36**(3), 403–431.

Koenker, R. and Bassett, G. (1978), 'Regression quantiles', *Econometrica* **46**(1), 33–50.

Koenker, R. and Hallock, K. F. (2001), 'Quantile regression', *Journal of Economic Perspectives* **15**(4), 143–156.

Landersø, R. and Heckman, J. J. (2017), 'The Scandinavian fantasy: The sources of intergenerational mobility in Denmark and the US', *The Scandinavian Journal of Economics* **119**(1), 178–230.

Lang, K. and Lehmann, J.-Y. K. (2012), 'Racial discrimination in the labor market: Theory and empirics', *Journal of Economic Literature* **50**(4), 959–1006.

Lang, K. and Manove, M. (2011), 'Education and labor market discrimination', *American Economic Review* **101**(4), 1467–96.

Lee, C.-I. and Solon, G. (2009), 'Trends in intergenerational income mobility', *The Review of Economics and Statistics* **91**(4), 766–772.

Lefranc, A., Ojima, F. and Yoshida, T. (2014), 'Intergenerational earnings mobility in Japan among sons and daughters: levels and trends', *Journal of Population Economics* **27**(1), 91–134.

Leslie, D. and Drinkwater, S. (1999), 'Staying on in full-time education: reasons for higher participation rates among ethnic minority males and females', *Economica* **66**(261), 63–77.

Lessard-Phillips, L., Boliver, V., Pampaka, M. and Swain, D. (2018), 'Exploring ethnic differences in the post-university destinations of Russell Group graduates', *Ethnicities* **18**(4), 496–517.

Li, Y. (2018), 'Against the odds?–a study of educational attainment and labour market position of the second-generation ethnic minority members in the UK', *Ethnicities* **18**(4), 471–495.

Li, Y. and Heath, A. (2018), 'Persisting disadvantages: a study of labour market dynamics of ethnic unemployment and earnings in the UK (2009–2015)', *Journal of Ethnic and Migration Studies* pp. 1–22.

Lindley, J. (2002), 'Race or religion? the impact of religion on the employment and earnings of Britain's ethnic communities', *Journal of Ethnic and Migration Studies* **28**(3), 427–442.

Longhi, S. and Brynin, M. (2017), *The ethnicity pay gap*, Research Report Series 108, Equality and Human Rights Commission.

Longhi, S., Nicoletti, C. and Platt, L. (2012), 'Explained and unexplained wage gaps across the main ethno-religious groups in Great Britain', *Oxford Economic Papers* **65**(2), 471–493.

Lundberg, S. J. and Startz, R. (1983), 'Private discrimination and social intervention in competitive labor market', *The American Economic Review* **73**(3), 340–347.

Machin, S., McNally, S. and Ruiz-Valenzuela, J. (2018), *Entry Through the Narrow Door: The Costs of Just Failing High Stakes Exams*, CESifo Working Paper Series 7008.

May, T. (2016), ‘Britain, the great meritocracy: Prime minister’s speech’. Delivered on 09/09/2016.
URL: *https://www.gov.uk/government/speeches/britain-the-great-meritocracy-prime-ministers-speech*

Mayer, S. E. and Lopoo, L. M. (2004), What do trends in the intergenerational economic mobility of sons and daughters in the United States mean?, *in* M. Corak, ed., ‘Generational income mobility in North America and Europe’, Cambridge: Cambridge University Press, chapter 5, pp. 90–121.

Mazumder, B. (2005), ‘Fortunate sons: New estimates of intergenerational mobility in the United States using social security earnings data’, *The Review of Economics and Statistics* **87**(2), 235–255.

Mazumder, B. (2015), ‘Estimating the intergenerational elasticity and rank association in the US: Overcoming the current limitations of tax data’, *FRB of Chicago Working Paper No. 2015-04* . http://dx.doi.org/10.2139/ssrn.2620727.

Mazumder, B. (2016), Estimating the intergenerational elasticity and rank association in the United States: Overcoming the current limitations of tax data, *in* L. Cappellari, S. W. Polachek and K. Tatsiramos, eds, ‘Inequality: Causes and Consequences’, Vol. 43, Emerald Group Publishing Limited, pp. 83–129.

McIntosh, S. (2006), ‘Further analysis of the returns to academic and vocational qualifications’, *Oxford Bulletin of Economics and Statistics* **68**(2), 225–251.

McIntosh, S. and Vignoles, A. (2001), ‘Measuring and assessing the impact of basic skills on labour market outcomes’, *Oxford Economic Papers* **53**(3), 453–481.

Metcalf, H. (2009), *Pay gaps across the equality strands: a review*, Research report 14, Equality and Human Rights Commission.

Modood, T. (2003), Ethnic differentials in educational performance, *in* D. Mason, ed., ‘Explaining Ethnic Differences: Changing Patterns of Disadvantage in Britain’, Bristol: The Policy Press, chapter 4, pp. 53–68.

Modood, T. (2004), ‘Capitals, ethnic identity and educational qualifications’, *Cultural Trends* **13**(2), 87–105.

Moffitt, R. A. and Gottschalk, P. (2012), ‘Trends in the transitory variance of male earnings methods and evidence’, *Journal of Human Resources* **47**(1), 204–236.

Narayan, A., der Weide, R. V., Cojocaru, A., Lakner, C., Redaelli, S., Mahler, D. G., Ramasubbaiah, R. G. N. and Thewissen, S. (2018), *Fair Progress? Economic Mobility across Generations around the World*, Washington, DC: World Bank.

Neal, D. (2006), Why has black–white skill convergence stopped?, *in* E. A. Hanushek and F. Welch, eds, ‘Handbook of the Economics of Education’, Vol. 1, Amsterdam: North Holland, chapter 9, pp. 511–576.

Neal, D. A. and Johnson, W. R. (1996), ‘The role of premarket factors in black-white wage differences’, *Journal of Political Economy* **104**(5), 869–895.

Nicoletti, C. and Ermisch, J. (2008), ‘Intergenerational earnings mobility: Changes across cohorts in Britain’, *The B.E. Journal of Economic Analysis & Policy* **7**(2), 1–38.

Nybom, M. and Stuhler, J. (2016), 'Heterogeneous income profiles and lifecycle bias in intergenerational mobility estimation', *Journal of Human Resources* **51**(1), 239–268.

Nybom, M. and Stuhler, J. (2017), 'Biases in standard measures of intergenerational income dependence', *Journal of Human Resources* **52**(3), 800–825.

OECD (2010), A family affair: Intergenerational social mobility across OECD countries, *in* 'Economic Policy Reforms: Going for Growth 2010', Paris, OECD Publishing, chapter 5, pp. 183–198.

OECD (2011), Growing income inequality in OECD countries: What drives it and how can policy tackle it, *in* 'OECD Forum on Tackling Inequality', OECD Paris.

OECD (2015), *In it together: Why less inequality benefits all*, OECD publishing.

OECD (2018*a*), *OECD Regions and Cities at a Glance 2018*, Paris: OECD Publishing.

OECD (2018*b*), *A Broken Social Elevator? How to Promote Social Mobility*, Paris, OECD Publishing.

OECD (2018*c*), *OECD Labour Force Statistics 2017*, Paris, OECD Publishing.

OECD (2019), *Education at a Glance 2019: OECD Indicators*, Paris, OECD Publishing.

Office for National Statistics (2018), *England and Wales 2011 Census*, Available at: https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/population-of-england-and-wales/latest. Accessed: 25 Jan 2020.

Office for National Statistics (2019*a*), *Statistical Bulletin: Regional gross disposable household income, UK: 1997 to 2017*, Available at: https://www.ons.gov.uk/economy/regionalaccounts/grossdisposablehouseholdincome /bulletins/regionalgrossdisposablehouseholdincomegdhi/1997to2017. Accessed: 16 Sept 2019.

Office for National Statistics (2019*b*), *Ethnicity pay gaps in Great Britain: 2018*, Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/ articles/ethnicitypaygapsingreatbritain/2018. Accessed: 27 Jan 2020.

Oster, E. (2019), 'Unobservable selection and coefficient stability: Theory and evidence', *Journal of Business & Economic Statistics* **37**(2), 187–204.

Pagan, A. (1984), 'Econometric issues in the analysis of regressions with generated regressors', *International Economic Review* **25**(1), 221–247.

Palomino, J. C., Marrero, G. A. and Rodríguez, J. G. (2018), 'One size doesn't fit all: A quantile analysis of intergenerational income mobility in the US (1980–2010)', *The Journal of Economic Inequality* **16**(3), 347–367.

Phelps, E. S. (1972), 'The statistical theory of racism and sexism', *The American Economic Review* **62**(4), 659–661.

Piketty, T. (2000), Theories of persistent inequality and integenerational mobility, *in* A. B. Atkinson and F. Bourguignon, eds, 'Handbook of Income Distribution', Vol. 16, Amsterdam: North Holland, chapter 8, pp. 429–476.

Platt, L. (2007), 'Making education count: the effects of ethnicity and qualifications on intergenerational social class mobility', *The Sociological Review* **55**(3), 485–508.

Platt, L. (2019), *Understanding inequalities: Stratification and difference*, Cambridge: Polity Press.

Raaum, O., Bratsberg, B., Røed, K., Österbacka, E., Eriksson, T., Jäntti, M. and Naylor., R. A. (2008), 'Marital sorting, household labor supply, and intergenerational earnings mobility across countries', *The B.E. Journal of Economic Analysis & Policy* **7**(2), 1–49.

Rios-Avila, F. (2020), 'Recentered influence functions (RIFs) in Stata: RIF regression and RIF decomposition', *The Stata Journal* **20**(1), 51–94.

Roemer, J. E. (1998), *Equality of Opportunity*, Harvard University Press.

Roemer, J. E. and Trannoy, A. (2015), Equality of opportunity, *in* A. Atkinson and F. Bourguignon, eds, 'Handbook of Income Distribution', Vol. 2, Amsterdam: North Holland, pp. 217–300.

Sacerdote, B. (2010), Peer effects in education: How might they work, how big are they and how much do we know thus far?, *in* A. Hanushek, S. Machin and L. Woessmann, eds, 'Handbook of the Economics of Education', Vol. 3, Amsterdam: North Holland, chapter 4, pp. 249–277.

Schnitzlein, D. D. (2016), 'A new look at intergenerational mobility in Germany compared to the US', *Review of Income and Wealth* **62**(4), 650–667.

Shah, B., Dwyer, C. and Modood, T. (2010), 'Explaining educational achievement and career aspirations among young British Pakistanis: Mobilizing 'ethnic capital'?', *Sociology* **44**(6), 1109–1127.

Shin, D. and Solon, G. (2011), 'Trends in men's earnings volatility: What does the panel study of income dynamics show?', *Journal of Public Economics* **95**(7-8), 973–982.

Social Mobility Commission (2016*a*), *The Social Mobility Index.*
URL: *https://www.gov.uk/government/publications/social-mobility-index*

Social Mobility Commission (2016*b*), *State of Nation 2016: Social Mobility in Great Britain*, London: Stationary Office.

Social Mobility Commission (2017*a*), *State of Nation 2017: Social Mobility in Great Britain*, London: Stationary Office.

Social Mobility Commission (2017*b*), *Time For Change: An Assessment of Government Policies on Social Mobility 1997-2017*, London: Stationary Office.

Social Mobility Commission (2019), *Social Mobility Barometer: Public attitudes to social mobility in the UK, 2019-20*, London: Stationary Office.

Social Mobility Commission (2020), *Monitoring social mobility. 2013–2020: Is the government delivering on our recommendations?*, London: Stationary Office.

Solon, G. (1989), 'Biases in the estimation of intergenerational earnings correlations', *The Review of Economics and Statistics* **71**(1), 172–174.

Solon, G. (1992), 'Intergenerational income mobility in the United States', *American Economic Review* **82**(3), 393–408.

Solon, G. (1999), Intergenerational mobility in the labor market, *in* O. Ashenfelter and D. Card, eds, 'Handbook of Labor Economics', Vol. 3, Elsevier, chapter 29, pp. 1761–1800.

Solon, G. (2002), 'Cross-country differences in intergenerational earnings mobility', *Journal of Economic Perspectives* **16**(3), 59–66.

Solon, G. (2004), A model of intergenerational mobility variation over time and place, *in* M. Corak, ed., 'Generational Income Mobility in North America and Europe', Cambridge: Cambridge University Press, chapter 2, pp. 38–47.

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

Spence, M. (1973), 'Job market signaling', *The Quarterly Journal of Economics* **87**(3), 355–374.

Strand, S. (2014), 'Ethnicity, gender, social class and achievement gaps at age 16: Intersectionality and 'getting it' for the white working class', *Research Papers in Education* **29**(2), 131–171.

Strand, S. (2015), Ethnicity, deprivation and educational achievement at age 16 in England: Trends over time, Technical report, Department for Education.

Strand, S., Deary, I. J. and Smith, P. (2006), 'Sex differences in cognitive abilities test scores: A UK national picture', *British Journal of Educational Psychology* **76**(3), 463–480.

Strand, S. and Demie, F. (2005), 'English language acquisition and educational attainment at the end of primary school', *Educational Studies* **31**(3), 275–291.

Strand, S., Malmberg, L. and Hall, J. (2015), *English as an Additional Language (EAL) and educational achievement in England: An analysis of the National Pupil Database*, Technical report, Education Endowment Foundation,Unbound Philanthropy, Bell Foundation.

Taylor, M. F., Brice, J., Buck, N. and Prentice-Lane, E. (2010), *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices*, Colchester: University of Essex.

Torche, F. (2015), 'Intergenerational mobility and gender in Mexico', *Social Forces* **94**(2), 563–587.

University College London, UCL Institute of Education, Centre for Longitudinal Studies (2018), *Next Steps: Sweeps 1-8, 2004-2016: Secure Access.[data collection]*, number SN: 7104, UK Data Service, http://doi.org/10.5255/UKDA-SN-7104-4.

Weiss, A. (1995), 'Human capital vs. signalling explanations of wages', *Journal of Economic Perspectives* **9**(4), 133–154.

Wilson, D., Burgess, S. and Briggs, A. (2011), 'The dynamics of school attainment of England's ethnic minorities', *Journal of Population Economics* **24**(2), 681–700.

Zimmerman, D. J. (1992), 'Regression toward mediocrity in economic stature', *The American Economic Review* **82**(3), 409–429.