University College London

Essays on Social Mobility and Education Reform

Andrew Eyles

A thesis submitted to the Department of Economics of University College London for the degree of Doctor of Philosophy.

London, 2024

Declaration

I, Andrew Eyles confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Chapter 1 is a joint project with Jo Blanden and Stephen Machin. Chapter 2 is a sole-authored paper of my own. Chapter 3 is a joint project with Stephen Machin. Chapter 4 is joint work with Stephen Machin and Sandra McNally, and Chapter 5 was jointly written with Stephen Machin, and Shqiponja Telhaj.

Chapter 1 has been published in the Journal of Economic Inequality, Volume 21, April 2023. DOI: https://doi.org/10.1007/s10888-023-09563-z.

Chapter 3 was published in the Journal of the European Economic Association Volume 17, Issue 4, August 2019. DOI: https://doi.org/10.1093/jeea/jvy021.

Chapter 4 as published in the Journal of Public Economics Volume 155, November 2017. DOI: https://doi.org/10.1016/j.jpubeco.2017.09.004.

Acknowledgements

I would like to thank many people who have supported while I worked on this thesis. My supervisors Stephen Machin and Richard Blundell provided invaluable guidance throughout as several colleagues at the Centre for Economic Performance, London School of Economics. I am also grateful for financial support from both UCL and the Economic and Social Research Council.

Many of the chapters contained have been the result of working with excellent coauthors: Jo Blanden, Stephen Machin, Sandra McNally, and Shqiponja Telhaj.

Finally, I would like to thank my friends and family for all their support throughout the PhD.

Abstract

In this dissertation, I use quasi-experimental and machine learning techniques across five separate studies. These studies concern the measurement of social mobility, the determinants of social mobility, and the effectiveness of a large-scale school reform program aimed at increasing prospects for disadvantaged children.

Chapter 1 uses rich data and house prices and their relationship to wealth, to estimate intergenerational wealth persistence in the UK. In line with intergenerational income transmission, it is shown that the intergenerational persistence of wealth has strengthened over time.

Chapter 2 uses cutting edge techniques from the machine learning explainability literature to understand the determinants of upward mobility for two UK birth cohorts. Cognitive ability in adolescence and educational outcomes are shown to be the key drivers of upward mobility. These variables act as sufficient statistics for a wide range of variables related to family background, socioemotional skills, and parental time investment.

Chapter 3 looks at the effectiveness of sponsored academies. It is shown that this radical reform gave already existing schools – particularly those serving disadvantaged pupils - greater operational autonomy. This led to large performance improvements for attendees especially for those in urban schools.

Chapter 4 studies the scaling up of the academies programme to primary schools. Unlike early academies, primary academies led to little performance improvements. Schools that gain academy status make use of their greater fiscal freedom, but do not invest in well-known drivers of school improvement.

Chapter 5 studies how academies affect the labour market for school leaders. Despite the level and variance of head teacher remuneration increasing rapidly in line with the growth of autonomous schools, only a weak link between the two is found. Schools of all types increasingly pay leaders outside of mandated pay scales reflecting a greater liberalisation of the UK head teacher labour market.

Impact Statement

This thesis studies many of the most pressing issues related to the UK economy. Increasing social mobility and the fluidity of wealth is seen as a key policy goal for any prospective government. This thesis highlights how intergenerational wealth transmission has changed over time and a key mechanism that drives the relationship between one's origin and one's destination – house prices.

This complements a well-established literature that looks at intergenerational transmission of income and occupation. Turning to the former, chapter 2 makes a substantive methodological contribution to this literature. Tools from machine learning explainability are introduced to characterise the process of upward mobility. These tools are combined with a latent factor framework and offer a lens through which patterns of mobility can be understood. In this chapter, I argue that these tools offer a substantial improvement on previous methods that have been used to understand trends in intergenerational mobility.

Adolescent cognitive ability and schooling are shown to be strong determinants of upward mobility and the remaining 3 chapters look at a large-scale school reform aimed at improving educational quality for disadvantaged children. By analysing where these reforms worked, where they did not, and why, these chapters provide some of the strongest evidence surrounding the academy school programme. They also add to a growing literature – such as that on US charter schools and Swedish free schools – that looks at the effectiveness of reforms that give state schools greater operational autonomy.

Table of Contents

List of Tables

List of Figures

Chapter 1: Intergenerational Home Ownership

Abstract

This paper studies intergenerational links in home ownership, an increasingly important wealth marker and a measure of economic status. Repeated cross sectional UK data show that home ownership rates have fallen rapidly over time, and markedly amongst younger people in more recent birth cohorts. Evidence from British birth cohorts data supplemented by the Wealth and Assets Survey show a significant rise through time in the intergenerational persistence of home ownership, as home ownership rates shrank disproportionately among those whose parents did not own their own home. Given the close connection between home ownership and wealth, these results on strengthening intergenerational persistence in home ownership are therefore suggestive of a fall in intergenerational housing wealth mobility over time.

1.1 Introduction

A large body of empirical research in social science has assessed the extent to which economic and social outcomes are transmitted across generations. In the economics literature, a heavy focus has been placed on studying earnings or income mobility, and on refining methods to accurately pin down the intergenerational earnings or income elasticity, a measure of how sensitive earnings or income of children (as adults) are to their parents' earnings or income. Some of the more recent work studies changes over time in the intergenerational persistence of earnings or income (see the reviews in Black and Devereux, 2011, Blanden, 2019, or Solon, 1999).

A smaller research focus to date has been on intergenerational housing, assets and wealth, even though intergenerational transmissions of these measures of economic status, and their change over time, are of considerable interest to researchers and policy makers.[1] First, they are key aspects of long-term living standards, and can be used to smooth consumption in the case of income shocks. Returns from housing and non-housing assets and wealth can be used to generate income flows and accumulate further wealth (Fagereng et al, 2020). Second, and highly relevant in the context of this paper, they can be directly passed on to the next generation (Black et al, 2020, Laitner, 2002, Fagereng et al , 2021). Third, wealth and the components of wealth are less equally distributed than income, for example with around half the population having no wealth at all (Keister and Moller, 2000; Piketty, 2014; Wolff, 2016, Crawford, Innes and O'Dea, 2016).

[1] Existing studies with a focus on wealth transmission are single point in time studies (see Mulligan, 1997; Piketty, 2000; Charles and Hurst, 2003; Adermon, Lindahl and Walderström, 2018; Black et al, 2020; and Fagereng, Mogstad and Rønning, 2021). Evidence on changes in the extent of transmission over time is virtually non-existent. The very few studies of relevance to trends in wealth transmission either tend to focus on the richest dynasties rather than the relationships found among the majority of the population (for example, Piketty, 2014; and Clark and Cummins, 2015) or study the impact of parental wealth on child economic or social outcomes (for example, Pfeffer, 2018, documents the growing importance of wealth for children's educational outcomes in the US).

The objective of this paper is to study intergenerational transmissions of home ownership in detail. Home ownership is associated with numerous positive outcomes including financial security, political engagement, higher quality accommodation, and better outcomes for children (Dietz and Haurin, 2003, Zavisca and Gerber, 2016, Goodman and Mayer, 2018). It is also a key marker of wealth; especially as higher house prices have made home ownership more valuable and a key marker of economic status in society. Indeed, housing equity is the largest component of overall wealth in the US (Wolff, 2017), Great Britain (Crawford, Innes and O'Dea, 2016), and in continental Europe (Jantti et al 2008).

Therefore, inequality in home ownership is potentially an important driver of economic inequality. This has received attention in the context of the black-white wealth gap in the US (Charles and Hurst, 2002, Boehm and Schlottman, 2004) and growing inequalities between older and more recent cohorts in the UK (Griffith, 2011; Cribb et al, 2016; Clarke et al, 2016). In addition, Aaronson (2000) and Pfeffer (2018) confirm the centrality of home ownership to the intergenerational impacts of wealth. This work shows that the connection between wealth and children's educational outcomes is well-proxied by home equity and home values, while Pfeffer and Killewald (2018) show that home value is an excellent proxy for net wealth when measuring the intergenerational persistence of wealth in the US.

The increased importance of home ownership is especially pertinent in the UK context where house prices have grown particularly fast by international standards, as demonstrated in Figure 1.1. As returns to housing tenure have outstripped returns to other financial assets, the importance of getting onto the 'housing ladder' has increased as a determinant of wealth accumulation over the course of one's life. This has led to concerns about younger individuals struggling to get onto the ladder when compared to previous generations.

A recent narrative is that young people's initial forays into the housing market are increasingly being funded by the 'Bank of Mum and Dad' (as discussed in Wood and Clark, 2018). In the UK, the proportion of first-time buyers who report receiving direct contributions from family and friends towards a deposit increased from 22% to 29% between 1996 and 2016 (Department for Communities and Local Government 2017, reporting on the English Housing Survey). An important role for parental background also emerges in the work of Lindley and McIntosh (2019) who show that, even among young people with professional and managerial occupations, those with parents from higher social classes have a higher probability of home ownership.

Figure 1.2 shows trends in home ownership over time from the UK Labour Force Survey between 1996 and 2016. These data reveal a dramatic fall in homeownership rates among the young (aged <35), which accelerated after the 2007 financial crisis as rates fell from 59 percent in 1996 to 54 percent in 2004, through to 46 percent in 2008 reaching a low of 34 percent in 2016. Falls among those aged 35-44 began later (only after 2007), but are also striking, falling from 78 percent in 1996 to 68 percent by 2016. [2]

Figure 1.2 is suggestive that trends in home ownership differ markedly by cohort, with successive cohorts becoming less likely to buy. To show this more clearly, Figure 1.3 presents coefficients on year of birth from three descriptive regression models of home ownership containing cohort, age and time effects. To identify cohort effects separately from age and year effects, the coefficient on the 1958 birth cohort is normalised to be zero (1958 is the first birth cohort used in the empirical analysis in this paper).

Coefficients from the first model, shown by the solid line in the Figure, do not account for any differences in factors that might predict home ownership, other than age

[2] The focus in this Figure is on people who are the head of their household (or the head's partner) so changes in home ownership rates among younger groups will be influenced by the age at which young people form independent households.

and year. They show that home ownership rates differed little for the older, 1936 to 1956, birth cohorts. This markedly contrasts with the sharp decline in ownership seen for those born later. The observed decline in ownership seen in the Figure, as those born in the early 1990s are 33 percentage points less likely to own a home than those born in 1958. The peak to trough differential – between birth cohorts 1946 and 1990 – is even larger at 37 percentage points. These large cohorts effects show a negative secular trend in home ownership for successive birth cohorts that only begins to plateau around 1990. Importantly, as shown by the other two set of cohort coefficient estimates in the Figure, which partition out the effects of family structure and income on ownership, these changes do not appear to be accounted for by changing family structure and/or the income distribution of the population.

These descriptives make clear the increasing difficulties that young people have been facing in accessing the housing market. The key focus and contribution of this paper is to hone in on the intergenerational dimension of this by asking to what extent buying has become especially difficult for those whose parents did not own their own home when they were growing up. It is perhaps surprising that this question has not received that much attention in social mobility research to date. This is all more the case as many data sources do contain housing tenure data for children and parents at different points in time, permitting analysis of trends in intergenerational correlations in home ownership.[3] This paper presents evidence on this from a variety of UK data sources over time. For different cohorts, an individual's home ownership status is related to that of their parents when they were young. A consistent picture emerges – those that reside in owner occupied housing as children are much more likely to themeslves be home owners in middle age.

[3] A notable exception is Jenkins and Maynard (1983) who investigate this issue using data from the Rowntree Study of families in York, with the second generation observed in the late 1970s.

As just noted, and importantly, it is possible to study trends. The analysis finds strong evidence of a significant rise in the intergenerational persistence of home ownership, in particular between 2000 and 2010, the period when younger people were finding it increasingly difficult to get into the housing market. By extending this cross-time analysis, beginning with wealth differences between home owners and renters, and studying empirical connections between home ownership, home value and wealth, we conclude that the intergenerational home ownership imply that the UK has likely also experienced a fall in intergenerational housing wealth mobility over time.[4]

1.2 Data and Methods

1.2.1 British Birth Cohort Studies

The earliest data we have available to study intergenerational home ownership comes from the British birth cohorts – the National Child Development Study (NCDS), a cohort born in 1958, and the British Cohort Study (BCS), a cohort born in 1970. The target sample for each cohort consisted of all babies born in a single week, with around 18,000 included at the start. They have been followed up regularly from birth, throughout childhood and into adulthood with the most recent surveys occurring at around age 62 for the NCDS (but not yet released) and age 46 (in 2016) for the BCS. These data have been extensively used to examine intergenerational mobility in income (Dearden et al, 1997; Blanden et al, 2004; Gregg et al, 2017) and in social class (Erikson and Goldthorpe, 2010).

[4] This aspect of the paper has some cross-over with a recent working paper by Gregg and Kanabar (2021) who use two sample two stage least squares based on parental age, home ownership and education level to impute parental wealth and calculate the intergenerational transmission of wealth for the UK. Their estimate of the rank correlation of wealth based on the Wealth and Assets Survey is slightly lower than ours but their results confirm that intergenerational wealth persistence is rising, albeit over a much shorter period than the one considered here.

The analysis focusses on household tenancy which is collected at various points during childhood. We use the measure obtained at age 16, as that is more comparable with the other data used in the paper. The main outcome measure for the cohort members is a measure of owner occupancy at age 42, in 2000 for the NCDS and 2012 for the BCS, supplemented with data collected at age 50 and 55 in the NCDS and at age 46 in the BCS. We combine outright ownership and buying with a mortgage into the category 'owner occupation'.

In addition to information on housing tenure, we make use of information on wealth assets held in several types of savings and investments for NCDS cohort members in 1991 (at age 33). These include bank accounts, stocks and shares and property aside from the main residence. The British Cohort Study at age 42 also asks about home value, mortgage outstanding and the value of savings and debt. This allows us to generate a simple measure of wealth. However, the distribution of this variable compares poorly with the wealth data from the WAS in 2011 so we do not use it in our main analysis. However, results obtained based on the individual's percentile in this wealth distribution are broadly comparable with those from the WAS in 2011.

In forming our samples, we select all cohort members with information on the variables of interest, this is most commonly home ownership for the cohort members and their parents. We might be concerned about attrition given that the cohorts have been followed from birth and require information on their housing tenure at age 42. Table A1.1 gives information about initial and final sample sizes in both cohorts, detailing where observations are lost. The patterns in the two cohorts are somewhat different, with the NCDS experiencing a large sample loss up to age 11, and the BCS samples continuing to fall to age 16. It is notable that the final samples in the two cohorts are much larger than those used to measure intergenerational income mobility in, for example, Blanden, Gregg, and Macmillan (2013). The Appendix of Blanden, Gregg and Macmillan (2013) examines the attrition in the income samples and concludes that it is unlikely to be responsible for the increase in income persistence that is found, we are therefore confident that attrition is not driving the direction of travel found using these larger samples.

1.2.2 The Wealth and Assets Survey (WAS)

The WAS is a household survey that aims to provide a comprehensive overview of the total assets and liabilities of households in Great Britain. 30,959 households were sampled at the initial wave and these households were followed up in subsequent waves. Our analysis makes use of data from Waves 1-5. Each wave covers two years with wave 1 covering 2006-2008 and Wave 5 covering 2014-2016. The WAS collects extensive information on wealth and its sources, including housing tenure, so that owner-occupancy can be defined in the same way as in the cohort studies.

The WAS can used for intergenerational analysis because it collects retrospective information, for those aged over 25, about economic conditions as a teenager. We use the information about the tenancy status of one's parents at age 14 to estimate the intergenerational home ownership transmission for the individuals in the WAS.

The samples used in the WAS are motivated by the need to be comparable with the ages when the cohort members were surveyed. We select individuals who are 40-44 to be comparable with the age 42 data and age 32-36 to be comparable with the age 33/34 data that we use to investigate wealth as an outcome. Our analysis focuses on the household reference person. The focus on the household reference person leads to a slight oversampling of men. In our age 42 samples in 2011(wave 3) and 2015 (wave 5), 60% of our sample are male. This compares with 51% of the NCDS sample and 54% of the BCS. Nevertheless, controlling for gender in our basic specifications does little to alter our results.

As is common in data sets focused on wealth, there is substantial attrition in the WAS, but this is addressed using top-up surveys in later waves. WAS oversamples those living in the wealthiest areas. This is motivated by the fact that total wealth is highly concentrated amongst the wealthiest in society and oversampling this group is necessary to get a comprehensive overview of the nation's total asset holdings. We adjust for this by using cross sectional weights to calculate wealth percentiles. We do not use weights when computing our intergenerational estimates, as nationally representative weights are unsuitable when considering particular age groups as we do here. However, our results are largely unchanged when weights are applied.

1.2.3 British Household Panel Study (BHPS)

Beginning in 1991 the BHPS covered a representative sample of 5,500 UK households and 10,300 adults aged 16 and above. Since then, data covering original sample respondents, and the individuals who reside with them, have been collected on an annual basis. The sample is augmented when original members (including children) leave to form a different household or individuals move in with the original sample members. In 2008, Understanding Society – a larger and more comprehensive study - replaced the BHPS, incorporating the original sample.

While we report ownership correlations using the BHPS, our primary motivation for using the data is that it also collects self-reported data on the value of one's main property for both children and parents. This allows us to calculate the rank-rank relationship between child and parental house values. It is particularly advantageous to measure house values for both the parents and the offspring due to the strong link between wealth and the value of the main residence discussed in the introduction. In principle, one can also measure wealth in the BHPS. Previous work has used the wealth modules in the BHPS to paint a picture of how wealth is distributed in the UK (Crossley and O'Dea; 2010). Using the same data for intergenerational analysis is somewhat problematic. Once individuals are matched to their parents and non-missing or non-conflicting wealth data are removed, the resulting sample sizes are very small. Karagiannaki (2017) considers the impact of parental wealth on educational outcomes using the BHPS, but this requires data on wealth for only one generation.

The BHPS sample consists of those aged 32-36 (age 33/34 sample) and those aged 41-43 (age 42 sample) in 2015/2016/2017. We also estimate models for 32-36 year olds in 2010/2011/2012. Rather than average outcomes over the multiple years, we retain the 2011 and 2016 records when possible and the earliest record when not (so an individual observed in 2015 and 2017, but not 2016 would have the 2015 record retained). In each case we match with parental records in 1991/1992/1993. We retain parental variables from the earliest of the three years. As individuals must reside with their parents in at least one wave in order to be linked with their parents, our final sample consists of individuals who, at some point during the BHPS data collection, lived with their parents.

As we want to focus on those who match with their parents during childhood and their teenage years, we focus on the offspring of those in the original BHPS 1991 sample. These individuals are between the ages of 12 and 18 in 1991. We then look at the subsample of these aged 32-36 in 2011 (2010/2012 for those that are not observed in 2011) and 2016 (2015/2017 for those that are not observed in 2016) alongside those aged 41-43 in 2016 (2015/2017 for those that are not observed 2016). Our final samples are selected based on comparability with the BCS and NCDS samples (in terms of the age at which we measure outcomes), sample size[5], and the need to match with parents. Amongst those of the relevant age group who match with a parental record, we retain individuals who are household

[5] Focusing on a single age at measurement i.e. looking at only 42 years olds results in very small samples in the BHPS.

reference persons (or the partners of household reference persons). We also consider only those for whom one of their parents is a household reference person in the years when the parental variables are measured.

As we look at rank-rank slopes when assessing the relationship between parental housing wealth and child housing wealth, we need to assign individuals to a percentile of the distribution of house prices. In doing so, we set house values to zero for those who do not own before calculating percentiles on a wave-by-wave basis using the full BHPS sample. Following Chetty et al. (2014), we set the rank of those with zero reported housing wealth to one half of the fraction of the sample reporting zero, i.e. if 20% have no housing wealth this 20% of the sample all have a rank of ten. We do not use household weights when doing this due to BHPS household weights are undefined for large portions of the sample. As will be discussed later, applying weights when calculating percentiles does not affect our results.

1.2.4 Descriptive Statistics

The initial integenerational analysis studies individuals at age 42 and relates their home ownership status to that of their parents when they were growing up. Given home ownership-age profiles, this is a good age at which to study this, as people of earlier ages (certainly in their 20s, but probably also in their 30s) may not have aged enough for home buying opportunities to have yet arisen. A second rationale comes from intergenerational studies which show that age 42 income at this stage of the life cycle is a good measure of permanent income (see, for example, Haider and Solon, 2006), and it is a key point of observation in two of our datasets.

The specific years when we can observe 42 year olds and their parents are as follows: a) In 2000 from the National Child Development Study (NCDS), a cohort born in a week of March 1958, with parental home ownership measured at cohort member age 16 in 1974.

b) In 2012 from the British Cohort Study (BCS), a cohort born in a week of April 1970, with parental home ownership measured at cohort member age 16 in 1986.

c) In 2011 and 2015 from two waves of the Wealth and Assets Survey (WAS) that permit the matching of individuals aged around 42 (40-44) years with their parents' home ownership status recalled from when they were age 14; around 1983 to 1987.[6]

We strive for comparability in terms of the samples and variables used across the datasets, but we are constrained in this because the purpose and design of the datasets is fundamentally different. However we are confident that cross-cohort NCDS and BCS 2000-2012 comparisons and the within-WAS 2011-2015 comparisons are consistent. And, as will be shown below, the estimated intergenerational correlations from 2012 in the BCS and 2011 in WAS are remarkably similar.

Table 1.1 shows descriptive statistics for these main samples. The first two rows shows a fall in the owner-occupancy rate of 42 year olds between 2000 and 2015 from 81 percent to 69 percent.[7] The pattern for the cohort members' parents is notably different with a rise in owner-occupancy from just over 50 percent to over 70 percent between the NCDS observed in 1974 and the first WAS observation that is centred on 1983. It is notable that the statistics for the first WAS survey from 2011 and the BCS in 2012 are extremely similar,[8] giving us confidence that we can extend the trends observed in the NCDS and BCS cohort datasets with estimates based on the Wealth and Assets Survey.

The second block of numbers give an early indication of the extent of intergenerational links by presenting the home ownership rates of 42 year olds by parental

[6] Although we refer to the WAS data as being drawn from 2011 and 2015, the two waves cover multiple years, with the '2011' wave spanning 2010-2012 and the '2015' wave spanning 2014-2016.
[7] This is in line with estimates derived from the Labour Force Survey that show an owner occupancy rate of 81% for 40-44 year olds in 2000 falling to 68% in 2015.
[8] This similarity is despite the oversampling of wealthy areas. This may be driven by high house prices in these areas driving slightly lower home ownership rates than might be expected based on wealth and income.

home ownership status. In all cases, there is a substantial and statistically significant gap between the home ownership rates of those with parents who are home owners and those who did not own their own home. This rose substantially from 2000 (the NCDS) and 2011/12 (the BCS and WAS) increasing from a gap of 14 percentage points to 22 percentage points. The data from the 2015 WAS shows a gap of almost 27 percentage points, indicating a further increase in more recent years.

In order to probe the sensitively of our results to the age of observation we estimate the intergenerational home ownership association at older ages. In the cohort studies we can explore additional information at ages 50 and 55 for the 1958 National Child Development Survey and at age 46 for the 1970 British Cohort Study. As the Wealth and Assets Survey covers the full population we estimate the intergenerational associations in that data up to age 59. Table 1.2 shows descriptive statistics for these samples, confirming that the patterns in home ownership over cohorts and time in these datasets are broadly in line with those observed in Figure 1.2 from the Labour Force Survey.

Some of the same, plus additional, data sources can be used to hone in on the changing relationship between home ownership and wealth. The best source of wealth data is the WAS, which asks detailed information on a comprehensive list of wealth components. The information obtained from the existing five waves of the WAS is largely consistent with the information obtained from adminstrative data (Blanden et al, 2021).

The cohort studies also feature rudimentary information on wealth components, but these are collected sporadically and their quality is variable. We make use of information on the wealth held in several types of savings and investments for NCDS cohort members in 1991 (at age 33). We are also able to examine information on housing wealth for both parents and children for some cohorts in the BHPS, although sample sizes are small.

Table 1.4 shows descriptive statistics for these wealth measures. Columns 1 and 2 provide information on individuals aged 42 in the 2011 and 2015 WAS data. These show mean net wealth of £323k in 2012 prices in 2011, rising to £380k in 2015, with the average value of the main residence and the value of savings and investments also rising (albeit by a smaller amount) over this period. Panels 3-6 provide information on wealth for 33 year olds across the four years when we can observe this group. As we have data from 2007 and 2011 we can observe the decline in household wealth associated with the financial crisis. This is quite steep with mean net wealth declining from £220k in 2007 to £157k in 2011. After 2011 average wealth, home value, and the value of saving and investments stay constant. Panels 7-10 gives the information on housing wealth in the BHPS samples that we make use of, and reassuringly the data on housing values is comparable with the information available from the WAS for the same age groups and years.

In our discussion section we conclude our analysis with a discussion of plausible values for wealth mobility, informed by our results to that point. This exercise requires knowledge of the relationship between total wealth and housing values, for both parents and children. To understand the relationship for the parents of older cohorts we make use of the small samples available in the BHPS data from the 1995 and a one off collection of the English Housing Condition Survey (EHCS) from 1986.[9] Using all the datasets available, we are able to measure the extent to which total wealth and housing wealth correlate for intermittent years between 1986 and 2015.

1.2.5 Methods

In the first, core set of analyses the home ownership status of 42-year-olds in the four survey years between 2000 and 2015 is related to the home ownership status of their

[9] The EHCS is a precursor to the English Housing Survey.

parents when they were a teenager. We use linear probability models of the determinants of home ownership (HO^{42}) for individual *i* in the cohort aged 42 in year *t*.

$$\mathrm{HO}_{it}^{42} = \alpha_t + \beta_t \mathrm{HO}_{it}^{parent} + u_{it}^{42} \qquad (1)$$

where HO_{it}^{42} is a dummy that equals 1 if individual i is a home owner at age 42 in year *t*, each cohort is defined by this year. $\mathrm{HO}_{it}^{parent}$ is the home ownership status of individual i's in cohort *t*'s parents when i was a teenager. The cohort specific intergenerational estimate in equation (1) is given by $\beta_t = \Pr[\mathrm{HO}_{it}^{42} = 1 | \mathrm{OO}_{it}^{parent} = 1 \,|\,]$. The temporal change in intergenerational transmission between t and t' is $\Delta\beta_{t't} = \beta_{t'} - \beta_t$.

Initially we follow the standard approach in the intergenerational literature and do not include any additional controls. We are not attempting to capture the causal effect of parental home ownership on own home ownership, but rather estimating an omnibus statistic that captures the consequence of all the mechanisms that lead to a link between the two, these could include associations in human capital, direct financial transfers and preferences. It is not a goal of this paper to separate out the influence of these transmission mechanisms. However, we also estimate the slightly expanded equation (2) which accounts for basic factors that we know are strongly related to home ownership,

$$\mathrm{HO}_{it}^{42} = \alpha_t + \beta_t \mathrm{HO}_{it}^{parent} + \sum_{j=1}^{J} \gamma_{j,t} X_{it}^{42} + \sum_{j=1}^{J} \varphi_{j,t} X_{it}^{parent} + u_{it}^{42} \qquad (2)$$

where X_{it}^{42} are a set of basic controls related to family structure at age 42 and X_{it}^{parent} considers comparable information for the parents during the child's teenage years. These compositional controls include the sex of the individual, whether they have a partner, whether the father lived with them when they were a teenager, each parent's age, and the square of these. While the choice of controls is to some extent arbitrary, we aim to control for secular changes in family structure that are related to homeownership and are likely to be correlated across generations. Our choice of controls strikes a balance between purging our estimates of the independent effect that changing household composition has played on homeownership and keeping the usual descriptive interpretation of intergenerational estimates. In order to check the robustness of the estimates to lifecycle concerns we also present estimates of (1) for the older observations available in the Cohort Studies and perform a more comprehensive assessment of their sensitivity to age in the WAS as the data structure is less restrictive.

The focus of the paper is on intergenerational mobility in home ownership and the data we have does not enable us to also fully study trends in intergenerational wealth mobility. This is largely because we do not have much information on parents' wealth. However, the data sources used can enable some connections to be made to wealth. First, as already noted, the National Child Development Study and the Wealth and Assets Survey provide some direct information about accumulated wealth for the individuals in the second generation. And second, both the BHPS and the WAS also contains some information on housing values.

These enable the study of three, related issues that connect our intergenerational home ownership analysis to wealth:

i) The first supplements and further contextualises the intergenerational home ownership analysis with models which relate wealth in the early 30s (because this is the age when the data is available for the NCDS cohort members) to parental home ownership. Wealth is measured by rank within the distribution of wealth in the sample, and the analysis relates child wealth to parental home ownership as follows:

$$\mathrm{W}_{\mathrm{it}}^{\mathrm{30s}} = \delta_{0t} + \delta_{1t}\mathrm{HO}_{\mathrm{it}}^{\mathrm{parent}} + u_{\mathrm{it}}^{\mathrm{30s}} \quad (3)$$

ii) The second estimates point in time BHPS intergenerational home value transmission parameters, η_1, as:

$$\mathrm{HV}_{\mathrm{i}}^{42} = \eta_0 + \eta_1 HV_{\mathrm{i}}^{parent} + u_{\mathrm{i}}^{42} \tag{4}$$

iii) The third uses data from several sources to relate housing value data (HV) to wealth rank for both generations by estimating the following measurement equations for age 42 individuals, and their parents, in their respective generations, as:

$$HV_{\mathrm{i}}^{\mathrm{k}} = \pi_0^{\mathrm{k}} + \pi_1^{\mathrm{k}} W_i^{\mathrm{k}} + \omega_{\mathrm{i}}^{\mathrm{k}} \ \{\mathrm{k} = \text{parent}, 42\} \tag{5}$$

These can be combined with the estimate of $\eta_1$to provide an indication of the level of intergenerational wealth mobility, $\theta_1 = \frac{\pi_{1t}^{42}}{\pi_{1t}^{parent}} \eta_1$ where $\mathrm{W}_{\mathrm{i}}^{42} = \theta_0 + \theta_1 W_{\mathrm{i}}^{parent} + e_{\mathrm{i}}^{42}$.

The estimates of δ_{1t}, π_1 and η_1 in equations (3) to (5) allow us to use a patchwork of data to end the paper with a suggestive picture of the intergenerational transmission of housing wealth and its trend over time. This needs the caveat that more research with better data on parental wealth for multiple generations is needed to shed more light on the temporal evolution of θ_1, and that this offers an important challenge for future research.

1.3 Trends in Intergenerational Home Ownership

Table 1.4 reports trends in the intergenerational persistence in home ownership by presenting estimates of β_t at or around age 42 from equations (1) and (2) for four years (t = 2000, 2011, 2012 and 2015) and of $\Delta\beta_{t't}$ between 2000 and 2015. Panel A shows estimates of the basic unconditional intergenerational transmission. Panel B adds a set of composition variables measuring characteristics of individuals and their parents. The first four columns of Panel A show the extent of intergenerational transmission of home ownership. For the earliest cohort of 42 years olds – the 1958 birth cohort observed in the year 2000 – home ownership is around 14 percentage points higher for those whose parents owned their own property in 1974.[10] This increases to 22 percentage points in both 2011 and 2012 and even further to 27 percentage points by 2015.[11] Column (5) indicates that by 2015, the dependency between the home ownership status of 42 year olds and that of their parents is much stronger than it was in 2000. [12]

Panel B confirms that these patterns are robust to the inclusion of basic composition controls. The change over time between both 2000 and 2011/12 and from 2011/12 to 2015 reduces slightly on their inclusion, but the overall increase in intergenerational persistence is still strongly significant.[13]

Table 1.5 reports information based on a wider sample of the WAS, to check trends for robustness across age groups. The columns report results in four five-year age bands from 40-44 to 55-59 with the rows reporting estimates of the unconditional persistence in home ownership for each waves 1-5. The pattern over time is extremely consistent with the results shown in Table 1.4, revealing a rise in the intergenerational association in home ownership for all age groups observed in 2014-16 compared to 2006-2008. Owing to fairly small sample sizes the change over time is only significant at the 95% level for the 50-54 age group where it rises from 0.156 (0.020) to 0.223 (0.024) over the (approximately) eight year period of observation. Overall, the estimated coefficients decline as we look at older groups.

---

[10] When parental home ownership at age 10 is the main explanatory variable the coefficients are 0.120 and 0.200 for the NCDS and BCS respectively, the change is almost identical to the results based on measures at 16. It is notable that associations are slightly stronger for ownership at 16 as owner occupation in the teenager years is available for the majority of our datasets.

[11] The log odds ratios for the upper panel are 0.946 (0.059), 1.011 (0.113), 1.063 (0.065), and 1.167 (0.134).

[12] Using longitudinal weights in the WAS 2011 sample (which adjust for attrition between waves 1 and 3) inflates our estimate of homeownership persistence to 0.236 (0.037). Cross sectional weights applied to the same sample shift the coefficient to 0.230 (0.030). Looking at wave 5, applying weights leads to two estimates that sandwich our unweighted coefficient - longitudinal weights increase our estimate to 0.300 (0.061), while cross sectional weights shrink the coefficient to 0.243 (0.035). Even in the latter case, there remain a large discrepancy between the intergenerational relationship measured in 2000 using the NCDS and the relationship measured 15 years later in WAS.

[13]The slight reduction in the change in coefficients is driven by the inclusion of the individual's partnership status. Those with parents who are owner occupiers are more likely to be in a partnership at age 42, and those with partners are more likely to own their own home.

Table 1.5 also reports the supplementary information available from the NCDS and BCS for older ages. Whilst the sparser data points available means that it is not possible to compare these cohorts at the same age beyond age 42, the evidence we have supports the finding of a substantial rise in intergenerational persistence between these cohorts. In the cohort data, there is little evidence of a decline in the intergenerational association of home ownership as individuals' age. This contrasts with patterns by age in results for intergenerational income mobility which show a clear rise in persistence as individuals move into their late 40s and 50s (Gregg et al, 2017). This difference might be a consequence of our measurement's limitations as a binary variable; as home ownership is more prevalent among older groups it is more difficult for it to pick up more nuanced measures of economic wellbeing as people age.

1.4 Home Ownership and Wealth

The British cohort studies only contain limited information on wealth and asset values. Therefore, to comment on the implications of intergenerational associations in home ownership for mobility we must look to a broader set of data. Links between wealth and home ownership are studied primarily using the Wealth and Assets Survey. Several aspects are considered, beginning with wealth differences between home owners and renters, before considering the relationship between wealth and housing value. We also use the WAS to consider changes in the relationship between wealth and parental home ownership over time and are able to supplement our findings with partial information on wealth for an earlier cohort, which again can be related to parental home ownership. Using the BHPS we use information on home value to get close to estimating the intergenerational transmission in wealth, before considering the implications our findings for trends in wealth mobility.

1.4.1 Home Owners and Renters

First consider differences in wealth between home owners and renters.[14] Figure 1.4 draws on 2011 and 2015 WAS data to show real (2012 prices) levels of household wealth across the four possible combinations of individual and parental home ownership status. The Figure shows that home owners whose parents also owned their home have the highest mean wealth levels in both years and that, if anything, there are bigger wealth gaps connected to intergenerational home ownership in 2015.

1.4.2 Housing Wealth

Figure 1.5 considers connections between wealth and more detailed measures of housing wealth – the value of the main residence (home value) and the same value less any outstanding mortgage on the property (home equity) – showing mean wealth percentile rank plotted against home value or equity percentile rank. There are strong associations, and, whilst home equity has the strongest relationship with a rank-rank slope of 0.829, there is also a strong relationship between the value of one's main residence and wealth. Moving up ten percentiles in the distribution of house values moves a household, on average, 7.4 percentiles up the wealth distribution. The relationship is shown as strongly linear, offering support for our measurement framework which focuses on linear relationships between wealth ranks and rank in home value.

1.4.3 Wealth and Parental Home Ownership

The strength of these contemporaneous relationships between housing tenure, housing wealth, and total wealth suggests that trends in the intergenerational assocations between parental and child housing variables may be indicative of trends in wealth mobility. Ideally, we would have wealth data for multiple cohorts of individuals matched to the wealth

---

[14] In practice, those who do not own a home could live rent free, squat, or report 'other' as a form of housing tenure. For simplicity, this group is referred to as renters as renting is by far the largest form of tenure amongst those who do not own their own home.

of their parents. This does not exist, but the Wealth and Assets Survey does allow us to look at the relationship between the percentile rank of an individual in the wealth distribution and their parents' home ownership status.[15]

The results shown in Table 1.6 focus on 42 year olds in 2011 and 2015 and extend the intergenerational model to look at the relationship between wealth and parental home ownership. The upper Panel A of Table 1.6 reproduces the home ownership results, while Panel B considers the relationship between wealth percentiles and parental home ownership. Whilst it comes as no surprise that those whose parents owned their home are significantly wealthier, it is also shown that the association between wealth percentile rank and parental home ownership rises across the two years: going from 15 to 19 percentile points.[16]

The data sources other than the WAS are more limited in the data they contain on wealth. The NCDS does contain information on the value of investments and savings, but only collects this in the 1991 wave at age 33 (rather than age 42, - the primary age of interest in this paper). Despite this, the information is useful as it can be used to generate a further cross-time comparison point prior to the WAS. Results for 33/34 year olds are shown in Table 1.7. As the main analysis reported earlier was presented only for the 42 year olds, the upper Panel shows the intergenerational home ownership transmission trends for this younger age group. A similar finding arises, with there being a sizeable increase in intergenerational home ownership persistence over time. In the NCDS in 1991, there is an 18 percentage point gap in ownership between the two groups, which rises to 32 percentage points by 2007 and further to 35 by 2015. In general, the intergenerational associations are

[15] Strictly speaking the WAS asks about owner occupancy of parents during teenage years but prompts individuals to use age 14 as a benchmark.
[16] A similar result holds if we consider the logarithm of total wealth. Unlike housing values, there are less concerns about individuals with zero wealth meaning that partial elasticities, with log of wealth as the dependent variable, are less problematic than elasticities that focus on housing wealth alone. Nevertheless, we focus on ranks so that our results are comparable across our various specifications. Results using the Log of total wealth are available in the notes to Table 4.

slightly stronger among this younger group, indicating that as individuals get older those who do not come from home-owning families catch up slightly (in terms of home ownership) with those who do.

Panel B of Table 1.7 considers the relationship between savings and investments and parental ownership. In 1991 savings and investments were 13 percentile points higher for NCDS cohort members whose parents were home-owners, and this rises to 17 percentile points higher in the 2015 WAS. The 4 percentile point rise shown is column (5) is on the margins of statistical significance (with low precision due to small WAS sample sizes), but in line with the results of Tables 1.4, 1.5A and 1.5B is suggestive of a strengthening relation between wealth and parental home ownership.

1.4.4 Intergenerational Wealth and Asset Correlations

The results so far show an increase in the intergenerational transmission of home ownership and, at the same time, a strengthening empirical association between wealth and parental home ownership. Figure 1.5 showed an almost one-to-one relationship between housing values and net wealth. There is one UK data source - the British Household Panel Survey (BHPS) - where it is possible to look directly at intergenerational correlations in housing values to study housing wealth correlations. Although there are clear limitations owing to limited sample size, this is potentially informative about overall wealth persistence.

As previously discussed, the BHPS began in 1991 and allows intergenerational matching between original sample members and their offspring from then onwards. Table 1.8 shows results from the BHPS for a sample focused around age 42 in 2016 (i.e. people born in 1974 who would be aged 17 in 1991 and who are intergenerationally matchable as they would still be living in the parental BHPS household) and for those around age 33/34 in 2011 and 2016.

For these samples, panel A of Table 1.8 shows what happens when we reproduce the earlier intergenerational home ownership regressions. Despite the small sample sizes, the results for the BHPS are strikingly consistent with the results presented earlier from the other datasets. The estimates are numerically extremely close. And, as with the earlier analysis, the coefficient from the linear probability regression of home ownership on parental home ownership is larger for those observed in their 30s as compared to those observed in their 40s. Moreover, there is again evidence of increasing persistence from 2011 to 2016 but, with sample sizes of 330 in 2011 and 211 in 2016 this increase is very imprecisely determined.

The strong similarity of the intergenerational home ownership transmission here and in other data gives us confidence to look more closely at the BHPS asset value data in these samples. Results in Panel B show the relationship between parental home ownership and individuals' home value. Those whose parents owned their own home are 25-30 percentiles higher in the distribution of housing value in early middle age than those whose parents rented. Results for 42 year olds are broadly comparable with and corroborate the WAS estimate in Table 1.6. Finally, the results in Panel C measure the intergenerational association in home values between the two generations. The results show a rank correlation in a range of 0.36 to 0.42 between housing value across generations.[17]

In the methods section above we showed that $\theta = \frac{\pi_1^{42}}{\pi_1^{parent}}\eta$ where θ is the intergenerational wealth correlation, η is the intergenerational correlation in housing value and the πs project house value onto wealth in each generation. Table 1.9 reports estimates of π_{1t} from the WAS and two earlier sources, the EHCS in 1986 and the earliest data at

[17] As noted in the data section, household weights are frequently undefined in the BHPS. However, applying weights when calculating percentiles leads to an identical point estimate for the rank slope for 42 year olds despite the sample size falling from 168 to 116.

which wealth data is collected in the BHPS - 1995. The WAS provides the most reliable estimates of π_{1t}^{42} (i.e. the projection from wealth to housing value for the adult children) and suggests that there was little change in π_{1t}^{42} in the 2010s. In order to gauge π_{1}^{parent} we need to go further back in time, and make use of data with smaller sample sizes and therefore less reliability. The EHCS data from 1986 and the BHPS data from 1995[18] provide alternative but similar estimates of π_{1}^{parent} for those aged 42 years old in 2016.[19] The 1995 BHPS gives an estimate of π_{1}^{parent} for those who were in their 30s in the BHPS in the 2010s. Looking across the whole of Table 1.9 the results indicate no substantive difference between π_{1}^{42} and π_{1}^{parent} implying that the level of the intergenerational transmission of housing wealth is a good indicator of the level of the intergenerational persistence of total wealth; in the range of 0.36 to 0.42 as shown in Table 1.8.

These results indicate that the point in time intergenerational housing wealth persistence is higher than comparable estimates of intergenerational income persistence in the UK (Blanden et al 2013, Gregg et al, 2017, and Rohenkohl, 2020 suggest that income rank persistence is around 0.30-0.35).[20] It is notable that this pattern is in line with results in Charles and Hurst (2003) for the US.

Finally, in the light of the findings so far, what about trends in intergenerational wealth mobility? The discussion in the earlier methods sub-section of the paper made it clear that what we can do here is limited as we do not have data on multiple child-parent wealth over

---

[18] The BHPS measure of wealth excludes pension wealth, but includes savings, investment assets such as ISAs, debt outstanding and home equity. We do not use BHPS wealth data in our main sample due to the low sample size once individuals are matched to their parents. Longitudinal matching on wealth data is made difficult in the BHPS as wealth data is only collected sporadically. The earliest collection is 1995 and the latest is in wave 12 (2016/17). Once individuals are matched and those with non-missing wealth observations are retained, sample sizes become too small for meaningful analysis.

[19] This cohort would have been 12 in 1986 and 21 in 1995.

[20] Estimates from our own age 42 sample, in the 2016 BHPS, accord closely with a coefficient and associated standard error of 0.317 (0.085)

time. However, we can say something in the spirit of the patchwork discussion about wealth presented earlier in the paper.

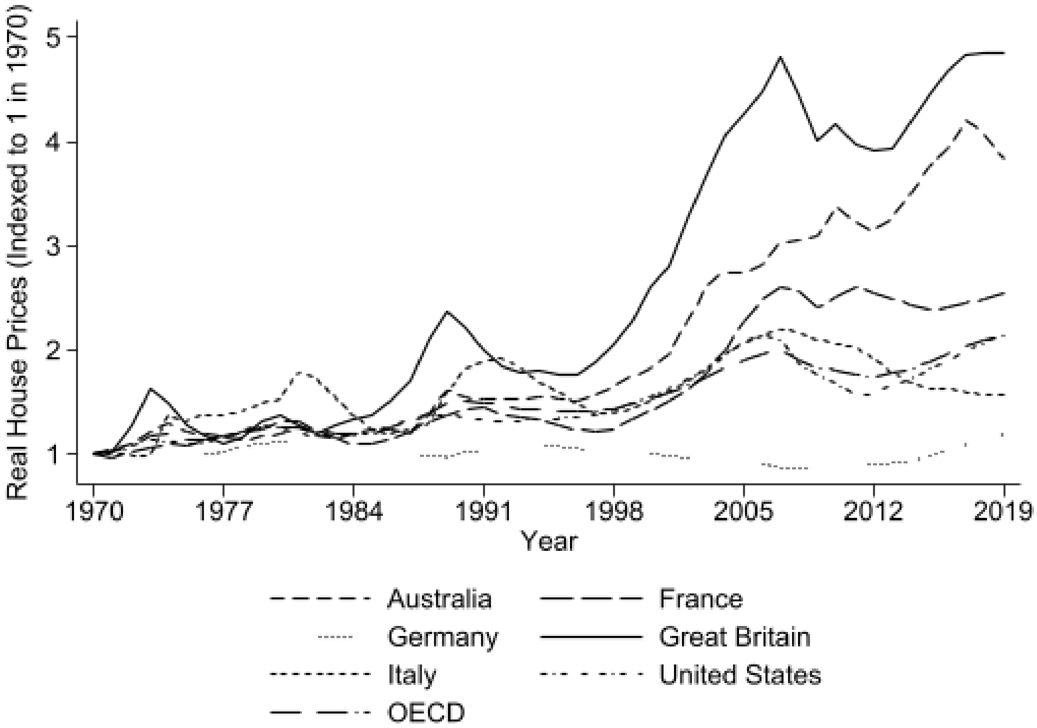
First of all, the evidence in Table 1.9 showed the relationship between home value and wealth to be steady over the sample period studied, i.e. $\frac{\pi_{1t}^{42}}{\pi_{1t}^{parent}}$. Under the admittedly strong assumption that the relationship between home ownership and wealth has also remained constant over this period, it is possible to say something about trends in the intergenerational transmission of wealth.

Therefore, because the results for β_t in Tables 1.4, 1.5A and 1.5B show there to have been a clear and marked increase in the extent to which parental home ownership determines children's home ownership in midlife, this implies that if we had data on child and parent wealth for similar cohorts, it too would reveal a rise in the persistence of intergenerational wealth (unless something else that we have not considered here is going on to offset this direction of travel). However, at this juncture this can only be taken as suggestive. We clearly need more research on this question to better validate this conclusion.

1.5 Conclusion

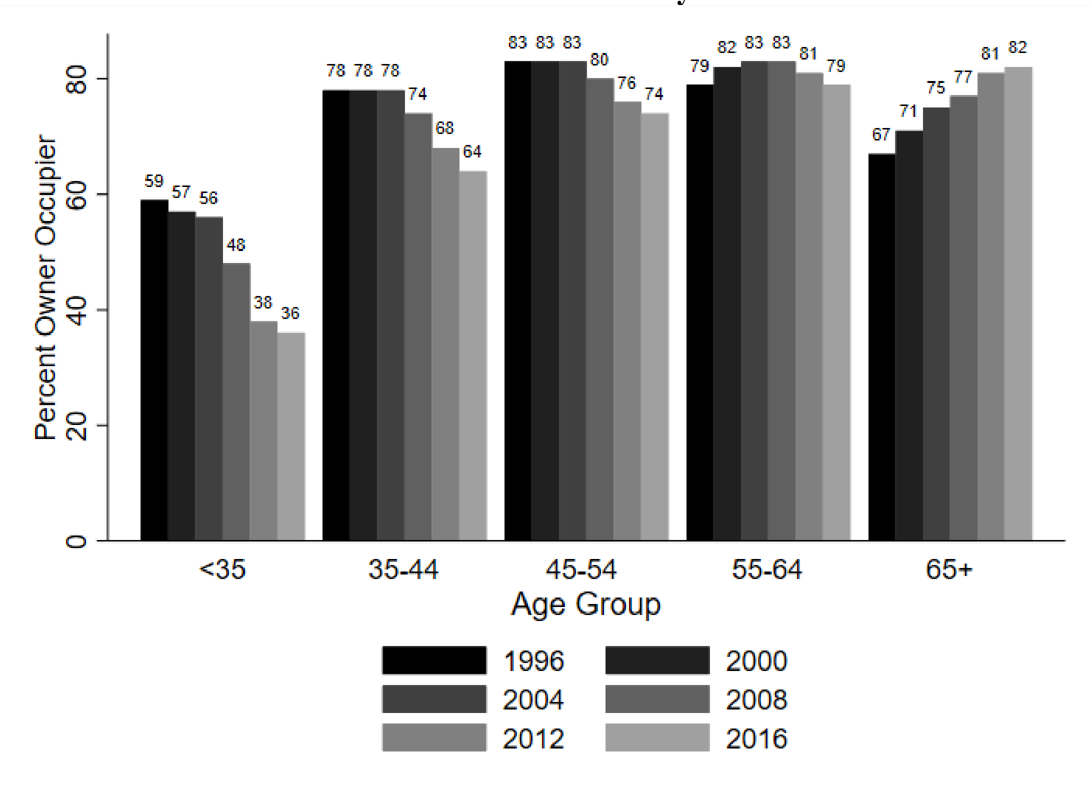
This paper focuses on an understudied area of social mobility and inequality research by studying intergenerational home ownership. Using UK data on home ownership of parents and children, it uncovers a strong intergenerational persistence that has become stronger over time. Indeed, the intergenerational persistence of home ownership status increased substantially between 2000 and 2016, as UK house prices rose sharply and young people's position in the labour market weakened (Costa and Machin, 2017). These made getting on the housing ladder much more difficult for people from more recent birth cohorts whose parents did not own their own home. Given the close connection between home ownership and wealth, these results on strengthening intergenerational home ownership are therefore also suggestive of a fall in intergenerational housing wealth mobility over time, though this latter question should be firmly on the agenda for future research to further probe and assess.

Figure 1.1: House Price Growth, 1970-2019



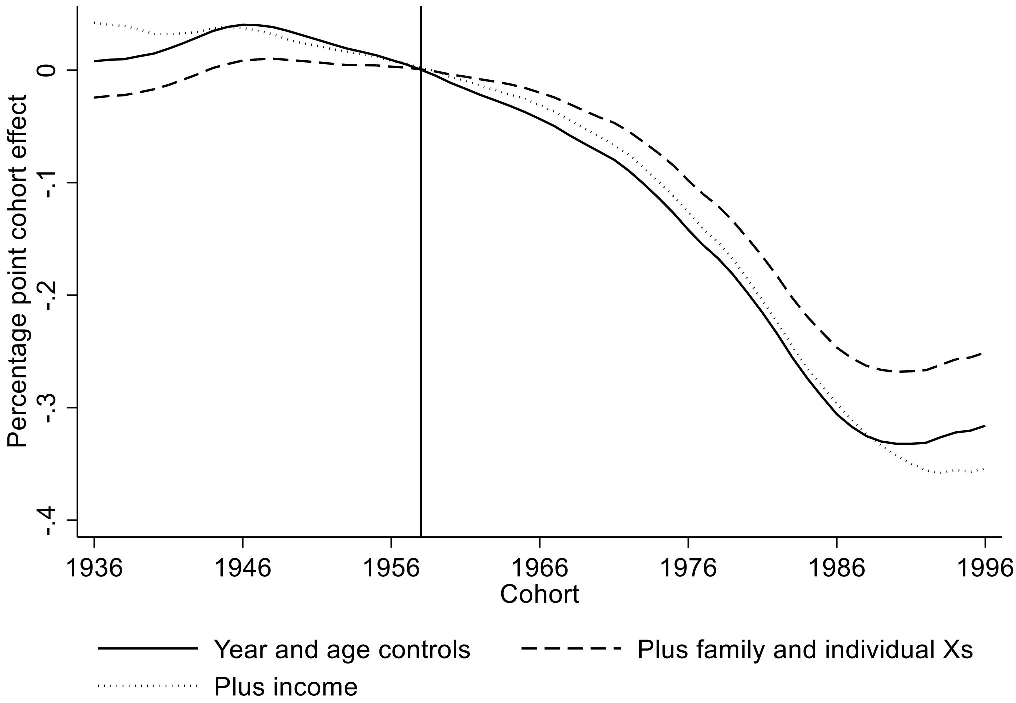
Notes: Author's own calculations using OECD house price indices. Figure refers to real house price growth.

Figure 1.2: Patterns of Home Ownership in the UK across time and age group, Labour Force Survey



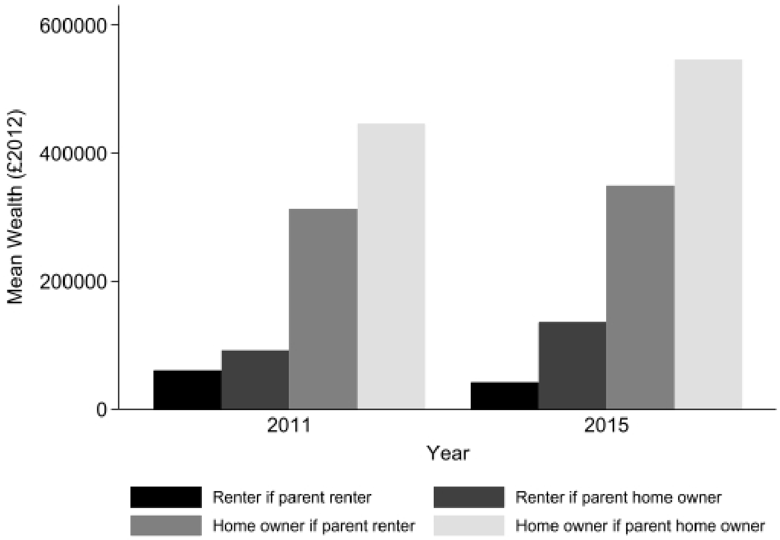
Notes: Labour Force Survey data from 1996 to 2016. The sample of observations is limited to household reference persons. Data are weighted using person weights provided by the LFS.

Figure 1.3: Cohort Effects on Home Ownership from the Labour Force Survey



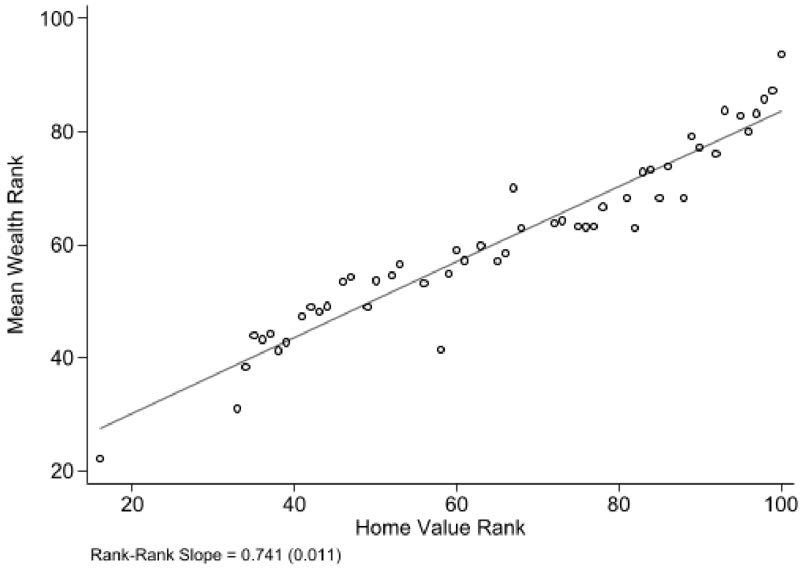
Notes: Labour Force Survey data from 1996 to 2016. The sample of observations is limited to household reference persons aged 20-69. Individual controls are gender, maritial status, number of dependent children, ethnicity and, in the case of the dahsed line, gross weekly income entered as a percentile in the annual wage distribution. Percentiles are calculated using LFS income weights. All three lines are based on coefficients from the common sample of individuals with full data on characteristics and income. In order to seperately identify the effect of cohort from age and year, we normalise the cohort effect to be 0 for individuals aged 42 in the year 2000 (those born in 1958 as indicated by the vertical line in the Figure). Coefficients are smoothed over a using a 5 year rolling window.

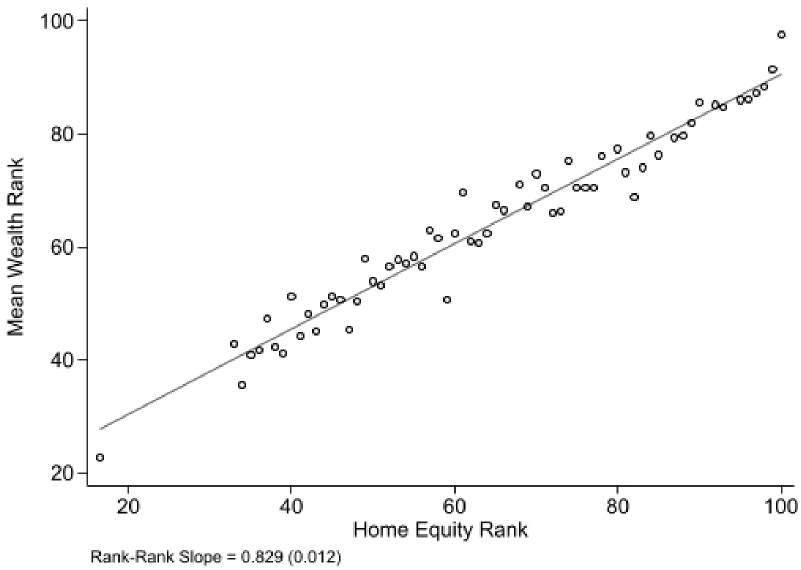
Figure 1.4: Wealth and Parental Home Ownership



Notes: Figure 1.4 uses total net wealth data provided by waves 3 and 5 of the Wealth and Asset Survey. Age and ownership are measured with respect to the household reference person. Results are averaged over ages 40-44 to avoid small sample sizes. Total wealth is in 2012 prices

Figure 1.5: Wealth and Home Value or Home Equity





Notes: Figure 1.5 plots the average percentile of wealth within each percentile bin of home equity and home values using data from the 2015 WAS. Bins are not of equal size because percentiles are calculates using all ages and household weights. As a result of this, we remove bins with fewer than five observations. Rank-rank slopes are calculated from the underlying microdata.

Table 1.1: Data to Study Trends in Intergenerational Home Ownership, Descriptive Statistics

	NCDS	WAS	BCS	WAS
	(1)	(2)	(3)	(4)
% Home owner	81.0	71.1	75.4	68.6
% Parent home owner	51.3	72.8	76.5	74.1
% Home owner if parent home owner	87.9	77.1	80.5	75.5
% Home owner if parent not home owner	73.7	55.1	58.8	48.9
Percentage point gap	14.2 (0.9)	22.0 (2.6)	21.7 (1.4)	26.6 (3.1)
Home ownership year	2000	2011	2012	2015
Parent home ownership year	1974	1983	1986	1987
Sample Size	8352	1771	6181	1271

Notes: The NCDS and BCS are single year birth cohorts matching cohort members at age 42 to parents at age 16. The WAS are multiple year birth cohorts matching individuals aged 40-44 (with centred age 42) to parents at age 14. Standard errors are reported in parentheses.

Table 1.2: Descriptive Statistics for Older Samples

	Wealth and Assets Survey			
	Ages 40-44	Ages 45-49	Ages 50-55	Ages 55-59
Wave 1: 2006-2008				
% Home owner	75.8	79.0	79.5	82.1
% Parent home owner	65.7	61.8	53.8	47.5
Wave 2: 2008-2010				
% Home owner	73.1	75.8	78.8	81.2
% Parent home owner	70.1	63.0	56.2	49.7
Wave 3: 2010-2012				
% Home owner	71.1	73.9	78.5	78.6
% Parent home owner	72.8	65.4	59.9	52.1
Wave 4: 2012-2014				
% Home owner	69.7	74.3	75.5	78.9
% Parent home owner	73.1	68.9	62.5	54.5
Wave 5: 2014-2016				
% Home owner	68.6	74.2	73.7	79.8
% Parent home owner	74.1	71.8	64.6	57.0

	Cohort Studies			
NCDS	Age 42	Age 50	Age 55	
% Home owner	81.0	83.8	77.4	
BCS	Age 42	Age 46		
% Home owner	75.3	77.6		

Table 1.3: Data to Study Links Between Wealth and Home Ownership, Descriptive Statistics

	WAS (1)	WAS (2)	NCDS (3)	WAS (4)	WAS (5)	WAS (6)
Mean net wealth (2012 prices)	£323,745	£380,285	Not available	£221,785	£157,501	£176,950
Mean value of main residence, for home owners (2012 prices)	£255,393	£275,764	Not available	£238,275	£203,808	£190,765
Saving and investment (2012 prices)	£42,069	£43,380	£11,929	£27,940	£19,651	£19,899
Sample Size	2011	2015	1991	1269	1159	898
Year	1771	1271	6774	2007	2011	2015
Age	40-44	40-44	33	33/34	33/34	33/34

	BHPS (7)	BHPS (8)	BHPS (9)
Mean value of main residence, for home owners (2012 prices)	£267,540	£229,689	£214,695
Sample Size	168	334	211
Year	2016	2011	2016
Age	42	34	34

Notes: The NCDS is a single year birth cohort matching cohort members at age 33 to parents at age 16. The WAS are multiple year birth cohorts matching individuals aged 33/34 and 40-44 (with centred age 42) to parents at age 14. The BHPS data are multiple years (2015-2017 in columns (7) and (9) and 2010-2012 in column (8)) centred around age 42 (41-43 year olds) and age 34 (32-36 year olds).

Table 1.4: Trends in Intergenerational Home Ownership Transmission

	NCDS 2000 (1)	WAS 2011 (2)	BCS 2012 (3)	WAS 2015 (4)	Change (4)-(1) (5)
A. Basic Intergenerational					
Parent home owner	0.141	0.220	0.217	0.265	0.124
	(0.009)	(0.026)	(0.014)	(0.031)	(0.032)
B. Compositional Controls					
Parent home owner	0.135	0.186	0.188	0.231	0.096
	(0.008)	(0.025)	(0.014)	(0.031)	(0.034)
Home ownership year	2000	2011	2012	2015	
Age when observed	42	42	42	40-44	
Parent home ownership year	1974	1983	1986	1987	
Sample size	8352	1771	6181	1271	

Notes: Panel (B) adds controls for age, age squared, average age of parents, the square of this, gender, the presence of a father during childhood, and the presence of a partner. All parental variables in the WAS are retrospectively asked and individuals are prompted to report values as they were at age 14. For this reason, parental age at observation is unobserved. For obvious reasons, we do not control for age in the two cohort regressions (Columns (1) and (3)). Robust standard errors are reported in parentheses.

Table 1.5A: Trends in Intergenerational Home Ownership Transmission at Older Ages, Wealth and Assets Survey

		Ages 40-44 (1)	Ages 45-49 (2)	Ages 50-54 (3)	Ages 55-59 (4)
Wave 1: 2006-2008	Parent home owner	0.205	0.132	0.156	0.117
		(0.023)	(0.022)	(0.020)	(0.018)
	Sample size	1665	1583	1569	1779
Wave 2: 2008-2010	Parent home owner	0.206	0.180	0.154	0.134
		(0.024)	(0.022)	(0.020)	(0.019)
	Sample size	1832	1795	1666	1723
Wave 3: 2010-2012	Parent home owner	0.220	0.221	0.207	0.139
		(0.025)	(0.023)	(0.021)	(0.020)
	Sample size	1771	1786	1783	1737
Wave 4: 2012-2014	Parent home owner	0.266	0.184	0.182	0.176
		(0.028)	(0.024)	(0.022)	(0.020)
	Sample size	1492	1728	1698	1651
Wave 5: 2014-2016	Parent home owner	0.265	0.187	0.223	0.167
		(0.031)	(0.026)	(0.024)	(0.021)
	Sample size	1271	1554	1697	1543

Table 1.5B: Trends in Intergenerational Home Ownership Transmission at Older Ages, Cohort Studies

		Age 42 (2000) (1)	Age 50 (2008) (2)	Age 55 (2013) (3)
National Child Development Study	Parent home owner	0.141 (0.008)	0.103 (0.009)	0.140 (0.013)
	Sample size	8352	7203	4146

		Age 42 (2012) (1)	Age 46 (2016) (2)
British Cohort Study	Parent home owner	0.217 (0.014)	0.212 (0.013)
	Sample size	6181	5537

Notes: Models contain no control variables. Parentheses include robust standard errors .

Table 1.6: Wealth and Parental Home Ownership, Wealth and Asset Survey

	WAS 2011 (1)	WAS 2015 (2)	Change (2)-(1) (3)
A. Home Owner			
Parent home owner	0.220	0.265	0.045
	(0.026)	(0.031)	(0.040)
Sample size	1771	1271	
B. Wealth Percentile			
Parent home owner	0.151	0.194	0.043
	(0.013)	(0.012)	(0.010)
Sample size	1771	1271	

Notes: Total wealth is the percentile in the total weighted wealth distribution and includes financial wealth, property wealth, and pension assets. Robust standard errors are reported in parentheses. Comparable estimates with the log of total wealth as the dependent variable are 0.813 (0.083), and 1.143 (0.105) with a statistically significant change across the waves of 0.330 (0.134).

Table 1.7: Wealth and Parental Ownership, Age 33/34

	NCDS 1991 (1)	WAS 2007 (2)	WAS 2011 (3)	WAS 2015 (4)	Change (4)-(1) (5)
A. Home Owner					
Parent home owner	0.181	0.317	0.341	0.345	0.164
	(0.009)	(0.031)	(0.033)	(0.037)	(0.038)
Sample size	6774	1269	1159	898	
B. Saving and Investment Percentile					
Parent home owner	0.125	0.152	0.168	0.166	0.041
	(0.079)	(0.016)	(0.015)	(0.016)	(0.026)
Sample size	6774	1269	1159	898	

Notes: Robust standard errors are reported in parentheses. Our measures of savings and investments exclude investment in property and refer to gross financial wealth and savings. The measure therefore includes formal investments, such as bank or building society current or saving accounts, investment vehicles such as Individual Savings Accounts, stocks and shares, and informal savings.

**Table 1.8: Intergenerational House Value Transmission,
British Household Panel Survey**

	BHPS 2016, Age 42	BHPS 2011, Age 34	BHPS 2016, Age 34
A. Home Owner			
Parental home owner	0.267 (0.118)	0.319 (0.070)	0.369 (0.076)
Sample size	168	334	211
B. House Value Rank			
Parental home owner	0.246 (0.074)	0.284 (0.042)	0.265 (0.045)
Sample size	168	334	211
C. House Value Rank			
Parental house value rank	0.415 (0.081)	0.363 (0.052)	0.390 (0.060)
Sample size	168	334	211

Notes: House value ranks come from self-reported values for the main residence. These are ranked in the BHPS sample. Robust standard errors are reported in parentheses.

Table 1.9: Estimates of π_{1t}, WAS, BHPS and EHCS

	EHCS	BHPS			WAS		
	1986	1995	2007	2009	2011	2013	2015
		(1)	(2)	(3)	(4)	(5)	(6)
House Value – Wealth rank slope (π^k_{1t})	1.084*** (0.014)	0.949*** (0.040)	0.975*** (0.012)	1.009*** (0.017)	1.033*** (0.016)	1.070*** (0.018)	1.041*** (0.018)
Age	40-44	40-44	40-44	40-44	40-44	40-44	40-44
Sample size	343	403	2,987	1,898	1,931	1,637	1,361

Notes: House value ranks come from self-reported values for the main residence. All wealth measures refer to housing wealth and the total value of savings less nonmortgage related debt. The exception if the 1986 that does not collect debt – mortgage or otherwise. Robust standard errors are reported in parentheses.

References

Aaronson, D. (2000) A Note on the Benefits of Homeownership, *Journal of Urban Economics*, 47, 356-369.

Adermon, A., M. Lindahl and D. Walderström (2018) Intergenerational Wealth Mobility and the Role of Inheritance: Evidence from Three Generations, *Economic Journal,* 128 (July), F482–F513.

Banks, J., Z. Smith and M. Wakefield (2002) The Distribution of Financial Wealth in the UK: Evidence from 2000 BHPS Data, IFS Working Paper no. 02/21, Institute for Fiscal Studies.

Black, S. and P. Devereux (2011) Recent Developments in Intergenerational Mobility, in in O. Ashenfelter and D. Card (eds.) Handbook of Labor Economics, Volume 4B, North Holland.

Black, S.E., Devereux, P.J., Lundborg, P. and Majlesi, K., 2020. Poor little rich kids? The role of nature versus nurture in wealth and other economic outcomes and behaviours. *The Review of Economic Studies*, *87*(4), pp.1683-1725.

Blanden, J. (2019) Intergenerational Income Persistence, IZA World of Labor, Institute of Labor Economics (IZA), 176.

Blanden, J., A. Goodman, P. Gregg and S. Machin (2004) Changes in Intergenerational Mobility in Britain, in M. Corak (ed.) *Generational Income Mobility*, Cambridge University Press.

Blanden, J., P. Gregg and L. Macmillan (2013) Intergenerational Persistence in Income and Social Class: The Impact of Within-Group Inequality, *Journal of Royal Statistical Society: Series A*, 176, 541-63.

Boehm, T. and A. Schlottmann (2004) The Dynamics of Race, Income, and Homeownership, *Journal of Urban Economics*, 55, 113-130.

Bourquin, P. R. Joyce and D. Sturrock (2021) Inheritances and inequality within generations, Institute for Fiscal Studies Report R173.

Charles, K. and E. Hurst (2002) The Transmission to Homeownership and the Black-White Wealth Gap*, The Review of Economics and Statistics*, 84, 281-297.

Charles, K. and E. Hurst (2003) The Correlation of Wealth across Generations*, Journal of Political Economy*, 111, 1155-1182.

Chetty, R., N. Hendren, P. Kline and E. Saez (2014) Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States, *Quarterly Journal of Economics*, 129, 1553-1623.

Clark, G. and N. Cummins (2015) Intergenerational Wealth Mobility in England, 1858-2012: Surnames and Social Mobility, *The Economic Journal,*125 (February), 61–85.

Clarke, S., A. Corlett and L. Judge (2016) The Housing Headwind: The Impact of Rising Housing Costs on UK Living Standards, Resolution Foundation Report, Resolution Foundation.

Corlett, A. and F. Odamtten (2021) Hope to buy: The decline of youth home ownership. Briefing. Resolution Foundation.

Costa, R. and S. Machin (2017) Real Wages and Living Standards in the UK, Centre for Economic Performance Election Analysis No. 36.

Cribb, J., A. Hood and R. Joyce (2016) The Economic Circumstances of Different Generations: The Latest Picture, IFS Briefing Note BN187, Institute for Fiscal Studies.

Crawford, R., D. Innes and C. O'Dea (2016) Household Wealth in Great Britain: Distribution, Composition and Changes 2006–12, *Fiscal Studies*, 37, 35-54.

Crossley, T. and C. O'Dea (2010) The Wealth and Saving of UK Families on the Eve of the Crisis, Institute for Fiscal Studies.

Dearden, L., S. Machin and H. Reed (1997) Intergenerational Mobility in Britain, *Economic Journal*, 107, 47-66.

Dietz, R., and D. Haurin, D. (2003). The social and private micro-level consequences of homeownership. *Journal of Urban Economics*, 54, 401-450

.

Department for Communities and Local Government (2017) English Housing Survey: First Time Buyers, 2015-2016.

Erikson, R. and J. Goldthorpe (2010) Has Social Mobility in Britain Decreased? Reconciling Divergent Findings on Income and Class Mobility, *British Journal of Sociology*, 61, 211-230.

Fagereng, A., L. Guiso, D. Malacrino and L. Pistaferri (2020) Heterogeneity and Persistence in Returns to Wealth, *Econometrica*, 88, 115–170.

Fagereng, A., Mogstad, M. and Rønning, M. (2021). Why do wealthy parents have wealthy children?. *Journal of Political Economy*, *129*(3), pp.703-756.

Goodman, L., and C. Mayer. 2018. Homeownership and the American Dream. *Journal of Economic Perspectives*, 32, 31-58.

Gregg, P., L. Macmillan and C. Vittori (2017) Moving Towards Estimating Sons' Lifetime Intergenerational Economic Mobility in the UK, *Oxford Bulletin of Economics and Statistics*, 79, 79-100.

Gregg, P., & Kanabar, R. (2021). Intergenerational wealth transmission in Great Britain (CEPEO Working Paper No. 21-06). Centre for Education Policy and Equalising Opportunities, UCL.

Griffith, M. (2011) Hoarding of Housing: the Intergenerational Crisis in the Housing Market, London: The Intergenerational Foundation.

Haider, S. and G. Solon (2006) Life-Cycle Variation in the Association between Current and Lifetime Earnings, *American Economic Review*, 96, 130-1320.

Jantti, M., E. Sierminska and T. Smeeding (2008) The Joint Distribution of Household Income and Wealth: Evidence from the Luxembourg Wealth Study", OECD Social, Employment and Migration Working Papers, No. 65, OECD Publishing, Paris.

Jenkins, S. and A. Maynard (1983) Intergenerational Continuities in Housing, *Urban Studies*, 20, 431-438.

Karagiannaki, E. (2017) The effect of parental wealth on children's outcomes in early adulthood The Journal of Economic Inequality, 15, 217–243.

Keister, L. and S. Moller (2000) Wealth Inequality in the United States, *Annual Review of Sociology*, 26, 63-81.

Lindley, J. and S. McIntosh (2019) The Social Mobility of Homeownership: To What Extent have the Millennials Fared Worse?, Sheffield Economic Research Paper Series Number 2019012.

Mulligan, C. (1997) Parental Priorities and Economic Inequality, University of Chicago Press.

Pfeffer, F. (2018) Growing Wealth Gaps in Education, *Demography*, 55, 1033–1068.

Pfeffer, F. and A. Killewald (2018) Generations of Advantage. Multigenerational Correlations in Family Wealth, *Social Forces,* 96(4) 1411–1442,

Piketty, T. (2000) Theories of Persistent Inequality and Intergenerational Mobility in A. Atkinson and F. Bourguignon (eds.) *Handbook of Income Distribution*, Volume 1, North Holland.

Piketty, T. (2014) Capital in the Twenty-First Century, Belknap Press of Harvard University Press.

Piketty, T., and G. Zucman (2014) Capital is Back: Wealth-Income Ratios in Rich Countries, 1700-2010, *Quarterly Journal of Economics*, 129, 1155-1210.

Roine, J. and D. Waldenström (2015) Long Run Trends in the Distribution of Income and Wealth, in A. Atkinson and F. Bourguignon (eds.) *Handbook of Income Distribution*, Volume 2, North Holland.

Rohenkohl, B. (2020) Essays in Social Mobility PhD Thesis, University of Sheffield.

Solon, G. (1999) Intergenerational Mobility in the Labor Market in O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics*, Volume 3, North Holland.

Nybom, M. and J. Stuhler (2017) Biases in Standard Measures of Income Dependence, Journal of Human Resources 50(3): 800-825.

Wagner, S,. D. Boertien and M. Gortz (2020) The Wealth of Parents: Trends Over Tie in Associative Mating Based on Parental Wealth, IZA Discussion Paper No 12999.

Wolff, E. (2016) Household Wealth Trends in the United States, 1962 to 2013: What Happened over the Great Recession?, *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 2, 24–43.

Wood, J. and S. Clarke (2018) House of the rising son (or daughter): The Impact of Parental Wealth on their Children's Homeownership, Resolution Foundation Report, Resolution Foundation.

Zavisca, J. and Gerber, T. (2016). The socioeconomic, demographic, and political effects of housing in comparative perspective. Annual Review of Sociology, 42, 347-367.

Appendix

Table A1.1: Sample Selection in the NCDS and BCS

	Number of Observations	
	National Child Development Study	British Cohort Study
In the first sweep	18,558	17,196
In at age 11/10	10,934	14,875
In at age 16	11,661	11,615
With housing information at age 16	11,624	9,378
In at age 33/34	8,472	9,665
With housing information at 33/34	7,714	9,602
With housing information at 16 and 33/34	7,687	6,392
In at age 42	8,433	9,841
With housing information at 42	8,375	9,754
With housing information at 16 and 42	8,352	6,267

Chapter 2: Inside the Black Box of Mobility: Understanding Upward Mobility in Two British Cohorts.

Abstract

I use tools from the machine learning explainability literature, combined with a latent factor framework, to highlight how patterns of upward mobility have changed across two British birth cohorts. The methodology allows for arbitrary patterns of non-linearity, guards against overfitting, is invariant to monotonic transformations of the factors studied, and, unlike previous studies, is not sensitive to the order in which predictors enter the model. I find that origin and gender predict mobility. Those born in the bottom quintile are most likely to be upwardly mobile due to having more 'room to move up'. Females are much less likely than their male counterparts to be upwardly mobile. Years of schooling and cognitive ability are also important. These act as sufficient statistics for a host of early childhood factors; namely, household composition, socioemotional skills, parental time investments, and school quality. Despite the relationship between parental and child income strengthening between the birth cohorts, the determinants of upward mobility are stable.

2.1 Introduction

In this paper, I ask whether patterns of social mobility have changed over time. Making inferences about trends in mobility is difficult due to a multiplicity of measures that can be used.[21] In the UK, debates about whether social mobility has changed over time centre around the metric used to measure mobility. Measures of occupational change show stasis (Goldthorpe and Bukodi, 2018) whereas income-based measures, such as rank correlations between parental income and that of their offspring, suggest a decline (Blanden et al, 2005).

Even if the aforementioned statistics are stable, it may still be the case that the process linking later life outcomes to one's origin is changing. For instance, the drivers of upward mobility - be they educational attainment, household demographics, or early childhood investments - may vary in importance across cohorts. In this paper, I use tools from the machine learning explainability literature to assess whether drivers of mobility have changed over time.

To do so, I use data from two British cohort studies - the 1958 National Child Development Study (NCDS) and the 1970 British Cohort Study (BCS). Each contains a wealth of measurements on conditions from birth until adulthood. In accordance with the literature on human capital formation, I combine these measurements using a latent factor framework to produce measures of cognitive skills, non-cognitive skills, and parental investments at various stages of development. I supplement these with variables capturing family structure and completed years of schooling. These 'features' are then used as inputs in forest-based predictive models that aim to predict upward mobility.[22] I use tools from coalitional game theory to

---

[21] Simply by varying how, and for whom, income mobility is measured, Engzell and Mood (2023) estimate 82,944 specifications relating parental and child income using Swedish registry data.
[22] In following the ML literature, I will refer to model inputs as features throughout.

measure how each 'feature' contributes to the predictions made by the models. Examining how features contribute differentially to mobility, across the two cohorts, is the contribution of this paper.

This paper adds to the literature that aims to understand the mechanisms linking parental occupation or income to the occupation and income of their offspring[23]. It also adds to the growing literature that uses machine learning methods in applied economics research (Mullainathan and Spiess, 2017; Athey, 2019).

The papers closest in spirit to this one are Blanden, Gregg, and Macmillan (2007) who look at the factors that mediate the intergenerational earnings elasticity (IGE) for NCDS and BCS cohort members. Similarly, Bolt et al (2021) combine multilevel mediation analysis with a latent factor framework to decompose the fraction of the IGE explained by a number of early life factors. Methodologically, Blundell and Risa (2018) make use of machine learning methods and argue that the outputs from machine learning models can be used as measures of mobility.[24]

This paper differs from these in both conceptual terms - I aim to assess the relative importance of different contributors to upward mobility rather than explain which factors mediate the IGE - and methodologically.[25] On the latter, I argue that machine learning (ML) methods are particularly appropriate for the problem at hand. Forest-based methods allow for inputs to interact with each other in a manner that linear regression does not. A wealth of

---

[23] Blanden, Doepke, and Stuhler (2022) provide a detailed overview of the link between educational inequality and social mobility. Heckman and Mosso (2014) survey the links between early life skills and parental investments and mobility.
[24] Blundell and Risa (2018) employ many of the same models that I use to predict income for BCS and NCDS cohort members based upon parental income and family characteristics. Their focus is on whether the fraction of variance explained by the models differs across the cohorts.
[25] The final section does include a mediation analysis to help interpretation of the ML results.

evidence highlights the importance of interactions between stocks of skills and investments at different ages (Cunha and Heckman, 2008; Cunha et al, 2010).

Understanding feature importance necessitates a predictive model that allows for non-linearities such as these. The methods used rely on cross validation as a means of model selection. While not a panacea for the small-sample sizes that are inherent with cohort studies, this form of model selection is less prone to overfitting than methods which simply use linear methods on all the available data.

Despite these advantages, machine-learning methods are often opaque and difficult to interpret. Models that provide the best fit are often 'ensemble' methods that fit multiple models and aggregate the predictions (Hastie, Tibshirani, and Friedman, 2009). As such, these methods have no counterpart to a regression coefficient by which the effect of a unit change of some variable can be mapped to a change in the outcome.[26] For this paper - where the primary aim is to understand how the importance of different features changes over time - this presents a problem. To overcome this, I use Shapley additive explanations (SHAP) as introduced in Lundberg and Lee (2017). SHAP is inspired by the use of Shapley values in game theory that measure the contribution of a player to a coalition as the average of all the marginal contributions a player makes to all possible coalitions of players. By replacing 'players' with ML model features, relative contributions of different features to model predictions can be compared in an intuitive manner.

[26] As will be explained in the methodology section, the SHAP approach does not suffer from the problem that different scales of the variables can lead to different conclusions regarding the relative importance of the variable.

Despite patterns of mobility being different across the two cohorts, the determinants of upward mobility are shown to be surprisingly stable. Being born in the bottom 20% of the income distribution makes upward mobility ‘easier’. Intuitively, those at the bottom have more space to move up while those born into the middle of the distribution can only achieve upward mobility by becoming a top 20% earner. For the 1958 and 1970 birth cohorts, the wage premium given to males makes upward mobility less likely for female cohort members.

Cognition and years of schooling are strong predictors of upward mobility in both cohorts. Once one knows a cohort members’ cognitive abilities at adolescence, information on family background has little predictive power over upward mobility outcomes. Multi-step mediation analysis highlights that school quality, parental time investments, and family background variables operate indirectly by increasing early stage cognition and completed years of schooling.

The paper proceeds as follows. Section 1 describes the data used. Section 2 discusses the measurement framework linking the numerous measurements taken in the cohort studies to the latent constructs underlying them. Section 3 gives a brief overview of the ML algorithms used for prediction and explains how SHAP enables feature importance to be assessed. Section 4 presents the results of the analysis. Section 5 concludes.

2.2 Data

I use data from two British cohort studies - the 1958 National Child Development Study (NCDS) and the 1970 British Cohort Study (BCS). These two studies are ideal for the purposes of this study and have been used to examine intergenerational mobility in income (Dearden et al. 1997; Blanden et al. 2004; Gregg et al. 2017), social class (Erikson and Goldthorpe 2010), and home ownership (Blanden et al 2023). They track two birth cohorts through their childhood, adolescence, and their working life. Importantly, the studies are structured similarly, and a number of variables are harmonised across the studies.

For the purposes of this study, I use parental income observations at age 16 and cohort members' income at age 42. Upward mobility is taken to mean residing in a higher income bracket - measured by income quintiles - than one's parents. Someone born into the bottom 20% of the income distribution but whose earnings place them above the 20th percentile is upwardly mobile. By construction, those born to parents in the top 20% of the distribution cannot be upwardly mobile. These individuals, unless noted, are removed from what follows.

Tables 2.1, 2.2, and 2.3 present some well-known facts about income mobility from the cohort studies. A linear regression of log parental income on income at age 42 highlights that between 2000 and 2012 (when age 42 income is measured), the intergenerational persistence of income went from 0.274 to 0.354. This strengthening of the relationship between parental income and one's own income holds when income earlier in the life cycle (age 33/34) is considered. In principle, this could represent a widening of the income distribution as well as changes in the dependency between origin and destination income. The bottom row of Table 2.1 follows Chetty et al (2014) and estimates the rank slope between parental and child income. A rise in rank-slopes is evident irrespective of whether age 33/34 or age 42 income is used as an outcome. One's position in the income distribution relates more strongly to parental income in the BCS than the NCDS.

Tables 2.2 and 2.3 look at transition probabilities between income quintiles where the origin is the quintile one was born into and the destination is income quintile at age 42. There are several changes between the cohorts. Income persistence at the top strengthens between the NCDS and BCS. Those born into the top 20% of household incomes have a 30% likelihood of remaining there in 2000 and a 36.8% chance in 2012. When looking at the likelihood of upward mobility, this translates into a fall in mobility prospects for those born into the second highest income quintile. These individuals must move into the top quintile to be upwardly mobile, but only 21.79% do so in the BCS as compared to 24% in the NCDS.

In general, even where upward mobility is unchanged, the form it takes differs across cohorts. 54% of those born into the 2nd income quintile achieve upward mobility in both cohorts, but those born in the BCS are more likely to do so by moving to a nearby income quintile i.e. their prospects of reaching the top 20% have declined over time (going from 14.42 to 11.81). Those born into the lowest income quintile have declining upward mobility prospects between the cohorts and, even for those that are upwardly mobile, the prospects of reaching the top two quintiles falls from 30.71 to 26.92.

In total, persistence in income increases at the top and, for those that are upwardly mobile, they are more likely to move to neighbouring income quintiles than make extreme moves. These patterns mean that the IGE and rank-rank slope in income increase while upward mobility remains stable.

In this paper, my focus is on the drivers of upward mobility. While patterns of mobility are well understood, the determinants of upward mobility are less well studied. Several studies have highlighted factors that may shape mobility outcomes. Kearney (2023) shows how family composition - particularly the presence of two married parents - advantages some children over others. Chetty et al (2014) discusses how family structure can partially explain cross sectional variation in mobility rates across US commuting zones. Heckman and Mosso (2014) discuss how family background and parental investments combine to produce stocks of cognitive and socio-emotional skills that influence schooling choices and outcomes.

As such, I collect measurements in the cohorts on family composition, cognition, parental investments, and socioemotional skills. For cognition, I use the plethora of tests discussed in Moulton et al (2020). These tests are administered at ages 7, 11, and 16 in the NCDS and ages 5, 10, and 16 in the BCS. While the tests are not fully comparable, they do aim to capture the same underlying constructs such as mathematical ability, verbal reasoning, and vocabulary. For socio-emotional skills, I follow Attanasio et al (2020) and use identical Rutter scale items that are reported by cohort member's mothers. For investments, I consider variables that capture parental time investment (whether one reads to the child, number of weekly outings, interest taken in child's education) and school level investments (pupil/teacher ratio, the presence of a parent teacher association in the school). These are chosen to match investment variables used elsewhere in the literature on human capital formation (see Agostinelli and Wiswell, 2023 for a recent example of this kind of work). In each case, the measures are merely proxies of some underlying construct. In the next section, I discuss how I combine these using a latent factor framework to produce interpretable indices.[27][28]

For family background, I choose a set of variables that capture conditions at birth and throughout childhood. Tables 2.4 and 2.5 show averages across three distinct groups - those who are upwardly mobile, those who are not, and those born into the top 20% of the income

[27] Appendix tables A2.1-A2.14 give a full breakdown of the confirmatory factor analysis underlying the indices.
[28] The structure of the latent factor framework is also useful when imputing missing data. For many observations, full data are not available but measures at earlier and/or later ages are alongside information on family background. I use Multiple Imputation by Chained Equations (MICE) to impute missing data for the sample of individuals who have age 42 income data and parental income at age 16 data.

distribution. The latter serve as a reference group as, by definition, they cannot achieve upward mobility. Alongside these, residualised differences between those who do and do not achieve upward mobility are presented. The differences condition out the income quintile group means due to the mechanical effect of one's origin quintile on the likelihood of upward mobility.

While most of the differences are significantly different, a few stand out numerically. Firstly, most of the variables that capture household demographics are small in the NCDS sample. Having a teen mother, an absent father, and parents who are not married at birth have limited effect on upward mobility. This is not entirely surprising given the limited variability in these variables in the sample. Few cohort members are born to teen mothers and even fewer are born into single parent families. Family structure appears uniform across the income distribution. Having a parent attend university is more prevalent amongst the upwardly mobile as is home ownership. The majority of those in the top 20% have parents who own a home and almost a quarter have a parent with a degree. Amongst the bottom 80%, the upwardly mobile, while not matching those in the top 20%, are edging towards them in terms of home ownership rates and parental education. Upwardly mobile individuals, conditional on origin quintile, are 3.3 percentage points more likely to have a university educated parent and are 8.2 percentage points more likely to grow up in owner occupied housing.

Numerically, the results from the BCS are somewhat similar. The differences between the upwardly mobile and not upwardly mobile in terms of parental education dwarfs differences in other family background characteristics such as parental divorce and the absence of a father figure. This is in spite of the fact that parental divorce increases sharply between the cohorts. In the NCDS, 8.7% of parents divorce and this rises to 17.1% in the 1970 cohort. Many more cohort members are born into owner occupied housing in the latter cohort but the difference between those who are upwardly mobile and those who are not remains similar in magnitude (the residualised difference in the BCS is 6.5 percentage points). One of the largest changes is in parental education. The proportion of parents with degrees doubles across the cohorts across the income distribution. 54% of those born into the top quintile have a parent with a degree in the BCS (as compared with 23% in the NCDS), while 16% of those in the bottom 4 quintiles do (as compared with 6% in the NCDS). The gap in parental education between the mobile and the not widens between the cohorts. Upwardly mobile individuals are 8.8 percentage points more likely to have a degree educated parent than those who are not upwardly mobile. The corresponding difference in the NCDS is 3.3 percentage points.

In both cohorts, gender plays a large role in mobility prospects. Males are more likely to be upwardly mobile than their female counterparts reflecting a sizable gender pay gap in the cohort data. While all the differences in Tables 2.4 and 2.5 are 'origin' variables, I do include one's own education as well. In each cohort, upwardly mobile sample members are much more likely to be degree educated than their non-mobile peers. The difference in the fraction of individuals having degrees doubles between the cohorts. In the NCDS, upwardly mobile individuals are 18.2 percentage points more likely to attend university than non-mobile individuals. This gap rises to 32 percentage points in the BCS.

2.3 Latent variable measurement

While family background can be aptly summarised as above, many of the predictors that I consider are less well defined. The cohort studies collect many different measurements that proxy underlying latent variables of interest. I follow the approach taken by Heckman et al (2013) and reduce the plethora of proxy variables down into low dimensional indices.[29]

$$1) M_{\omega,i,t,j} = \mu_{\omega,t,j} + \lambda_{\omega,t,j}\omega_{i,t} + \varepsilon_{\omega,i,t,j}$$

Here M denotes measure, ω denotes the underlying construct (time investment, cognition, socio-emotional skill, and school quality), t denotes age, and j indexes the measure. Error terms are assumed to be uncorrelated across individuals, measures, and age.

As the latent variables have no natural scale, normalizations are needed to identify the model. For the continuous measures that I use in the cognition equation, I set the constant to zero and the intercept to 1 for one of the test score loadings.[30] For the equations governing investments, socio-emotional skills, and school quality, the variables have an ordinal scale. Here I take the approach discussed in Muthén (1984) that uses polychoric correlation matrices to derive factor loadings. In doing so, I normalise the model by fixing the variance of the latent factors to 1.

Exploratory factor analysis suggests that for the NCDS there is a single cognitive construct at each age (7, 11, and 16), a single socio-emotional construct at age 7, two socio-emotional constructs at ages 11 and 16 (that I label internalising and externalising skills), a parental time investment factor at all ages, and a school quality factor at age 7. The BCS yields a

[29] Appendix Tables A2.1-A2.14 give the results of the exploratory factor analysis. I first conduct an EFA to determine the underlying structure of the data i.e. how many constructs are needed to describe covariation in the data. I then estimate a series of dedicated factor models where each measure loads onto a single factor.
[30] The cognition equation uses standardised test scores as the measures.

similar structure with a cognitive factor, two socio-emotional factors, and a parental time investment factor at each age (5, 10, and 16), alongside a school quality factor at age 10.[31]

Factor scores, which later serve as inputs to the ML algorithm described in the next section, are plotted in Figures 2.1 through 2.5. As the scores alone are difficult to interpret, I plot the distribution separately for upwardly mobile, not upwardly mobile, and those born into the top quintile. I also report rank differences between both groups - those are mobile and those who are not - and those born into the top 20%.

Figures 2.1 and 2.2 show the distribution of scores for the socio-emotional skill measures. Several studies highlight the role that non cognitive skills play in shaping outcomes and the increasing importance of non-cognitive skills in determining wages (Heckman et al, 2006; Deming, 2017). Interestingly, there appears to be less of a role played by these skills in enhancing mobility prospects. The distributions of both externalising and internalising skills are similar for those who are and who are not upwardly mobile in both cohorts. Both groups have lower scores, on average, than those born in the top quintile. In line with previous studies that document an increase in inequality in socio-emotional skills (Attanasio et al, 2020), the gap between those born outside of the top 20% of the income distribution and the rest widens between the NCDS and the BCS. This is particularly the case for externalising skills where the difference in mean rank between those who are not upwardly mobile and the top 20% goes from 3.94 to 9.78. The differences between the top 20% and those who achieve mobility exhibits a

[31] The existence of a single school quality factor alongside multiple parental time investment factors (on for each age) is due to data collection. Only at ages 7 and 10 are detailed questions given to teachers and heads about the school. Conversely, parents are asked multiple questions about their engagement with their children in all childhood/adolescent waves.

similar rise from 2.30 to 10.20. Internalising skills exhibit a somewhat weaker relationship with socioeconomic background.

A different picture emerges when cognition at age 16 is looked at (Figures 2.3 and 2.4). For both cohorts, the score distribution of the upwardly mobile is shifted away from those who remain in the quintile into which they were born. For the NCDS, the cognitive score distribution of mobile individuals looks more like the distribution amongst those born into the top quintile. A regression of ranked factor scores on group dummies (with top quintile as the omitted category) gives coefficients of -18.02 and -4.60 for those who are not and who are upwardly mobile in the NCDS cohort. For the BCS, the same coefficients are -23.37 and -11.91. While the top quintile pulls further away from the bottom 80% in the BCS, inequality in cognitive achievement between those who move up the income distribution and those who do not remains stable across the cohort.

Finally, I plot scores for two well-known inputs into cognition - school quality and time investments. For both these investments, and across both cohorts, there is a clear divide between the bottom 80% of the parental income distribution and the top 20%. Those in the top quintile have a score distribution that lies clearly to the right of the rest. Looking within the bottom 80% and splitting by mobility status shows little difference between those who are upwardly mobile and those who are not. For parental time investments, the mean rank of scores is 9.02 lower for those who are not mobile and 7.98 lower for those who are when compared with those born into the top quintile in the NCDS. These gaps widened further to 17.31 and 17.96 in the BCS cohort. With a notable exception, a similar pattern follows with the school quality factor scores. Here the gap between those born in the top quintile and those who are not/are upwardly mobile respectively are -15.44 and -11.34 in the NCDS. These gaps widened in the BCS cohort to -

22.97 and -22.41. In the NCDS cohort, the rank difference is significant here at 4.10. The score distribution is bimodal for all three categories with a greater density around the second (higher school quality) maxima for those who are upwardly mobile.

Figures 2.6 and 2.7 consolidate the evidence above as follows. I run a logistic regression of an upward mobility dummy on each individuals' cognitive score, their socio-emotional scores, their school quality score, and their parental investment score. I use the fitted model to vary each factor from the 5th to 95th percentile holding other factors at their mean values. The resultant curves trace out how the predicted probability of upward mobility changes with each factor. The curves highlight that raising cognition has an outsized impact on mobility relative to socioemotional skills. Varying cognition from the 5th to 95th percentile raises mobility prospects from 7% to 82% in the NCDS and 8% to 68% in the BCS. School quality raises mobility prospects more than parental investments of time, but again the change is more pronounced in the NCDS than the BCS. In the NCDS the impact is a 20 percentage points rise (going from just under 30% to 50%) while the BCS rise is 12 percentage points (36% to 48%).

Taken together, the evidence shows that on several determinants of wages - cognition and externalising skills - those born into the top quintile of household income have pulled away from their counterparts between the two cohorts. Similar holds true when one considers factors that serve as inputs into cognition and socio-emotional skills - school quality and parental time investments.[32] Amongst those that are born into the bottom 80% of the income distribution, individuals who are and who are not upwardly mobile are more comparable to each other than

[32] Later I will formally link these inputs in a production function framework.

they are to those in the top quintile. Cognition, and to a lesser extent school quality, are exceptions to this suggesting that these variables are important determinants of upward mobility.

2.4 Methodology

2.4.1 XGBoost

The descriptive evidence above suggests that certain factors play a more crucial role in driving mobility than others; however, it is desirable to combine them in a holistic way when predicting the likelihood of upward mobility. To do so, I use an algorithm called extreme gradient boosting (XGBoost).

There are several reasons why this method is preferred to, for instance, running a linear regression. Firstly, the fitted model allows for arbitrary patterns of nonlinearity. Features such as socioemotional skills are allowed to interact with other features to predict outcomes. Secondly, although not specific to this method, the algorithm has a number of 'tuning' parameters that can be chosen to prevent the model from overfitting. This is particularly important given the data that is used to estimate intergenerational mobility. Cohort data typically has few observations relative to the number of features included. Not explicitly cross validating the models can lead to overfitting.

For this paper, I aim to explain what drives upward mobility. Any kind of model explainability will ultimately explain *model fit*. Understanding models that are less sensitive to outliers and generalise well to other samples is of more interest than understanding models that perform poorly when new data is introduced. Finally, unlike parametric models, the fit of XGBoost is invariant to monotonic transformations of the features used. This is useful given that many of the features are latent constructs that have no natural scale.

The algorithm fits a series of decision trees before averaging their predictions. Each fitted tree is a weak learner that performs marginally better at classifying the observation than a guess. Intuitively, weak learners have low variance in that they perform similarly on different datasets. Each tree is fit sequentially on the errors of the previous tree so that each additional tree is able to predict where the previous one performed poorly. Formally:

$$Y = F_0(x) + e_0$$

$$\hat{Y} = \hat{F}_0(x)$$

$$e_0 = F_1(x) + e_1$$

$$\hat{e}_0 = \hat{F}_1(x)$$

Here $F_n(x)$ is a prediction from the nth tree based model.[33] A final prediction is then made by summing the initial prediction along with the predicted errors:

$$2) \quad \text{Prediction} = \hat{Y} + \hat{e}_0 + \hat{e}_1 + \ldots$$

In practice there are a number of decisions to be made about how the underlying trees are built. I vary the maximum depth of the underlying trees i.e. how deep the $F(x)$ predictions can go, the number of boosting rounds (the size of n), and the learning rate. The latter governs the weight that each additional tree adds to the final prediction. To select these parameters, I

[33] Trees splits are selected so as to minimise the log loss criterion in each node. Log loss penalises predictions that are confidently wrong. In this case, instances where upward mobility has a high probability, but observations remain in the class into which they were born.

perform 5-fold cross validation and a grid search over the parameter space before selecting the parameters that provide the highest predictive accuracy.

2.4.2 Shapley Additive Explanations (SHAP)

While the above algorithm provides a convenient way to predict mobility, model predictions are often opaque. Tree based models, and models that average over tree-based predictions, give no counterpart to a regression coefficient that explains how a unit change in a feature changes the probability of the outcome studied.

To understand the output of the model I use a tool from coalitional game theory - Shapley values - as suggested in Lundberg and Lee (2017). Each feature used in the model is assigned a value that is that features marginal contribution across all possible combinations of features used in the model. Formally,

$$3) \quad \phi_i(f) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!}[f(S \cup \{i\}) - f(S)]$$

Here, the contribution of feature i depends upon the difference in prediction between a model with and without that feature $f(S \cup \{i\}) - f(S))$ for a given subset of features - S - that is weighted and summed across all feature combinations.

Shapley values have several properties that make them well suited to explaining model predictions. Firstly, for a categorisation problem such as predicting upward mobility, the log odds of the model's predicted probability can be expressed as a sum of the baseline prediction (the sample average probability) and the Shapley value for each included feature. The sum of Shapley values for the features thus account for the deviation of the model's prediction from the sample average of the outcome. In other words, the values show how to distribute the differences between individual predictions and the sample average among all features within the model.

Secondly, unlike previous analysis that use regression-based methods to assess feature importance in similar contexts (Blanden et al 2007), SHAP is, by construction, invariant to the order in which variables enter the model. SHAP is also invariant to any variable transformations that leave the model fit unchanged. The underlying model that I explain is invariant to monotonic transformations of the model features and computed Shapley values will inherit this characteristic.

2.5 Results

2.5.1 Model fit

Table 2.6 presents several fit measures for the model. As a point of comparison, fit measures are also presented for a naive model - one that predicts each observation will fall in the most common mobility class (no mobility in both samples) - and a linear probability model (LPM) that predicts upward mobility for those with fitted probabilities above 0.5. The fit metrics reported are the fraction of correct predictions (accuracy), the ratio of correct positive predictions to total positive predictions (precision), and a measure of the model's ability to distinguish between positive and negative outcomes (AUC).

The results highlight that the features considered have a significant predictive power for upward mobility. Both the LPM and XGBoost perform well in terms of overall accuracy. Interestingly, there is little difference in predictive accuracy between a linear model and the more flexible boosting model in terms when predicting the main upward mobility measure used in this paper - whether one resides in a higher income quintile than the quintile one was born in. This suggests a limited role for interactions between the features considered.[34]

In addition to the simple mobility definition, I also consider a second definition model that predicts the likelihood of moving from either of the bottom two quintiles to the top two. This ‘rags to riches’ style of mobility is rare in both the cohorts. 8%/9% of those in the bottom quintiles reach the top as compared with 47%/45% who are upwardly mobile according to the first definition. In this case, there is a large performance discrepancy between the tuned ML model and the LPM. While the LPM is accurate in this case, its accuracy is driven by the naive prediction that no one will achieve upward mobility. In absence of parameter tuning, the ML model also predicts no upward mobility for either the NCDS or the BCS. The tuned model on the other hand does predict upward mobility for some observations and does so correctly in 59% (NCDS) and 86% (BCS).

2.5.2 Feature Importance

The tuned ML model does well at correctly classifying individuals as mobile vs not irrespective of the definition of mobility. In this section, I explain the model predictions using Shapley values.

Figures 2.8 and 2.9 plot the mean absolute SHAP value of each feature over all observations in the dataset. The features are ordered by their mean absolute value and features

[34] The result above is likely to be driven by data processing done prior to fitting the models. Reducing the large amount of cognitive, socio-emotional, and investment proxies into indices guards against overfitting at the outset and minimises the role played by hyperparameter tuning. This, coupled with weak interaction effects, drives the results.

with limited impact on model predictions are pooled at the bottom. The results are presented on a log odds scale.

A consistent picture emerges across both the cohorts. Despite evidence of different transition probabilities between the 1958 and 1970 birth cohorts, the determinants of upward mobility are stable. In each case, one's gender and origin are of crucial importance. This is due to a sizable wage penalty for females in the cohorts and greater ease of upward mobility for those in the bottom quintiles. Upward mobility for those born in the 'middle' means breaking into the top income quintile in adulthood whereas mobility can more easily be achieved by those who can, for instance, move from the bottom quintile into the middle of the income distribution.

Aside from these effects, education, and cognition (particularly at ages 10/11 and 16) play a large role. In the NCDS the effect of cognition variables is 0.39 while the effect of schooling variables is 0.19[35]. For the BCS, the total effect of cognition related variables is 0.22 whereas for education variables it is 0.43. Completed education has higher predictive power in the latter cohort than the former where cognition has more predictive power, but taken together gender, origin, cognition, and schooling are the main drivers of mobility in both cohorts.

Figures 2.10 and 2.11 give a more complete picture of how each feature contributes to model fit. The previous Figures plot absolute values whereas 10 and 11 colour the values according to whether they are positive or negative. As the model allows for arbitrary interactions it may be the case that high levels of a feature lead to high predicted probabilities for some individuals and lower probabilities for others. However, looking at the plots the features that are of most importance are uniform in sign. Higher levels of cognition have uniformly positive

[35] These are obtained by summing the SHAP values on the education dummies used in the model.

SHAP values for both the NCDS and BCS whereas being female pushes down the likelihood of upward mobility. Similarly, being born into lower income quintiles makes it easier to achieve mobility (at least according to the definition in this paper) as does going on to have more years of schooling. For the latter, educational levels of NVQ level 3 and above boost mobility prospects whereas NVQ2 and below decrease them. This result is consistent across both cohorts.

While education and cognition play a large role in upward mobility, they are somewhat secondary to the mechanical role played by origin (and to a lesser extent gender). Figures 2.12 and 2.13 assess whether the role of cognition varies by gender and origin. In each case, Shapley values for cognition are plotted separately for those born into the bottom 20% and for females. As the Figures show, the effect of cognition is uniform across quintiles and genders. Higher measured cognition is associated with higher Shapley irrespective of gender and income origin and the relationship between the level of cognition and the extent to which the model predicts upward mobility is constant across origin/gender.

2.5.3 Interpretation

The results above suggest that cognition and schooling play a primary role in shaping upward mobility. Several factors - particularly family background measures, school quality, and time investments - have minimal predictive value. There are a few potential explanations for this; firstly, these factors could have no explanatory power; or secondly, their effects can be fully captured by schooling and cognition i.e. educational outcomes act as a sufficient statistic for the effect of family composition and investments.

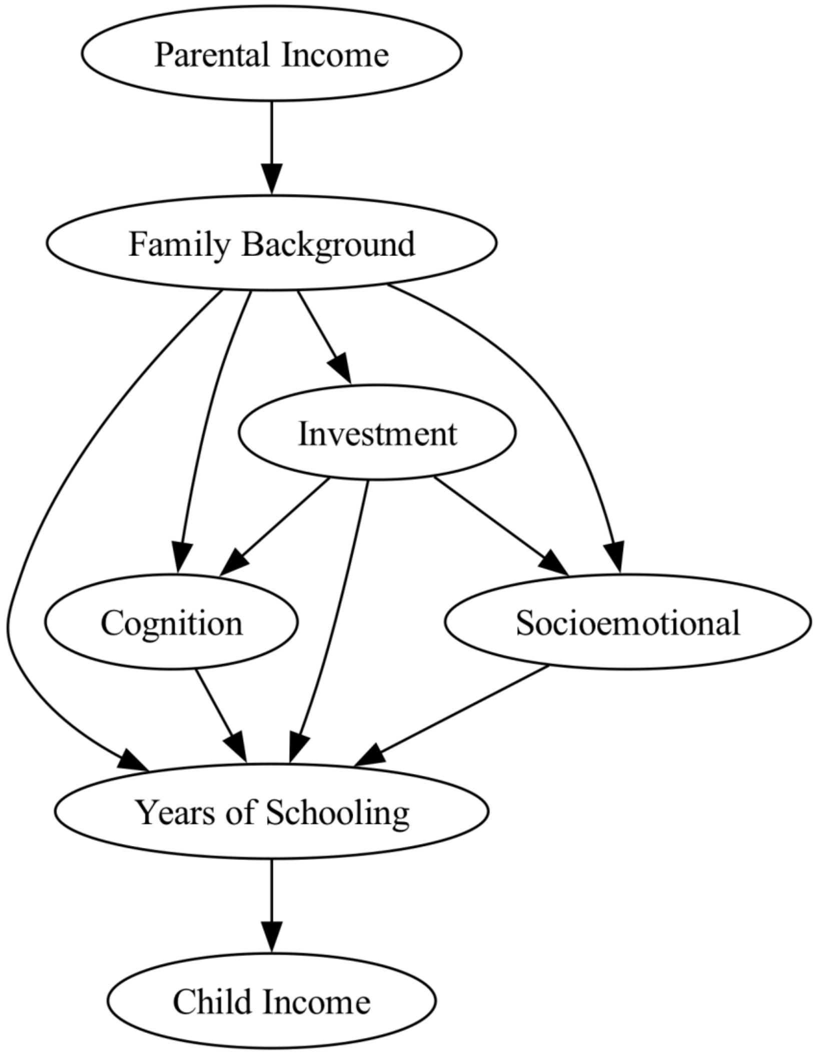
To better understand how investment and background contribute to mobility outcomes, I follow Bolt et al (2021) and run a multi-level mediation analysis to understand the relationship between parental income and one's own income for both birth cohorts. In the first stage, I allow only for direct mediation. In the second, I allow inputs such as time investment and school quality, to mediate the IGE both directly and indirectly. Formally, in the first stage I estimate:

$$4) \quad lnY_i = \alpha_S S_i + \alpha_C C_i + \alpha_{SE} SE_i + \alpha_I I_i + \alpha_F F_i + \alpha_{Y_P} \ln Y_{Parent,i} + u_i^Y$$

Where the own income is regressed onto years of schooling, the cognitive measures, socio-emotional measures, investment measures, family background characteristics, and parental income.[36] For the simple mediation, I then relate each measure to parental income. If ρ denotes the IGE and κ_j denotes the coefficient from a regression of measure j on log parental income, then the share of the IGE explained by measure j can be written as $\alpha_j \kappa_j / \rho$.

A more complex model allows for measures to have indirect as well as direct effects. Here, I allow for a richer set of dynamics as illustrated in the path diagram below:

[36] In practice, the multilevel mediation is made simpler by only including final levels of cognition and socio-emotional skills. As I aggregate these across ages, and because early skills operate through increasing later skills, the results are unchanged by their exclusion.



Factors such as investment operate through multiple channels. They have a direct effect on income. They also have an indirect effect through years of schooling, cognition, and socioemotional skills. The indirect effects themselves can be indirect with investments raising cognition, which then raises schooling, which itself shifts child income.

Table 2.7 shows the fraction of the IGE explained by each of the factors. Columns (1) and (3) show the direct effects estimated from equation 5. In line with the XGBoost results, both schooling and cognition play a large role in explaining the association between child and parent income. 75% of the IGE in the NCDS can be explained by these two factors alone, although this falls to 49% in the BCS. Interestingly, there is a larger role played by school quality and family background in the BCS. Each contributed equally with a total contribution of 22% against a statistically insignificant 9% in the BCS. This is not in direct contradiction with the findings presented earlier. The ML algorithm predicts upward mobility while the mediation analysis highlights factors that strengthen the persistence between parental and child income. The relatively larger role played by school quality and family background in the latter analysis highlight that while these variables are not strongly predictive of upward mobility, they do strengthen income persistence at the top. This is highlighted in the earlier discussion where school quality - particularly in the BCS - differs between those born into the top 20% and the bottom 80% but is very similar between those who are and who are not upwardly mobile. A similar point can be made about several family background variables in Tables 2.3 and 2.4.

Columns (2) and (4) allow for more indirect channels between measures. While the direct effect of schooling and cognition remain significant across both cohorts (a combined effect of 22% in the NCDS and 23% in the BCS), the roles of investments and family background grow in importance once indirect channels are allowed for. This is true for parental time investments in the BCS where the share explained increases from -2.2% to 21.4%. In both cohorts, the role played by family background variables rises so that around 25% of the IGE can be explained by these variables across both cohorts.

As noted earlier, variables that explain the IGE need not be the same as those that explain the likelihood of upward mobility. However, the mediation analysis does highlight an increased role for investments and family background once they are allowed to influence income via cognition and schooling. Table 2.8 adds to this by focusing solely on the 80% of observations on whom the ML model is estimated. For these, I estimate a simple linear model of skill production where years of schooling is regressed on standardised cognitive and socioemotional factors scores and family background variables. I then estimate how cognition, socioemotional skills, and investments combine to produce skills (in this case cognition at later ages).

The estimates show several patterns. Firstly, socioemotional skills only weakly relate to years of schooling once cognition is conditioned on. A standard deviation change in cognition increases years of schooling by between 1.8 (NCDS) and 1.2 years (BCS) whereas externalising skills increase years of education by 0.23 (NCDS) and 0.09 (BCS). Despite this, externalising skills are important in so far as they increase cognition. Across both cohorts externalising skills at age 11 have a positive effect upon cognition at age 16. A unit increase in the standardised measure leads to a 0.044 increase in cognition in the NCDS and a 0.149 increase in the BCS. For the BCS, this effect is sizable (the coefficient on age 11 cognition is 0.598 as a comparison).

In line with earlier studies (see for instance Cunha and Heckman, 2008), investments - in this case school quality and parental time - also shift cognition. However, they only have a quantitatively important effect on early cognition (age 11) which then increases later attainment. Investments at age 11 have minimal effect on cognition at age 16.[37]

[37] While investments at later age do increase socioemotional skills, these only weakly relate to years of schooling and (as shown in Figures 6 and 7) have limited effect on the likelihood of upward mobility.

2.6 Conclusion

In this paper, I use tools from the machine learning explainability literature to highlight how patterns of upward mobility have changed across two British birth cohorts. While mobility in the cohorts has been studied extensively, the approach here adds to an emerging literature that aims to understand the mechanisms driving mobility outcomes (Chetty and Hendren, 2018; Blanden et al, 2007; Bolt et al, 2021).

The approach makes use of the multiplicity of measures at multiple ages available in cohort studies by using a latent factor framework to condense multiple proxy measurements into low dimensional indices that capture parental time investments, skills, and school quality. These are used in a predictive ML model whose parameters are explicitly chosen to avoid overfitting - a problem inherent in cohort studies that typically have small numbers of observations relative to the number of 'features' used in the predictive model.

Tools from coalitional game theory - Shapley values - are then used to explain how each feature contributes to the model's prediction. One's origin plays an important role, but this is somewhat mechanical - those born in the bottom quintile have more 'room to move up' than those whose only means of being upwardly mobile is to break into the top 20 percent of income earners. Aside from this, cognition and years of schooling play an important role irrespective of gender and the income quintile one was born into. While education and cognition are equally important in each cohort, their relative roles reverse between 1958 and 1970. Years of education have less additional predictive power once cognition is conditioned on for NCDS members than BCS members. For the latter, years of education have significant additional predictive power even when cognition at adolescence is controlled for. Rising returns to education in the UK labour market (for example see Gosling et al 2000) may well contribute to this result.

For both the NCDS and the BCS, educational outcomes and cognition at adolescence act as sufficient statistics for a range of early childhood factors. Once these are known, there is little predictive role for family composition, school quality, parental time investments, or socioemotional skills. In line with the vast literature on human capital production, these factors are shown to influence cognition, and through this, educational outcomes; however, my results highlight that they play little to no *direct* role in predicting upward mobility even in a model that allows for arbitrary interactions between these factors, adolescent cognitive ability, and completed years of schooling.

Table 2.1: Mobility in the cohorts

	NCDS 1958		BCS 1970	
	Age 33	Age 42	Age 34	Age 42
Elasticity	0.289	0.274	0.389	0.354
	(0.035)	(0.037)	(0.033)	(0.035)
Rank Slope	0.163	0.158	0.280	0.229
	(0.016)	(0.016)	(0.020)	(0.020)
Sample Size	4096	3740	2348	2426

Notes: Robust standard errors in parenthesis. The independent variable in each case is the log/rank of parental income at age 16. For a detailed description of parental income data in the NCDS, see Dearden, Machin, and Reed (1997). Sample sizes refer to those used in the training of the ML models later. Ranks are computed based on the full sample of those with observed earnings.

Table 2.2: Transition matrices, NCDS

	Cohort member income quintile, age 42, 2000					% Upwardly Mobile
Parental income quintile	1	2	3	4	5	
1	23.59	22.40	23.29	17.95	12.76	76.41
	(23.59)	(45.99)	(69.29)	(87.24)	(100)	
2	22.83	23.36	21.23	18.16	14.42	53.81
	(22.83)	(46.19)	(67.42)	(85.58)	(100)	
3	21.40	19.33	20.75	20.62	17.90	38.52
	(21.40)	(40.73)	(61.48)	(82.10)	(100)	
4	18.44	19.43	18.19	19.93	24.01	24.01
	(18.44)	(37.87)	(56.06)	(75.99)	(100)	
5	16.12	15.18	16.94	21.68	30.08	0
	(16.12)	(31.30)	(48.24)	(69.92)	(100)	

Notes: Quintile 1 refers to those in the lowest income grouping. Numbers in parentheses are cumulative percentages across columns. The final column gives the percentage of individuals from each parental income quintile to achieve upward mobility. Overall, 47% of the NCDS sample and 45% of the BCS sample are upwardly mobile according to our main outcome measure. For our secondary measure, where we consider the likelihood of those in the bottom two quintiles reaching the top two, the likelihood of upward mobility is 23% in each cohort.

Table 2.3: Transition matrices, BCS

Parental income quintile	Cohort member income quintile, age 42, 2012: 1	2	3	4	5	% Upwardly Mobile
1	24.45	26.37	22.25	15.66	11.26	75.55
	(24.45)	(50.82)	(73.08)	(88.74)	(100)	
2	20.89	25.53	23.21	18.57	11.81	53.59
	(20.89)	(46.41)	(69.62)	(88.19)	(100)	
3	21.73	22.54	16.30	20.32	19.11	39.44
	(21.73)	(44.27)	(60.56)	(80.89)	(100)	
4	18.99	17.32	18.81	23.09	21.79	21.79
	(18.99)	(36.31)	(55.12)	(78.21)	(100)	
5	14.26	13.18	14.62	21.12	36.82	0
	(14.26)	(27.44)	(42.06)	(63.18)	(100)	

Notes: See Table 2.2.

Table 2.4: Descriptives, Demographics, NCDS

	Bottom 80% Parental Income				Top 20%
	Overall	Upwardly Mobile	Not Upwardly Mobile	Residualised Difference	Overall
Female	0.5	0.289	0.672	-0.471 (0.016)	0.527
Birthweight (Kg)	3.349	3.406	3.310	0.137 (0.021)	3.386
BMI (Age 10)	17.464	17.416	17.489	-0.121 (0.107)	17.383
Teenage mother	0.052	0.046	0.060	-0.016 (0.009)	0.041
Parents married at birth	0.973	0.975	0.970	0.014 (0.007)	0.98
Father figure	0.947	0.947	0.949	0.020 (0.009)	0.965
Parents divorced	0.087	0.090	0.076	-0.028 (0.010)	0.067
Home owners	0.432	0.435	0.433	0.082 (0.019)	0.709
Household size	5.074	4.958	5.071	-0.140 (0.062)	4.996
Parent attended university	0.060	0.068	0.053	0.033 (0.010)	0.232
Attended university	0.176	0.262	0.125	0.182 (0.017)	0.325

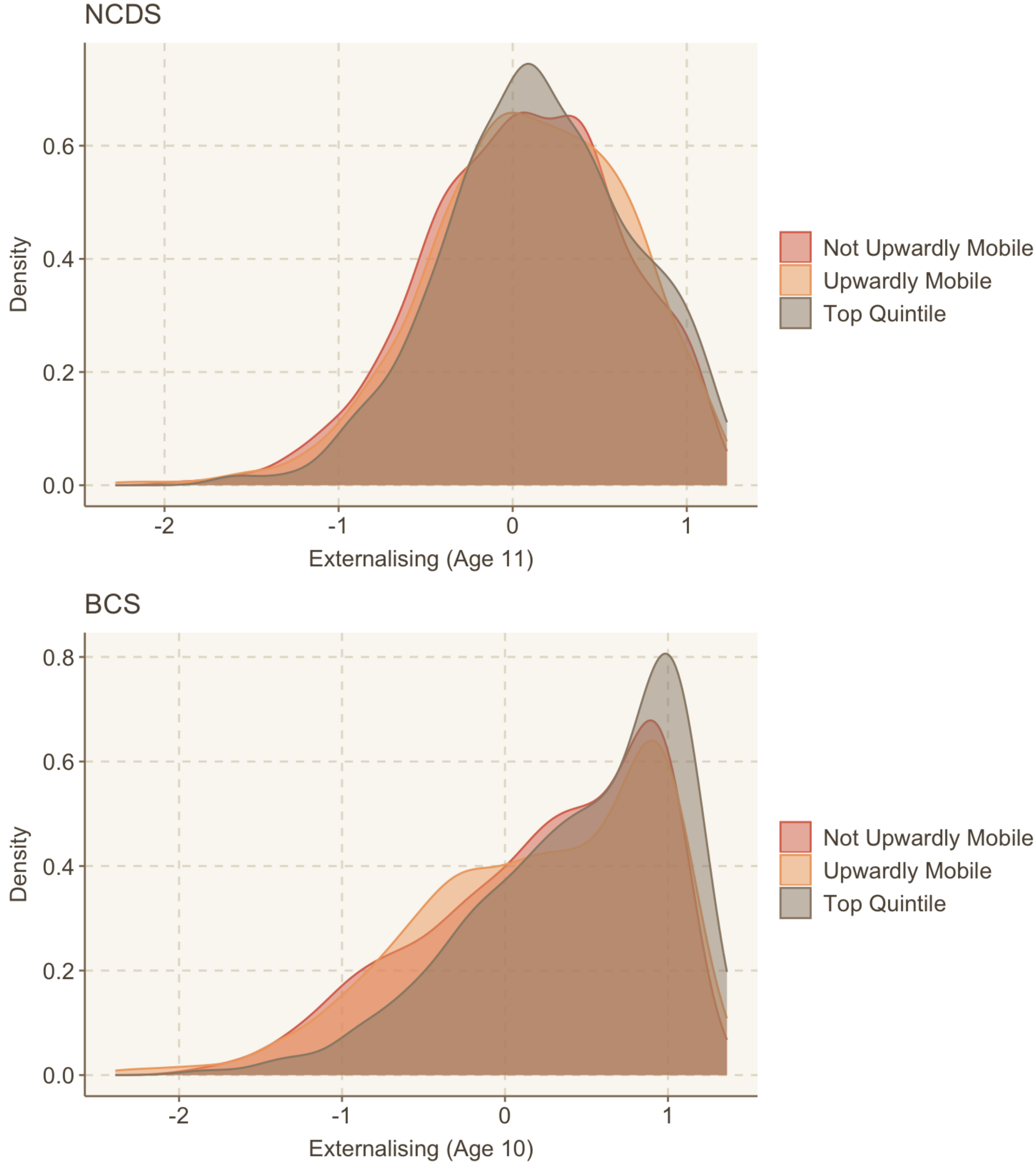
Notes: Upwardly mobile is measured as being in a higher income quintile (at age 42) than one's parents where parental income is measured at age 16. By construction, those born in the top quintile cannot be upwardly mobile. Differences in demographics by mobility status are confounded by origin. To account for this, I present regression adjusted differences that partial out the effect of parental income quintile. Robust standard errors on these adjusted differences are reported in parentheses.

Table 2.5: Descriptives, Demographics, BCS

	Bottom 80% Parental Income				Top 20%
	Overall	Upwardly Mobile	Not Upwardly Mobile	Residualised Difference	Overall
Female	0.54	0.36	0.705	-0.414 (0.022)	0.51
Birthweight	3.335	3.371	3.306	0.102 (0.026)	3.428
BMI (Age 10)	16.919	16.832	16.975	-0.215 (0.116)	16.715
Teenage mother	0.093	0.096	0.087	-0.005 (0.014)	0.046
Parents married at birth	0.956	0.959	0.956	0.019 (0.010)	0.974
Father figure	0.929	0.921	0.935	0.013 (0.013)	0.963
Parents divorced	0.171	0.214	0.133	0.010 (0.018)	0.097
Home owners	0.732	0.698	0.771	0.065 (0.020)	0.974
Household size	4.525	4.569	4.457	-0.036 (0.061)	4.313
Parent attended university	0.158	0.175	0.151	0.088 (0.019)	0.540
Attended university	0.351	0.482	0.270	0.320 (0.023)	0.621

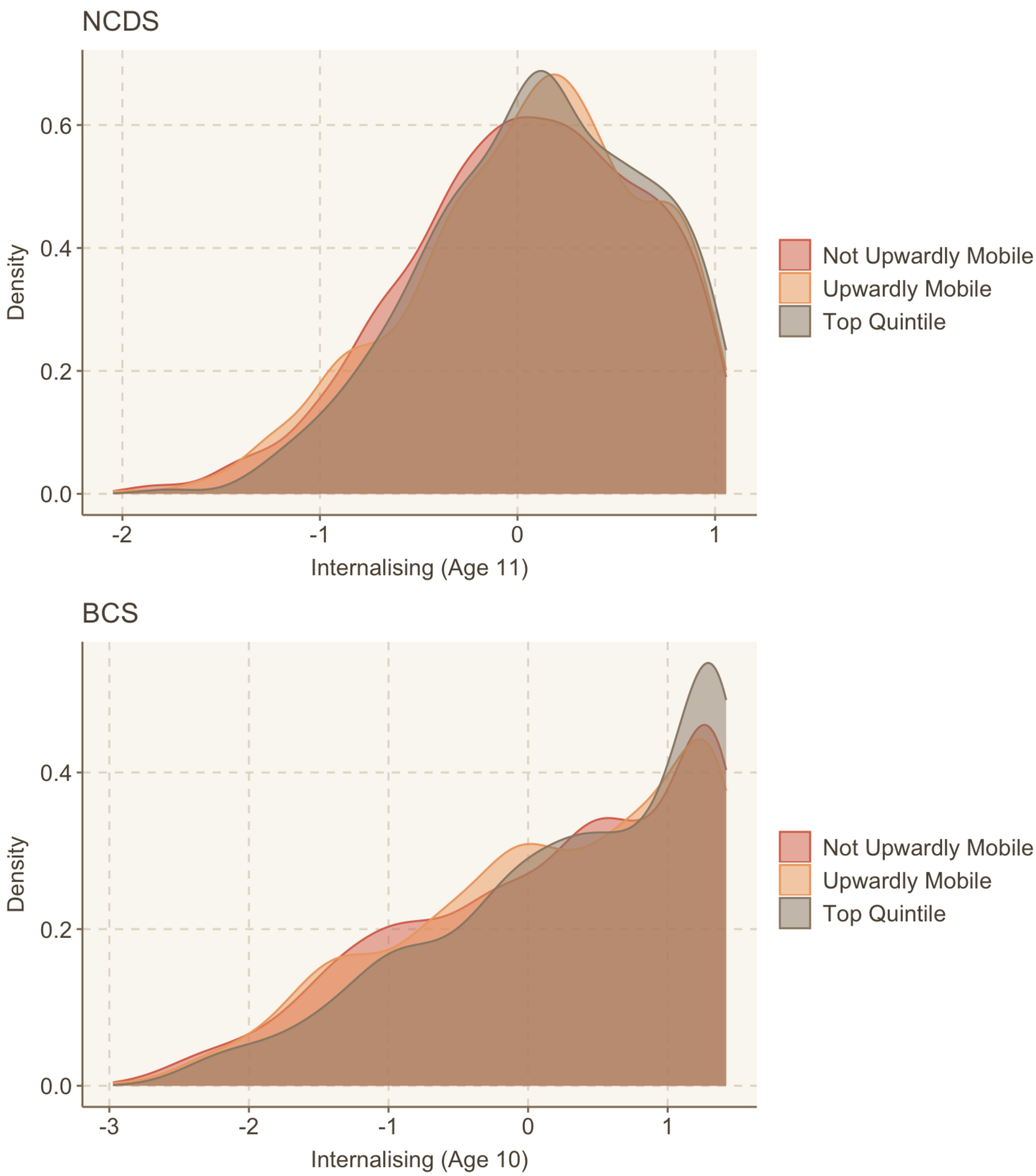
Notes: See Table 2.4.

Figure 2.1: Externalising Skills, Mid-Childhood.



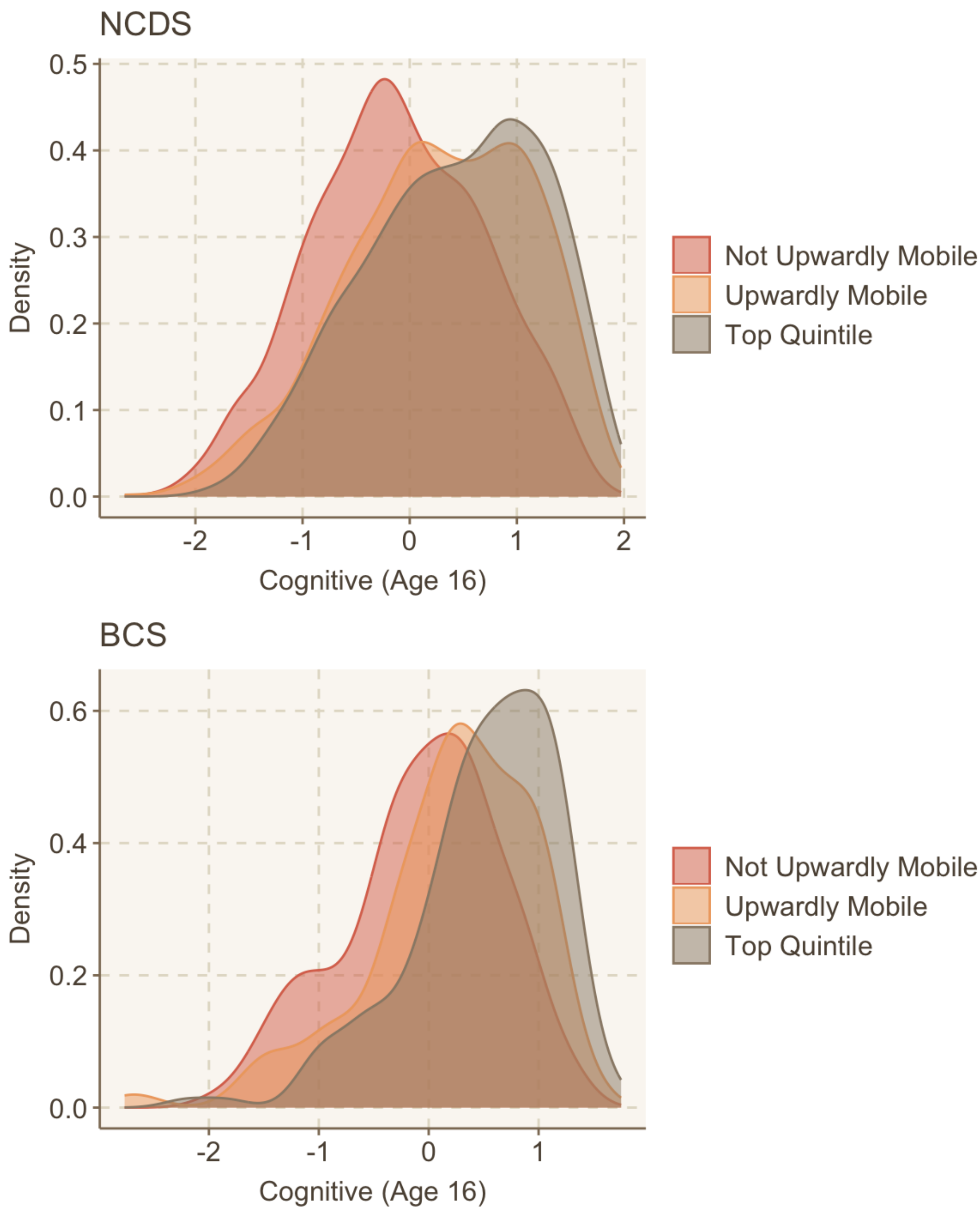
Notes: Distribution of factor scores are plotted for those who achieve upward mobility, those who do not, and those born into the top quintile of family incomes. See Appendix Tables A2.1 – A2.14 for a detailed breakdown of the factor analysis by which these scores are derived. A regression of ranked factor scores on two group indicators (with top quintile as the omitted category) results in coefficients of -3.94 (1.29) and -2.30 (1.30) for those who are not and those who are upwardly mobile respectively in the NCDS cohort. The same results for the BCS cohort are -9.78 (1.54) and -10.20 (1.62).

Figure 2.2: Internalising Skills, Mid-Childhood.



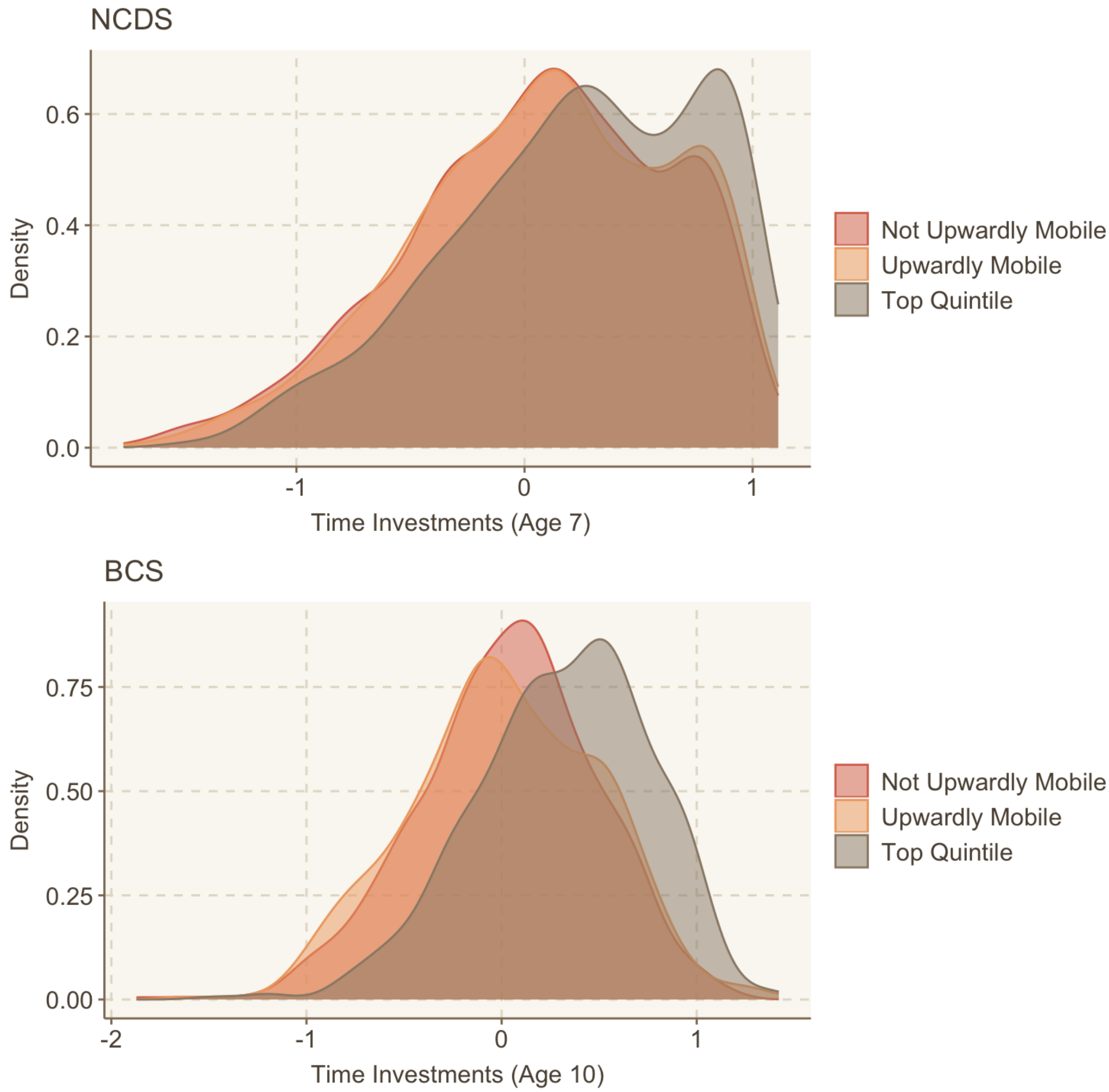
Notes: See Figure 2.1. A regression of ranked factor scores on group dummies (with top quintile as the omitted category) gives coefficients of -3.08 (1.29) and -1.80 (1.31) for those who are not and who are upwardly mobile in the NCDS cohort. For the BCS, these are -5.68 (1.58) and -5.69 (1.64).

Figure 2.3: Cognitive Skills, Age 16.



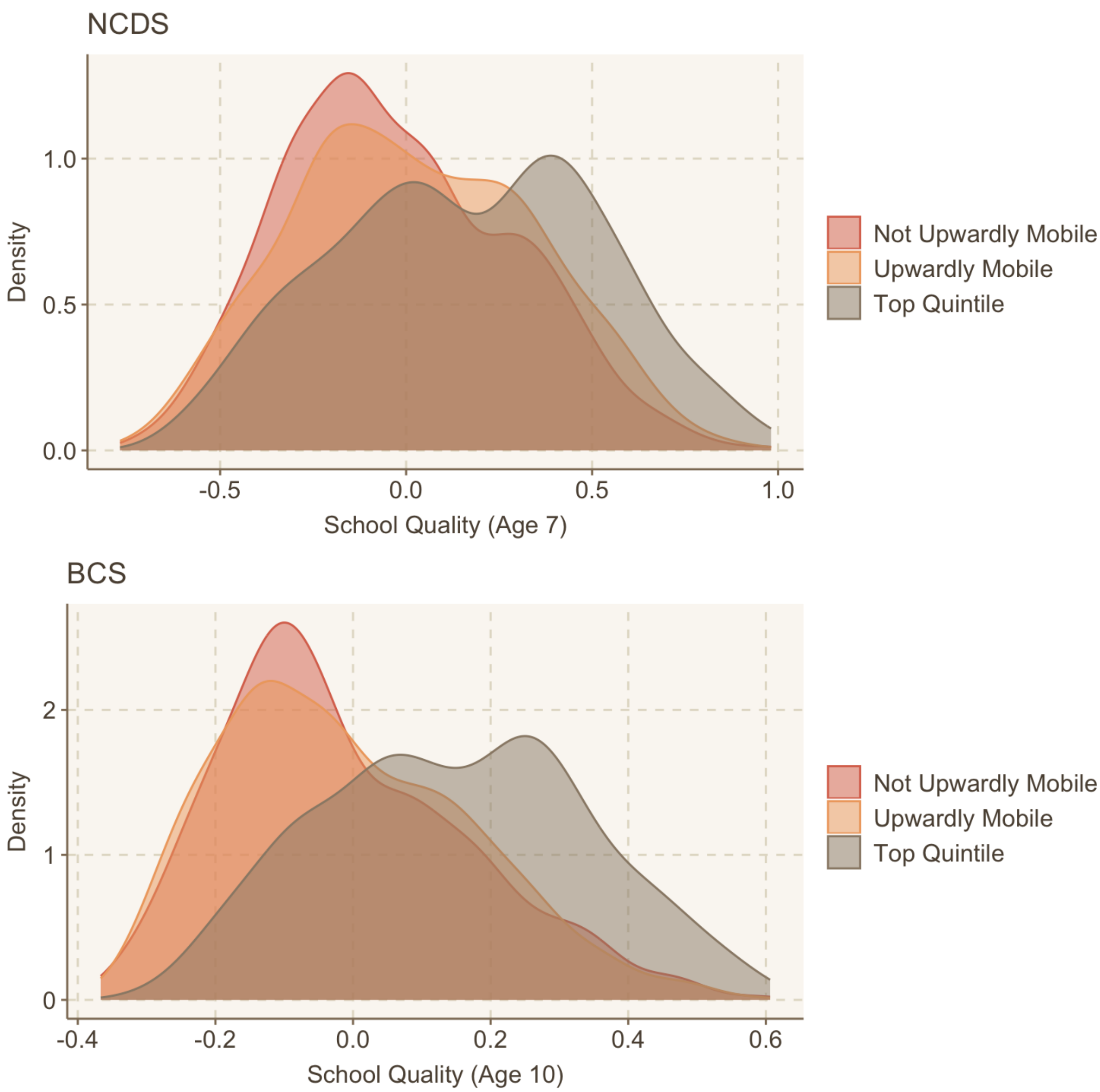
Notes: See Figure 2.1. A regression of ranked factor scores on group dummies (with top quintile as the omitted category) gives coefficients of -18.02 (1.22) and -4.60 (1.28) for those who are not and who are upwardly mobile in the NCDS cohort. For the BCS, these are -23.37 (2.74) and -11.91 (3.01).

Figure 2.4: Time Investments, Mid-Childhood



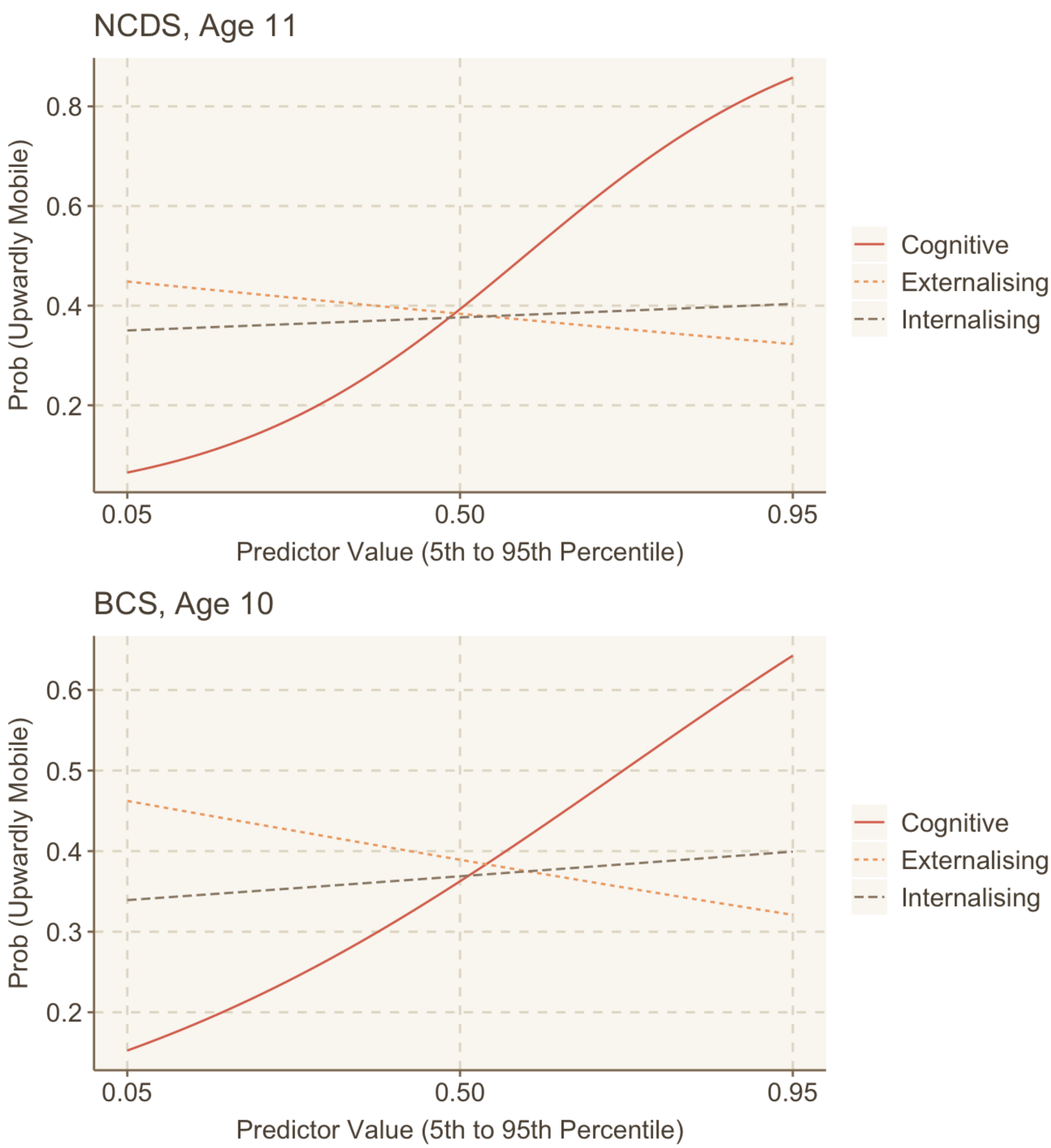
Notes: See Figure 2.1. A regression of ranked factor scores on group dummies (with top quintile as the omitted category) gives coefficients of -9.02 (1.34) and -7.98 (1.37) for those who are not and who are upwardly mobile in the NCDS cohort. For the BCS, these are -17.31 (1.51) and -17.96 (1.62).

Figure 2.5: School Quality, Mid-Childhood



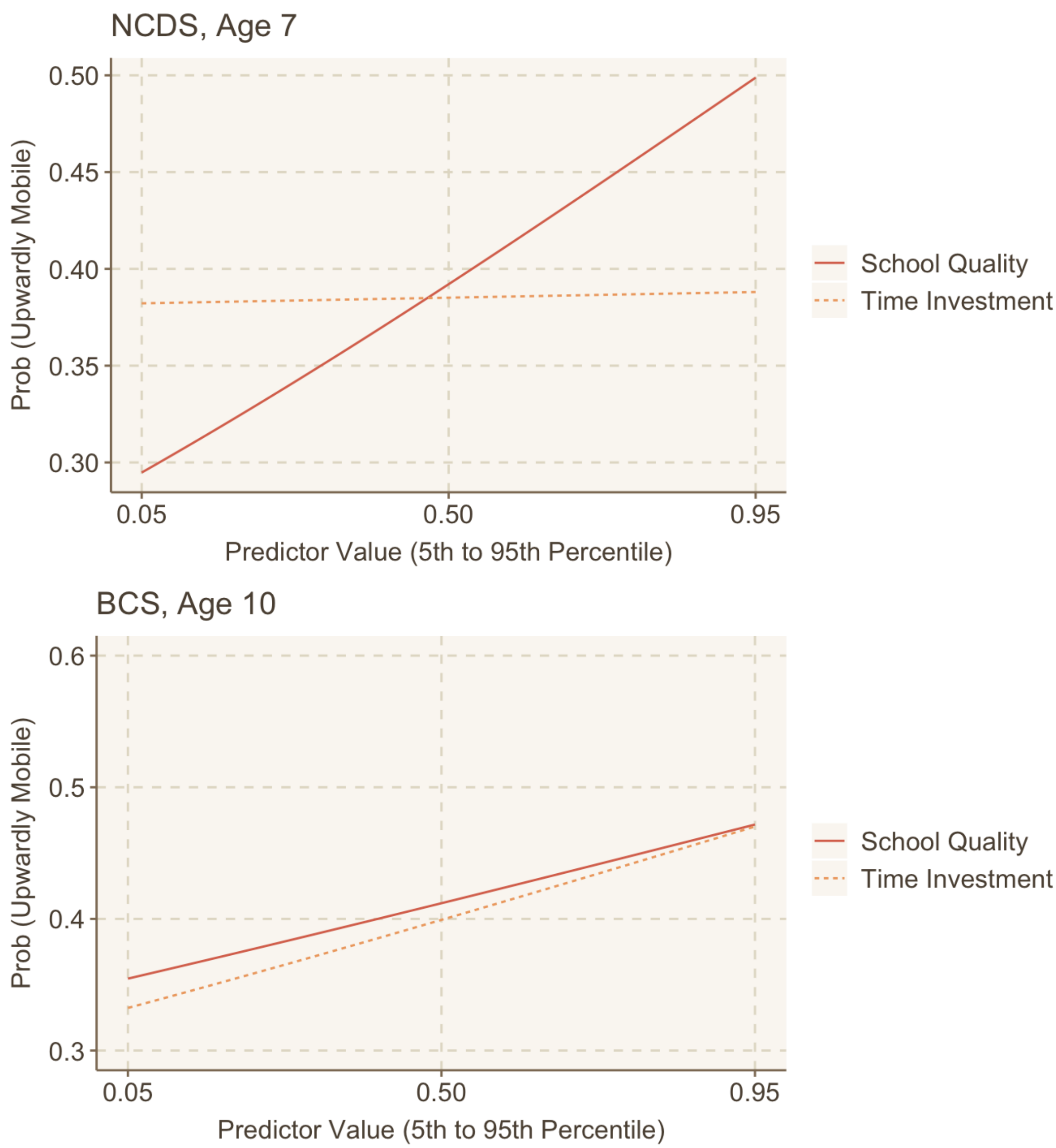
Notes: See Figure 2.1. A regression of ranked factor scores on group dummies (with top quintile as the omitted category) gives coefficients of -15.44 (1.33) and -11.34 (1.38) for those who are not and who are upwardly mobile in the NCDS cohort. For the BCS, these are -22.97 (1.46) and -22.41 (1.54).

Figure 2.6: The comparative roles of cognition and socio-emotional skills.



Notes: Probabilities are estimated via a logistic regression of an upward mobility dummy on cognitive ability, externalising and internalising skills, school quality, and time investments alongside parental income quintile. In each case, I hold all factors at their mean value and vary the factor of interest from the 5th percentile to the 95th percentile. The fitted values above are for those in the 3rd income quintile i.e. the middle of the income distribution. See the Appendix for details of the measures underlying the extracted factors.

Figure 2.7: The comparative roles of school and time investments.



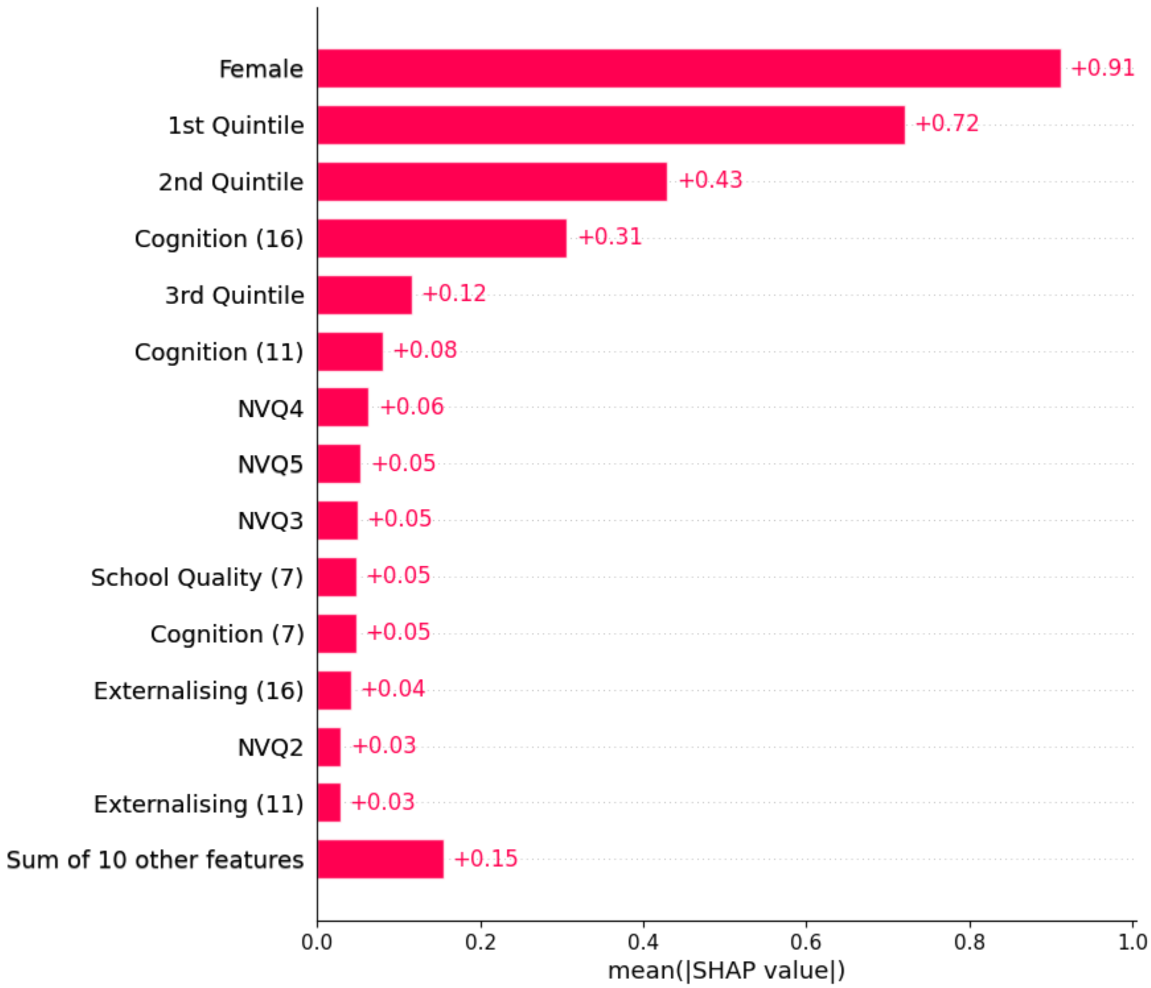
Notes: See Figure 2.6.

Table 2.6: Algorithm Performance, NCDS and BCS.

	NCDS				BCS			
	Naive	LPM	ML Boost	ML Boost	Naive	LPM	ML Boost	ML Boost
Mobility 1								
Accuracy	0.53	0.77	0.76	0.78	0.55	0.77	0.74	0.79
Precision	0	0.77	0.75	0.77	0	0.75	0.72	0.80
AUC	0.5	0.86	0.84	0.86	0.50	0.85	0.82	0.79
Mobility 2								
Accuracy	0.91	0.91	0.91	0.92	0.92	0.92	0.91	0.96
Precision	0	0	0.51	0.59	0	0	0.25	0.86
AUC	0.5	0.78	0.76	0.79	0.5	0.74	0.67	0.98
Tuned	No	No	No	Yes	No	No	No	Yes

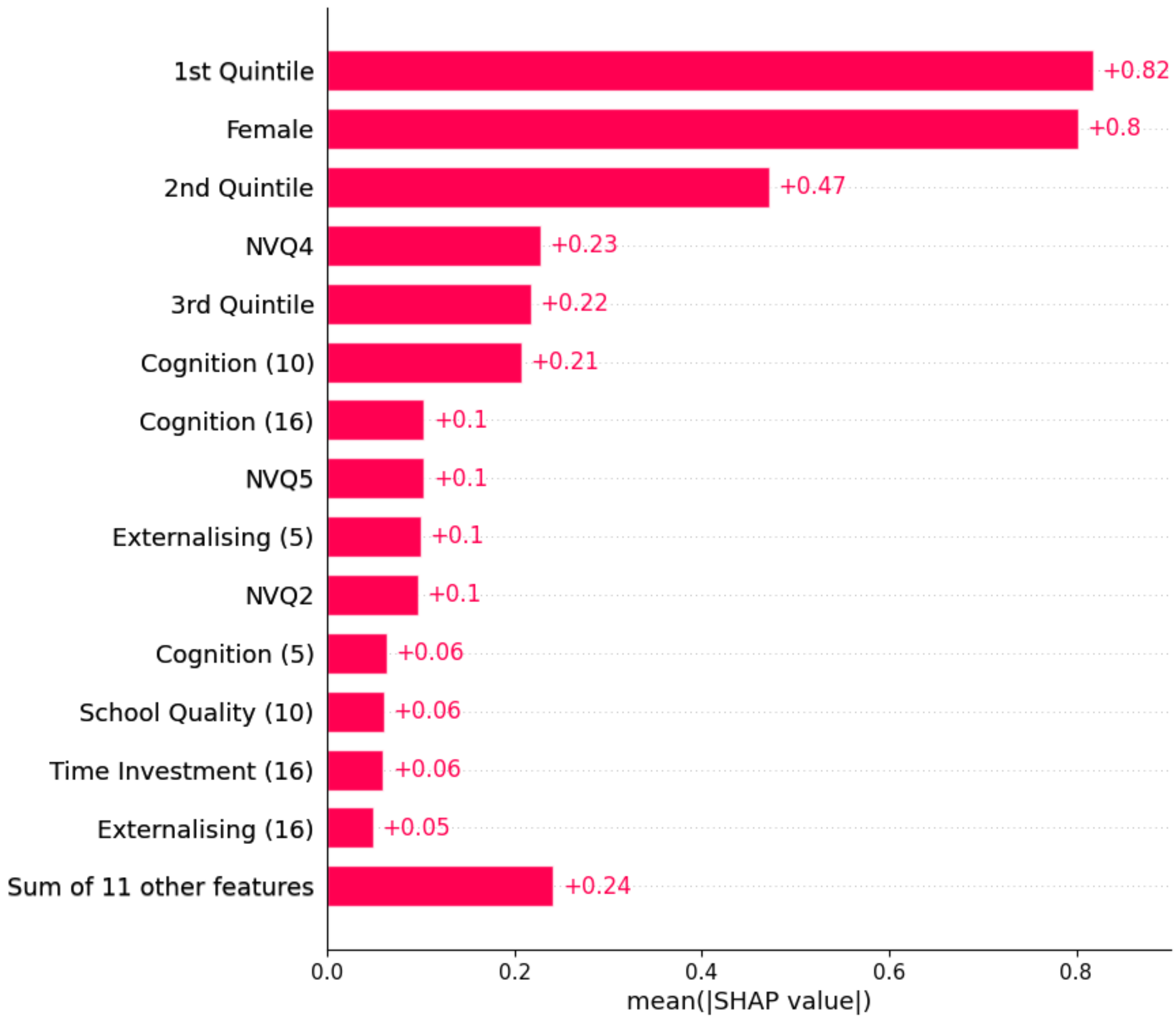
Notes: Sample sizes for the NCDS and BCS are 2844 and 1536 respectively. Mobility 1 takes value 1 if the cohort member is born in any of the bottom 4 income quintiles and resides in a higher quintile at age 42. Mobility 2 focuses on mobility from quintiles 1 and 2 (the bottom 40% of the income distribution) to quintiles 4 and 5 (the top 40%). The specific boosting algorithm used is XGBoost and parameter tuning is done via Bayesian hyperparameter optimization. In each case, performance is assessed using 5 fold cross validation. For the first mobility measure, hyperparameters (in this case the number of trees and the learning rate) are selected to maximise accuracy. In the second case, where the outcome is highly skewed for both cohorts, precision is maximised. The naive estimator simply assigns each variable to the most common class observed in the data. For the top panel, all observations are assumed to be upwardly mobile while for panel 2 all observations are assumed to not be mobile.

Figure 2.8: SHAP values, NCDS



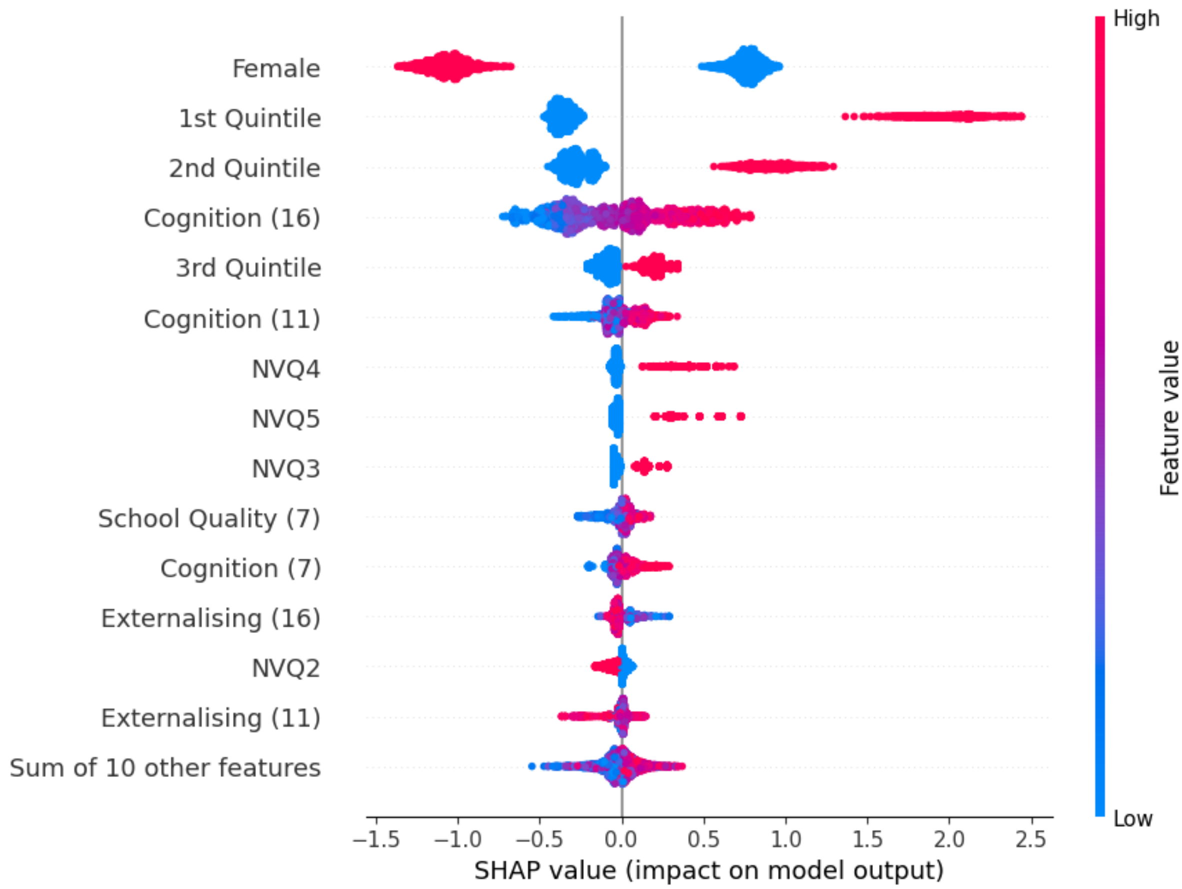
Notes: Shapley values are based on the boosting model fitted on the entire sample with optimal parameters. Features with average absolute Shapley values that are negligible are excluded from the plot. The Figure plots the average over absolute Shapley values for each feature/observation pair. The Shap values reflect the influence of each feature on the model's output where output is measured as the log odds ratio. Mobility is measured by whether the cohort member is born in any of the bottom 4 income quintiles and resides in a higher quintile at age 42.

Figure 2.9: SHAP values, BCS



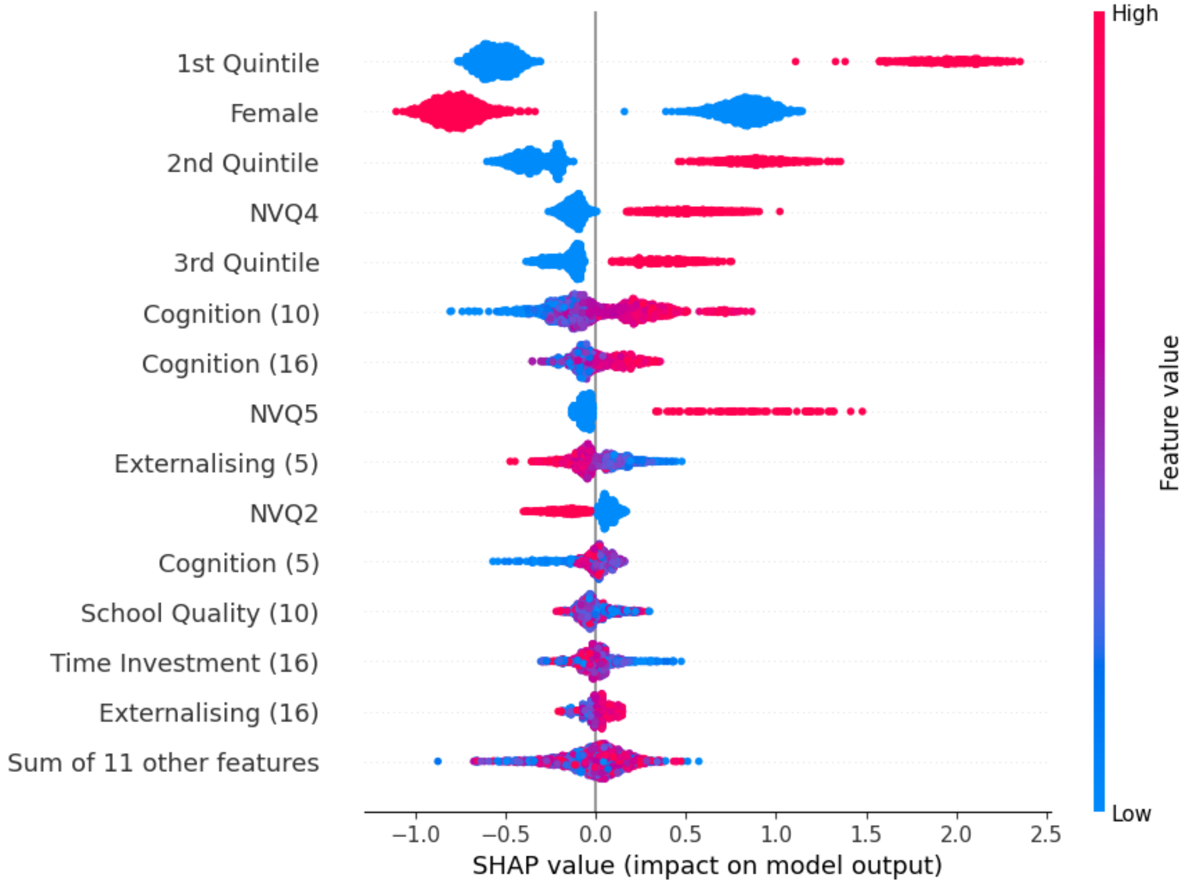
Notes: See Figure 2.8.

Figure 2.10: Beeswarm Plot, NCDS



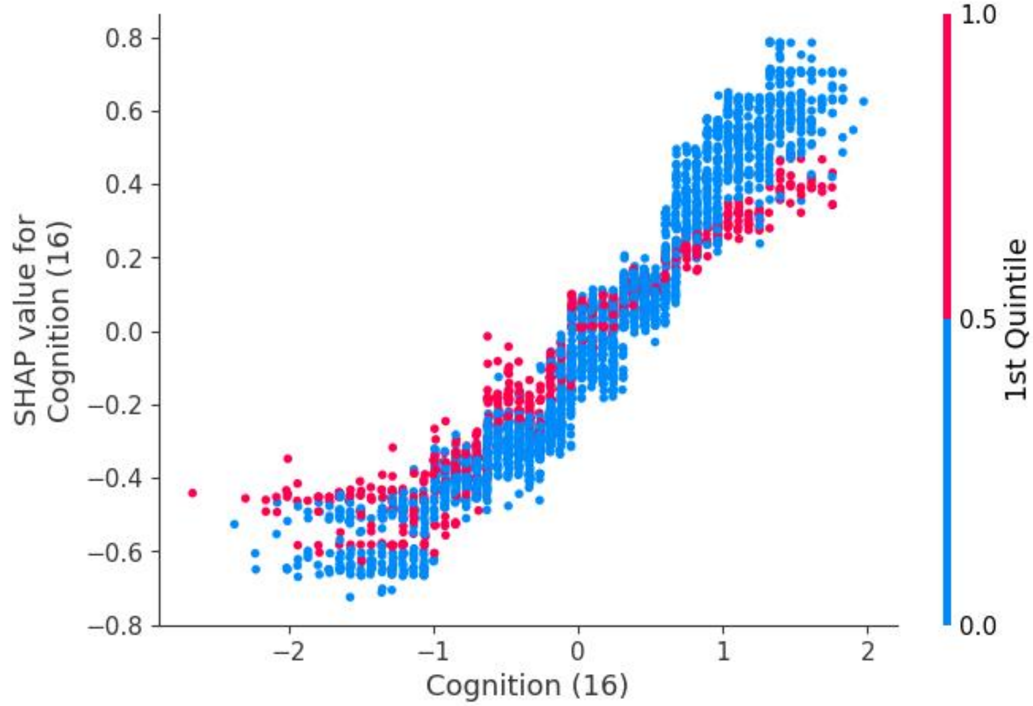
Notes: Shapley values are based on the boosting model fitted on the entire sample with optimal parameters. Features with average absolute Shapley values that are negligible are excluded from the plot. Each point represents an observation and the SHAP value for that feature for that observation. The SHAP values reflect the influence of each feature on the model's output where output is measured as the log odds ratio. Mobility 1 takes value 1 if the cohort member is born in any of the bottom 4 income quintiles and resides in a higher quintile at age 42.

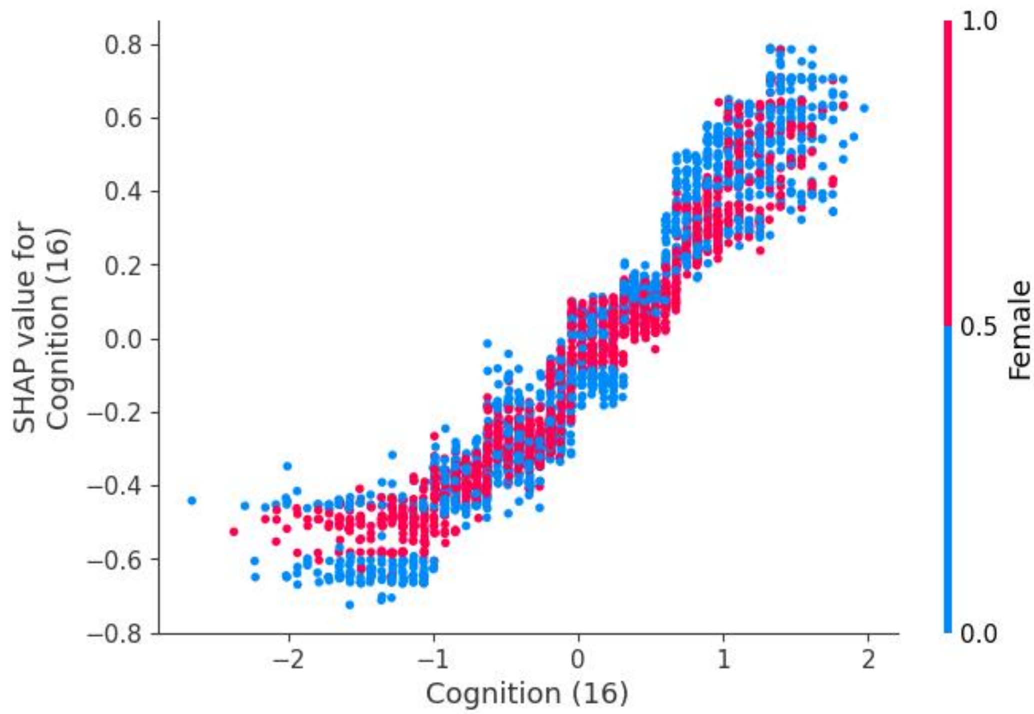
Figure 2.11: Beeswarm Plot, BCS



Notes: See Figure 2.10.

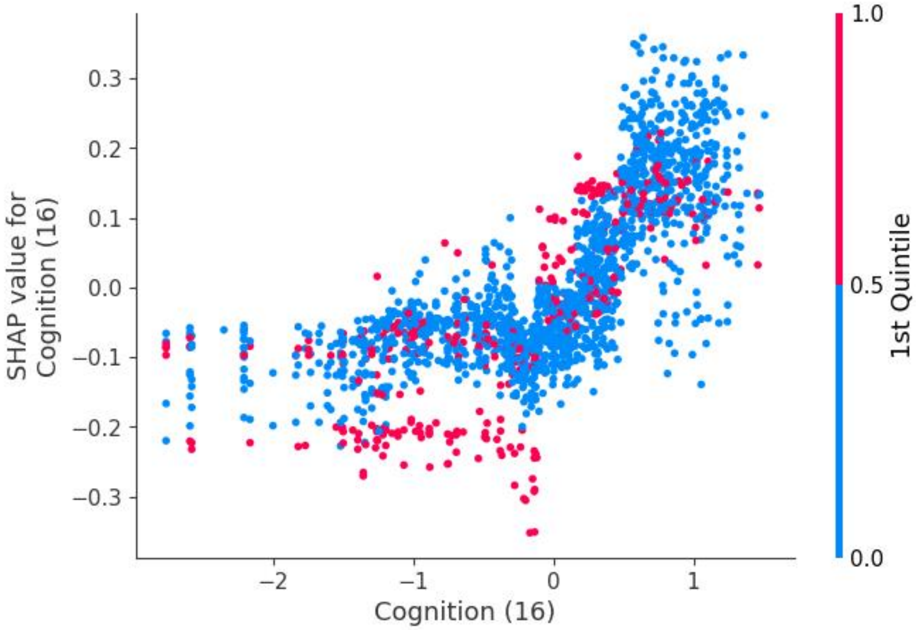
Figure 2.12: Interaction Plots, NCDS

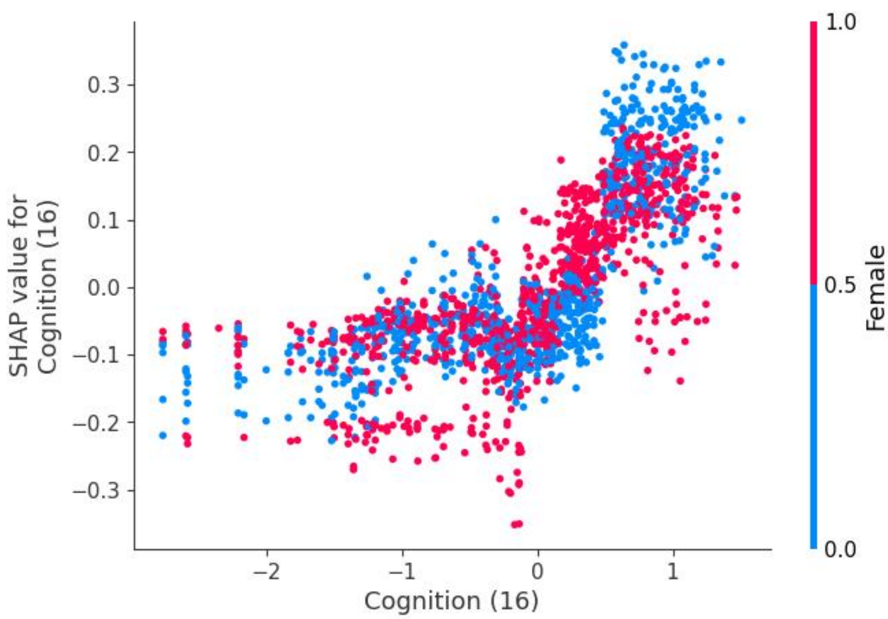




Notes: This plots the size of the interaction effects and as such does not account for the direct effect of either cognition, origin quintile, or gender.

Figure 2.13: Interaction Plots, BCS





Notes: See Figure 2.12.

Table 2.7: Mediation Analysis, NCDS and BCS.

	NCDS		BCS	
	Simple	Complex	Simple	Complex
Years of schooling	0.457	0.073	0.384	0.123
	[0.337, 0.606]	[0.021, 0.141]	[0.293, 0.493]	[0.07, 0.191]
Cognition	0.304	0.147	0.104	0.107
	[0.196, 0.444]	[0.086, 0.219]	[0.034, 0.187]	[0.061, 0.16]
Socio-emotional skills	-0.021	0.012	-0.032	-0.010
	[-0.054, 0.006]	[-0.01, 0.039]	[-0.09, 0.02]	[-0.038, 0.011]
Time investments	-0.022	0.214	0.003	0.003
	[-0.103, 0.062]	[0.139, 0.316]	[-0.01, 0.017]	[-0.006, 0.015]
School quality	0.038	0.064	0.115	0.138
	[-0.027, 0.105]	[0.022, 0.111]	[0.039, 0.194]	[0.081, 0.211]
Family background	0.054	0.248	0.101	0.250
	[-0.03, 0.157]	[0.141, 0.377]	[0.007, 0.200]	[0.145, 0.359]
Sample Size	2877	2877	1995	1995

Notes: the table shows the fraction of the IGE between parental income at age 16 and one's own income at 42 that is explained by each variable. Years of schooling is derived via converting NVQ levels into a continuous variable. Cognition, socio-emotional skills, and time investments are pooled estimates over the course of the cohort member's childhood. School quality is measured at age 10 in the BCS and age 7 in the NCDS. 95% confidence intervals are derived from 500 bootstrap replications. See the main text for an explanation of the difference between the simple and complex versions of the analysis.

Table 2.8: Linear Skill Production Function, NCDS and BCS.

	NCDS			BCS		
	Years of Schooling	Cognition (16)	Cognitive (11)	Years of Schooling	Cognition (16)	Cognitive (11)
Cognition (t-1)	1.816	0.666	0.648	1.177	0.598	0.440
	(0.058)	(0.010)	(0.014)	(0.097)	(0.015)	(0.022)
Internalising (t-1)	-0.059	-0.033	-	0.057	-0.104	0.004
	(0.073)	(0.012)		(0.117)	(0.015)	(0.025)
Externalising (t-1)	0.234	0.044	0.016	0.090	0.149	0.041
	(0.081)	(0.013)	(0.014)	(0.109)	(0.018)	(0.028)
Time Invest (t-1)	-	0.023	0.037	-	0.009	0.086
		(0.010)	(0.018)		(0.014)	(0.023)
School Quality (t-1)	-	0.031	0.132	-	0.018	0.153
		(0.010)	(0.018)		(0.014)	(0.021)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	2312	2312	2312	1534	1534	1534

Notes: estimation is based on those in the bottom 80% of the parental income distribution. Latent variables are standardised within the sample. Controls are all family background variables alongside dummies for parental income quintiles. Standard errors are in parentheses. For the NCDS, I use a single socio-emotional skill measure at age 7. The coefficient on this is reported in the externalising (t-1) row. Similarly, I only have a single school quality measure for the BCS (age 10) and NCDS (age 7). These are used as an input for both the age 16 and age 10/11 production function estimates.

References

Agostinelli F and Wiswall M (2024) Estimating the Technology of Children's Skill Formation, *Journal of Political Economy*, forthcoming.

Athey S and Imbens GW (2019) Machine Learning Methods That Economists Should Know About, *Annual Review of Economics*, 11(1), 685-725.

Attanasio O, Blundell R, Conti G, and Mason G (2020) Inequality in socio-emotional skills: A cross-cohort comparison, *Journal of Public Economics*, 191,1041-1071.

Blanden J, Doepke M, and Stuhler J, Educational Inequality. IZA Discussion Paper No. 15225.

Blanden J, Eyles A, and Machin S (2023) Intergenerational home ownership, *J Econ Inequal*, 21, 251–275.

Blanden J, Gregg P, and Machin S (2005) Intergenerational Mobility in Europe and North America, Sutton Trust.

Blanden J, Gregg P, and Macmillan L (2007) Accounting for Intergenerational Income Persistence: Noncognitive Skills, Ability and Education, *The Economic Journal*, 117, 43-60.

Blanden J and Machin S (2004), Educational Inequality and the Expansion of UK Higher Education, *Scottish Journal of Political Economy*, 51, 230-249.

Blundell J and Risa E (2018) Intergenerational Mobility as a Prediction Problem: Theory and Evidence from Norway and Britain, Unpublished manuscript.

Bolt U, French E, Hentall-MacCuish J, and O'Dea C (2021) Intergenerational altruism and transfers of time and money: a life cycle perspective, Institute for Fiscal Studies.

Bukodi E and Goldthorpe JH (2018) Social Mobility and Education in Britain: Research, Politics and Policy, Cambridge: Cambridge University Press.

Chetty R and Hendren N (2018) The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates, *The Quarterly Journal of Economics*, 133 (3), 1163–1228.

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, *The Quarterly Journal of Economics*, 129(4), 1553–1623.

Cunha F, and Heckman JJ (2008) Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation, *The Journal of Human Resources*, 43(4), 738–82.

Cunha F, Heckman JJ, and Schennach SM (2010) Estimating the Technology of Cognitive and Noncognitive Skill Formation, *Econometrica*, 78, 883-931.

Dearden L, Machin S, and Reed H (1997) Intergenerational Mobility in Britain, *The Economic Journal*, Volume 107, 47–66.

Deming D (2017) The Growing Importance of Social Skills in the Labor Market, *The Quarterly Journal of Economics*, 132(4), 1593–164.

Engzell P and Mood C (2023) Understanding Patterns and Trends in Income Mobility through Multiverse Analysis. *American Sociological Review*, *88*(4), 600-626.

Erikson R and Goldthorpe JH (2010) Has social mobility in Britain decreased? Reconciling divergent findings on income and class mobility, *The British Journal of Sociology*, 61, 211-230.

Gosling A, Machin S, and Meghir C (2000) The Changing Distribution of Male Wages in the UK, *Review of Economic Studies*, 67, 635-66.

Gregg P, Macmillan L, and Vittori C (2017) Moving Towards Estimating Sons' Lifetime Intergenerational Economic Mobility in the UK, *Oxf Bull Econ Stat*, 79, 79-100.

Hastie T, Tibshirani R, and Friedman J (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.) Stanford University.

Heckman JJ and Mosso S (2014) The Economics of Human Development and Social Mobility, *Annual Review of Economics*, 6(1), 689-733.

Heckman JJ, Pinto R, and Savelyev P (2013) Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes, *American Economic Review*, 103 (6), 2052-86.

Heckman JJ, 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.

Horn JL (1965) A rationale and test for the number of factors in factor analysis, *Psychometrika, 30*(2), 179–185.

Joshi H, Nasim B, Goodman A (2016) The Measurement of Social and Emotional Skills and their Association with Academic Attainment in British Cohort Studies, Non-cognitive Skills and Factors in Educational Attainment. Contemporary Approaches to Research in learning Innovations, Khine, M.S., Areepattamannil, S. (eds), SensePublishers, Rotterdam.

Kaiser HF (1960) The application of electronic computers to factor analysis, *Educational and Psychological Measurement*, 20, 141–151.

Kearney MS (2023) The Two-Parent Privilege, University of Chicago Press Economics Books, University of Chicago Press.

Ledesma RD, Valero-Mora P and Macbeth G (2015) The Scree Test and the Number of Factors: a Dynamic Graphics Approach, *Span J Psychol*, March 17, 18(11).

Lundberg SM and Lee S (2017) A unified approach to interpreting model predictions, *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777.

Moulton V, McElroy E, Richards M, Fitzsimons E, Northstone K, Conti G, Ploubidis GB, Sullivan A, and O'Neill D (2020) A guide to the cognitive measures in five British birth cohort studies, CLOSER.

Mullainathan S and Spiess J (2017) Machine Learning: An Applied Econometric Approach, *Journal of Economic Perspectives*, 31 (2), 87-106.

Muthén B (1984) A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators, *Psychometrika, 49*(1), 115–132.

Revelle W and Rocklin T (1979) Very Simple Structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research,* 14(4), 403–414.

Velicer WF (1976) Determining the number of components from the matrix of partial correlations. *Psychometrika,* 41*,* 321-327.

Appendix

Factor analysis

Table A2.1: Estimating the Number of Factors, Skill Measures

	NCDS Age 7	NCDS Age 11	NCDS Age 16	BCS Age 5	BCS Age 10	BCS Age 16
Optimal Coordinates	4	5	6	5	5	6
Acceleration Factor	1	1	1	1	2	1
Parallel Analysis	4	5	6	6	5	6
Kaiser	4	5	6	6	5	6
VSS 1	2	2	1	3	2	2
VSS 2	3	3	2	3	3	2
Velicer	3	2	3	3	4	4

Table A2.2: Estimating the Number of Factors, Investment Measures

	NCDS Age 7	NCDS Age 11	NCDS Age 16	BCS Age 5	BCS Age 10	BCS Age 16
Optimal Coordinates	3	2	6	2	4	2
Acceleration Factor	1	1	1	2	1	1
Parallel Analysis	5	5	8	3	6	6
Kaiser	5	5	8	3	6	6
VSS 1	3	4	6	6	8	2
VSS 2	5	5	6	6	8	2
Velicer	3	2	1	2	2	2

Notes: There are a number of methods to compute the optimal number of factors. Above are scree based approaches (Ledesma et al, 2015), parallel analysis (Horn, 1965), the Kaiser rule (Kaiser, 1960), Very Simple Structure based approaches (Revelle and Rocklin, 1979), and Velicer's method (Velicer, 1976).

Table A2.3: Factor Loadings, NCDS Age 7, Skill Measures

	Factors		
	Cognitive	Socio-emotional	Communality
Squirmy	-0.01	0.63*	0.4
Destructive	-0.16	0.50*	0.3
Fights	-0.12	0.36*	0.16
Worries	0.20	0.44*	0.2
Lonely	0.01	0.19	0.04
Irritable	0.03	0.62*	0.38
Miserable	0.12	0.58*	0.33
Twitches	0.04	0.38*	0.14
Sucks thumb	0.06	0.12	0.01
Disobedient	-0.05	0.58*	0.35
Concentrate	-0.14	0.54*	0.35
Upsets Easily	0.12	0.33*	0.11
Bullied	-0.03	0.42*	0.18
Bites nails	0.02	0.23	0.05
Copying Designs	0.49*	-0.01	0.24
Problem Arithmetic	0.63*	0.01	0.4
Human Figure Drawing	0.53*	0.01	0.28
Southgate Reading	0.71*	-0.02	0.51

Notes: Joshi, Nasim, and Goodman (2016) and Moulton et al (2020) provide comprehensive overviews of socio-emotional and cognitive skill measures in the NCDS/BCS. Measures that are retained in the fully dedicated measurement system are highlighted by an asterisk. I retain measures with a factor loading of at least 0.3 in absolute terms. In every case, measures that form the socio-emotional/internalising/externalising skill scales are taken from the SDQ. Cognitive tests are standardised to have mean 0 and standard deviation 1.

Table A2.4: Factor Loadings, NCDS Age 11, Skill Measures

	Factor			
	Cognitive	Externalising	Internalising	Communality
Squirmy	0.00	0.54*	0.14	0.352
Destructive	-0.08	0.65*	-0.05	0.434
Fights	-0.10	0.54*	-0.19	0.298
Worried	0.02	0.00	0.81*	0.660
Lonely	0.00	0.04	0.15	0.028
Irritable	0.04	0.60*	0.10	0.392
Miserable	0.05	0.46*	0.27	0.347
Twitches	-0.01	0.21	0.19	0.105
Sucks Thumb	0.04	0.15	0.09	0.038
Disobedient	0.07	0.70*	-0.08	0.442
Concentration	-0.16	0.46*	0.12	0.325
Upsets Easily	-0.05	0.00	0.58*	0.343
Bullied	-0.15	0.21	0.32*	0.226
Bites Nails	0.03	0.21	0.06	0.052
Reading	0.82*	-0.02	0.03	0.686
Math	0.90*	0.01	-0.04	0.808
General Ability	0.90*	0.00	0.01	0.801
Copying Designs	0.37*	0.00	-0.02	0.139

Notes: See Table A2.1.

Table A2.5: Factor Loadings, NCDS Age 16, Skill Measures

	Factors		
	Externalising	Internalising	Communality
Restless	0.47*	0.23	0.35
Squirmy	0.49*	0.26	0.4
Destructive	0.74	-0.04	0.52
Fights	0.7	-0.02	0.49
Disliked	0.38	0.32	0.33
Worries	-0.03	0.73*	0.52
Lonely	0.01	0.38*	0.14
Irritable	0.56*	0.18	0.42
Miserable	0.35	0.44	0.43
Twitches	0.24	0.29	0.19
Sucks thumb	0.14	0.20	0.08
Bites nails	0.24	0.11	0.09
Disobedient	0.77*	-0.09	0.55
Concentration	0.55*	0.23	0.44
Upsets easily	-0.07	0.65*	0.39
Fussy	0.02	0.32*	0.11
Lies	0.76*	-0.07	0.55
Bullies	0.70*	-0.08	0.45

Notes: See Table A2.3.

Table A2.6: Factor Loadings, BCS Age 5, Skill Measures

	Factor			
	Cognitive	Externalising	Internalising	Communality
Restless	0.01	0.54*	0.05	0.31
Squirmy	0.05	0.49*	0.15	0.29
Destructive	-0.06	0.69*	-0.11	0.47
Fights	0.	0.67*	-0.08	0.43
Disliked	-0.10	0.40*	0.24	0.3
Worried	0.03	-0.05	0.76*	0.57
Lonely	-0.07	0.09	0.34*	0.15
Irritable	0.04	0.50*	0.25	0.36
Miserable	-0.03	0.29	0.48*	0.38
Takes things	-0.07	0.60*	-0.13	0.37
Twitches	-0.06	0.21	0.21	0.12
Sucks thumb	0.05	-0.01	0.07	0.01
Bites nails	-0.06	0.13	0.13	0.05
Disobedient	0.05	0.72*	0.03	0.51
Concentration	-0.14	0.51*	0.10	0.36
Upsets easily	-0.05	-0.08	0.56*	0.3
Fussy	-0.01	0.05	0.45*	0.21
Lies	0	0.64*	0	0.41
Bullies	0.03	0.68*	0	0.45
Schonell Reading	0.29	-0.09	0.05	0.11
Complete Profile	0.36*	0.05	-0.01	0.12
EPVT	0.47*	-0.05	-0.05	0.24
Human Figure Drawing	0.57*	0.05	0.01	0.32
Copying Designs	0.67*	0	0	0.46

Notes: See Table A2.3.

Table A2.7: Factor Loadings, BCS Age 10, Skill Measures

	Factor			
	Cognitive	Externalising	Internalising	Communality
Restless	-0.01	0.61*	-0.02	0.37
Squirmy	0.07	0.61*	0.08	0.38
Destructive	-0.06	0.69*	-0.04	0.49
Fights	-0.05	0.75*	-0.11	0.55
Disliked	-0.01	0.46*	0.21	0.32
Worries	0.01	0.04	0.80*	0.66
Lonely	0.05	0.14	0.38*	0.20
Irritable	0.04	0.59*	0.23	0.48
Miserable	0.01	0.43	0.41	0.47
Twitches	0.05	0.28	0.16	0.13
Disobedient	0.01	0.75*	-0.04	0.54
Concentration	-0.13	0.57*	0.07	0.42
Upsets Easily	-0.04	-0.11	0.75*	0.53
Fussy	-0.04	0.08	0.43*	0.21
Bullies	0.00	0.74*	-0.05	0.52
Math	0.66*	-0.04	0.03	0.45
EPVT	0.86*	-0.03	0.01	0.76
BAS Digits	0.47*	0.02	-0.02	0.22
BAS Word 1	0.77*	0.02	-0.01	0.58
BAS Word 2	0.73*	0.03	-0.01	0.53
Spelling	0.64*	-0.05	0.03	0.43
Pictorial	0.73*	0.01	0.00	0.53
Friendly Math	0.84*	0.00	-0.04	0.71

Notes: See Table A2.3

Table A2.8: Factor Loadings, BCS Age 16, Skill Measures

	Factor			
	Cognitive	Externalising	Internalising	Communality
Restless	0.05	0.35*	0.3	0.31
Squirmy	-0.05	0.35	0.35	0.33
Destructive	-0.07	0.83*	0.07	0.71
Fights	0.07	0.64*	0.07	0.48
Disliked	-0.01	0.47*	0.26	0.38
Worries	0.03	-0.07	0.81*	0.61
Lonely	-0.08	0.15	0.38*	0.2
Irritable	-0.04	0.49	0.30	0.44
Miserable	0	0.39	0.46	0.51
Takes things	0.02	0.85*	-0.07	0.69
Twitches	-0.09	0.37*	0.25	0.26
Sucks thumb	-0.07	0.23	0.03	0.06
Bites nails	0.02	0.18	0.14	0.07
Disobedient	0.03	0.82*	-0.02	0.67
Concentrate	0.17	0.46*	0.29	0.48
Upsets easily	0.02	-0.05	0.63	0.37
Fussy	0	0.06	0.46*	0.23
Lies	0.08	0.82*	-0.06	0.68
Bullies	-0.01	0.81*	-0.12	0.58
Spell	0.63*	0.07	-0.06	0.42
BAS Matrices	0.53*	-0.03	0.03	0.28
EPVT	0.89*	-0.01	0.01	0.8
Reading	0.85*	-0.02	-0.02	0.71
Math	0.78*	0.02	0.04	0.62

Notes: See Table A2.3

Table A2.9: Factor Loadings, NCDS Age 7, Investment Measures

	Factor		
	Time Investment	School Quality	Communality
Class size	0.05	-0.06	0
PTA in school	-0.11	0.33*	0.09
Education meetings	-0.11	0.35*	0.11
Parent initiate meetings	-0.03	0.66*	0.43
Teacher initiate meetings	-0.09	0.32*	0.09
% parents higher managerial in school	0.02	0.36*	0.13
Mother reads to child	0.44*	0.12	0.25
Father reads to child	0.66*	0.04	0.45
Mother outings	0.65*	0.09	0.46
Father outings	0.95*	-0.03	0.88
Father role	0.54*	-0.06	0.28
Mother interest in child's education	0.02	0.82*	0.68
Father interest in child's education	0.10	0.53*	0.32

Notes: Where measures take ordinal values, factor loadings are based upon polychoric correlations. For variables related to involvement of the father, these are coded to their lowest (ordinal) value in the case of no father figure being present in the household. The measures above are measured on scales that make this natural i.e. an absent father is treated as never reading to his child, having no outings with his child, and taking no interest in his child's education. Measures are selected so as to capture non-monetary parental investments and school quality related variables.

Table A2.10: Factor Loadings, NCDS Age 11, Investment Measures

	Factor	
	Investment	Communality
% Taking GCE exams	0.29	0.09
Class size	0.01	0
Parent expect child to stay on after age 16	0.50*	0.25
Parent expects child to enter further education	0.42*	0.17
Mother outings	0.42*	0.17
Father outings	0.46*	0.22
Father role	0.30*	0.09
Parent initiate meetings	0.64*	0.41
Teacher initiate meetings	0.36*	0.13
Father interest in child’s education	0.74*	0.54
Mother interest in child’s education	0.83*	0.69
Pupil teacher ratio	0.01	0
% Teachers with 1-2 years experience	-0.01	0

Notes: See Table A2.9.

Table A2.11: Factor Loadings, NCDS Age 16, Investment Measures

	Factor	
	Investment	Communality
Number PT discussions	0.38*	0.14
% GCE Boys	0.26	0.07
% GCE Girls	0.24	0.06
% Boys staying on at school	0.43*	0.18
% Girls staying on at school	0.41*	0.17
Hours per week English lessons	-0.06	0
Hours per week math lessons	0.10	0.01
Private school	0.80*	0.65
Pupil teacher ratio	-0.20	0.04
Hope child stays on after 16	0.71*	0.50
Wish child attends university	0.68*	0.46
Expects child attends university	0.72*	0.51
Parents satisfied with child's education	0.08	0.01
School holds teacher training	0.04	0
PTA meetings	0.11	0.01
Parent discusses child's education	0.66*	0.43
Parent initiate meetings	0.40*	0.16
Father interest in child's education	0.64*	0.41
Mother interest in child's education	0.66*	0.43
% parents higher managerial in school	0.54*	0.29
English classes ability streamed	0.05	0
Math classes ability streamed	0.20	0.04

Notes: See Table A2.9.

Table A2.12: Factor Loadings, BCS Age 5, Investment Measures

	Factor	
	Investment	Communality
Parents read to child	0.35*	0.12
Mother help at school	0.29	0.08
Outing to friends' house	0.22	0.05
Outing to park	0.32*	0.10
Outing to shop	0.25	0.06
Mother met teacher	0.70*	0.50
Father met teacher	0.67*	0.45
Number of days watch TV	-0.18	0.03

Notes: See Table A2.9.

Table A2.13: Factor Loadings, BCS Age 10, Investment Measures

	Factor		
	School Quality	Time Investment	Communality
Parents met teacher	-0.08	0.50*	0.24
Wish carried on education	0.34*	0.30	0.26
Parent discuss child's education with teacher	0.14	0.44*	0.24
Class size	-0.33*	0.24	0.13
Class hours	0.18	-0.09	0.03
Father role	-0.14	0.45*	0.19
Walks with parents	-0.06	0.54*	0.28
Outings with parents	-0.08	0.60*	0.35
Chats with parents	-0.04	0.55*	0.29
Mother interest in child's education	0.28	0.49*	0.38
Father interest in child's education	0.26	0.46*	0.34
School provides homework	0.55*	-0.15	0.29
Private school	1.01*	-0.03	1
Parent spend time talking to child	0.04	0.22	0.06
% Parents higher managerial in school	0.68*	0.1	0.51
% Parents in privately owned housing	0.68*	0.03	0.48
% Pupils above average attainment	0.51*	0.07	0.28

Notes: See Table A2.9.

Table A2.14: Factor Loadings, BCS Age 16, Investment Measures

	Factor	
	Investment	Communality
Visit relatives	0.47*	0.22
Play indoor games with parents	0.55*	0.30
Go to pub/restaurant with parents	0.39*	0.15
Do recreational activities with parents	0.65*	0.42
Go to sports with parents	0.53*	0.28
Do outdoor hobbies with parents	0.54*	0.29
Do indoor hobbies with parents	0.60*	0.35
Shop with parents	0.37*	0.13
Holiday with parents	0.36*	0.13
Go to clubs with parents	0.53*	0.28
Go to church with parents	0.27	0.08
Go to theatre with parents	0.52*	0.27
Go for meal with parents	0.21	0.04
Go to cafe with parents	0.49*	0.24
Play musical instrument with parents	0.39*	0.15
Feel like can talk with parents	0.48*	0.23
Feels parents are loving	0.42*	0.18
Feels parents are helpful	0.44*	0.20
Feels parents are generous	0.37*	0.14
Spends time with mother	0.49*	0.24
Spends time with father	0.55*	0.30
Spends time with both parents	0.56*	0.32
Parents give career advice	0.21	0.04
Parents and child have shared interests	0.49*	0.24

Notes: See Table A2.9.

Chapter 3: The Introduction of Academy Schools to England's Education

Abstract

This paper studies the origins of what has become one of the most radical and encompassing programmes of school reform seen in the recent past in advanced countries – the introduction of academy schools to English education. Academies are independent state funded schools that are allowed to run in an autonomous manner outside of local authority control. Almost all academies are conversions from already existent state schools and so are school takeovers that enable more autonomy in operation than was permitted in their predecessor state. Studying the first round of conversions that took place in the 2000s, where poorly performing schools were converted to academies, a focus is placed on legacy enrolled pupils who were already attending the school prior to conversion. The impact on end of secondary school pupil performance is shown to be positive and significant. Performance improvements are stronger for pupils in urban academies and for those converting from schools that gained relatively more autonomy because of conversion.

3.1 Introduction

The English academy schools programme is turning out to be one of the most radical and encompassing programmes of school reform seen in a developed country. Unlike traditional community schools, academies are autonomous, state-funded schools that are managed and run outside of local authority control. In almost all cases, they are conversions of pre-existing schools that inherit pupils already enrolled in the school, but enjoy significantly more autonomy in operation than was permitted in their predecessor state.[38] At the time of writing, nearly 65% of England's secondary schools and a further 15% of primary schools had become academies.[39] The vast majority became academies after a change of government in May 2010 when legislation - the 2010 Academies Act - greatly widened the remit of the programme.[40]

The genesis of the English academies programme is what is studied in this paper. The programme was initiated under the 1997-2010 Labour government when strong concerns were being expressed that schools in particular local authorities (usually those serving disadvantaged urban neighbourhoods) were not delivering pupils a good enough education. A widespread recognition emerged that something needed to be done to improve standards in schools where it had been said that “teachers had lost control of the corridors”. The proposed solution was to replace an existing school with a new type of state school managed by a private team of

[38] They are different from most US charter schools which are typically, though not always, set up from scratch. A closer comparison to the typical charter school in England are free schools, recent additions that are brand new schools (often set up by parent or community groups). A closer US comparison to academies are ‘in-district’ charters where an already existent public school is converted to a charter as a school takeover – these are less commonplace than US charters, but there are places where conversions of public schools to charters have taken place (like Boston and New Orleans – see Abdulkadiroglu et al., 2016).
[39] In England, secondary schooling takes place from ages 11-16 and primary schooling from ages 5-11.
[40] Prior to the Act only secondary schools could become academies and to convert they were required to sign up a sponsor. Afterwards, primary schools were permitted to become academies, free schools were introduced, and a sponsor was no longer required for conversion to take place. See Eyles, Hupkau and Machin (2016) for more details.

independent co-sponsors. The sponsors of the new academy school delegate the management of the school to a largely self-appointed board of governors who have responsibility for employing all academy staff, agreeing levels of pay and conditions of service and deciding on the policies for staffing structure, career development, discipline, and performance management.

This paper studies the causal impact of academy school conversion on pupil performance. This line of enquiry answers two key policy questions. Did the programme have the desired effect on the population it was targeted at? And is the programme likely to benefit similar pupils in the future? As the discussion has already made clear, it was pupils in disadvantaged schools that formed the target population. This selection is accounted for in the research design by using a combination of differences-in-differences and instrumental variables. Namely, it compares outcomes of pupils enrolled in academy schools to outcomes of those enrolled in a group of comparison schools – a set of state schools that go on to become sponsored academies after the sample period ends. Potentially endogenous sorting into academies is circumvented by using enrolment in the academy before conversion – legacy enrolment - as an instrument for actual attendance. The approach has similarities to that taken in Abdulkadiroglu et al. (2016), who study charter school takeovers in New Orleans. They limit their study to what they term "grandfathered" pupils; that is, those who passively enrol in a charter, by virtue of already being enrolled in the school prior to the takeover.

Whilst this study informs the current policy debate in England, it also complements two often overlapping strands of research in the economics of education. The first focuses on how the type of school one attends influences test scores, while the second focuses on increasing the amount of autonomy in previously centralised education systems. The former is exemplified by the US literature on catholic school attendance (Altonji, Elder and Taber, 2005, Neal, 1997, and

Evans and Schwab, 1995) and charter school attendance (see Epple, Romano and Zimmer, 2016). Examples of the latter include cross country evidence on the contribution of greater school autonomy to international test score differences (Hanushek and Woessmann, 2011, 2015, and OCED, 2011) and the effects of private school voucher programs (Epple, Romano and Urquiola, 2017).

Focusing on pupils who enrolled in academies prior to conversion, the results suggest that academy schools considerably raise the achievement of their attendees. The preferred estimate is that pupils who attend an academy gain on average 0.12 of a standard deviation (σ) higher end of school test scores relative to otherwise similar pupils who attend traditional schools. This effect increases over time, with pupils who attend for four years reaping gains of 0.28σ. Suggestive evidence is also presented to show that improvements are concentrated amongst those schools that gained the most autonomy after conversion. Mirroring findings on charter schools, improvements appear stronger in schools in urban areas. Alongside performance improvements, there is evidence that schools change their intake upon conversion; in particular, incoming cohorts of students have higher baseline test scores after conversion. This legitimises the adopted research design, which uses pupil level data, and explicitly accounts for such compositional changes. Although the changes are sizeable, empirical tests also show that the estimated performance improvements do not seem to come about because performance effects and changes in peer composition are related.

3.2 Context

3.2.1 Academy Schools

Academy schools were first introduced in the early 2000s. With hindsight, their introduction can be viewed as a landmark in the history of education in England.[41] Academies are now the predominant school type in the English secondary sector but are not without controversy. The almost evangelical fervour for academisation shown by its advocates has been countered by an equal lack of enthusiasm by detractors.[42] Lord Adonis - the key player in setting up the Labour academies programme – eloquently describes this in his 2012 book (Adonis, 2012).

The first academies opened in the school year beginning in September 2002. Academies are independent, non-selective, state-funded schools that fall outside the control of local authorities. In most cases, they are conversions of already existing predecessor schools. The first tranche of academies that are studied in this paper are managed by a private team of independent co-sponsors. The sponsors of the academy school delegate the management of the school to a largely self-appointed board of governors which has responsibility for employing all academy staff, agreeing levels of pay and conditions of service and deciding on policies for staffing structure, career development, discipline, and performance management. Since the introduction of the Academies Act 2010, the programme's remit has changed. Converter academies – good and outstanding schools that gain academy status without a sponsor – now dominate the English educational landscape.[43] It is important to note that all the academies that studied in this paper,

[41] It is only England, and not in the other nations of the United Kingdom (Northern Ireland, Scotland and Wales) who run their own devolved education systems, where academies have been introduced. In the OECD's Programme for International Student Assessment (PISA) data, this has resulted in England becoming the highest ranked country in school autonomy over resource allocation in the 2012 PISA – see Eyles, Hupkau and Machin (2016) for more detail on this aspect of academisation of English schools, and the policy context more generally.

[42] For example, the anti-academies alliance (see the website at http://antiacademies.org.uk).

[43] Post-2010 the programme extended to cover primary schools (see Eyles, Machin and McNally, 2017).

both in the treatment and control group, are sponsored academies that – in the main analysis – were approved for opening prior to this change.[44]

3.2.2 Secondary School Types in England and Academy Introductions

The English secondary education system is composed of seven school types: independent schools, academy schools, city technology colleges (CTCs), voluntary aided schools, foundation schools, voluntary controlled schools, and community schools. Each school type is characterised by a unique set of features regarding their autonomy and governance. This is shown in Table 1. In the Table, the different school types are ordered by the amount of autonomy that their governing body/management body has, ranging from those with the most (private fee-paying independent schools that operate outside of the state sector) to those with the least (community schools that are largely operated under the remit of local authority control).

In the time period, under study, the main impetus of the programme was to replace failing schools with academies by moving away from the conventional school type that had populated the English secondary sector in the past.[45] The path to establishing an academy school in a local authority involved a number of steps. The key feature was the need to sign up a sponsor, who worked with the local authority (LA) where the school operates, and to complete a formal expression of interest (this made the case that an academy in the proposed area was both needed and feasible). The phase is completed when the LA and sponsor send the expression of interest to the Secretary of State for Education for his or her ministerial approval. After approval the process

[44] The latest opening date, for the control schools, is September 2010, which coincides with the first openings of converter academies. A focus is placed on those that follow the sponsor route as these are underperforming schools gaining academy status via the same route as those in the treatment sample. Converters voluntarily gain academy status and are not comparable to the schools that are studied.

[45] There were some other cases, for example where schools that already had more autonomy than a typical state community school became an academy, or as a means for fee-charging independent schools to broaden their intake of pupils by becoming academies (Department for Children, Schools and Families, 2007), but as the numbers discussed below will show, these were the exception rather than the norm.

moves on to the feasibility stage and beyond that to actual conversion of the already existing school to an academy.

Figure 1 shows the evolution of school types between the school years 2001/02 and 2009/10. The change in the composition of schools is modest relative to the vast post-2010 expansion of the programme. In 2001/02 there were no academies; by 2009/10, 203 academies were in operation. By 2008/09 – the final school year in the sample to be analysed - there were 133 academies open and operating. These had a gradual introduction, with the first three opening in the 2002/3 school year, and then in the subsequent school years as follows: 2003/04 - 9; 2004/05 - 5; 2005/06 - 10; 2006/07 - 19; 2007/08 - 37; 2008/09 - 50.

Appendix Table A3.1 shows that (at least) one school from every secondary school type converted to become an academy prior to 2008/09. Because of the research design to be implemented, the focus is limited to schools that convert from an already-existing school; furthermore, the analysis is based upon schools that enrol pupils at age 11 and have students sit their final compulsory schooling exams at the school (this corresponds to the conventional secondary school in England). The final sample consists of 94 treatment schools drawn from the seven cohorts of schools opening prior to the 2008/09 school year and 114 control schools opening in the 2009/10 and 2010/11 school years.[46] These latter two cohorts consist solely of sponsored academies that were approved to become academies prior to the new regime that arose after the 2010 Academies Act.[47]

[46] The two main discrepancies between the sample the 133 and 94 are the removal of newly built academies, of which there were 12, and of 12 conversions from City technology colleges. The rationale for this latter omission is given below.

[47] For inclusion in the analysis, the approval of 'future' academies had to have taken place before May 2010, when the government changed, and the new coalition introduced the Academies Act.

3.2.3 Related Literature

Whilst there is a sizeable body of research on the impact of different schooling systems on pupil performance, there are fewer studies that look at what happens when the type of school attended by pupils changes. One study that looks at schools changing status in England is Clark (2009). He looks at what happened when schools became grant-maintained (GM) – a school type that enjoyed substantial operational autonomy.[48] He utilises a regression-discontinuity design since the decision to change status was decided by parental vote. As narrow GM vote winners experienced a significant improvement in pupil performance (of about a quarter of a standard deviation) compared to the narrow GM vote losers, his results suggest that increased school autonomy can bring about performance improvements.

GM schools were introduced in the late 1980s and conversion to GM status involved little turnover in management; indeed, the process was voluntary and often instigated by the school's governors. The granting of greater autonomy to already successful schools contrasts with the initial academies programme, where managerial changes were imposed on schools deemed to be struggling. In this respect, the US work on charter schools is more relevant to the analysis undertaken in this paper.[49]

Initial findings from the literature on charters, based upon quasi-experimental research designs, produced mixed to negative results. For instance, Betts et al. (2006) find that charters perform roughly at a similar level to public schools in the 16 charters they study in San Diego

[48] GM schools were renamed as foundation schools (see Table 1) in the Schools Act of 1998.
[49] Epple, Romano and Zimmer (2016) provide an in-depth and up-to-date survey of the work on charter schools.

while two studies carried out by CREDO (2009, 2013) find little average effects when looking at charters across 16 and 27 states.

Concerns with non-random selection into charters subsequently led researchers to begin to look at lottery based estimates of the effect of charter attendance. These studies exploit the fact that some schools use lotteries to allocate places when the school is oversubscribed. The vast majority of these papers find substantial positive test score gains for pupils "lotteried" in to charters relative to those "lotteried" out (see Abdulkadiroglu et al., 2011, Angrist et al., 2010, Angrist et al., 2013, Dobbie and Fryer, 2011, Dobbie and Fryer, 2013, and Hoxby, Murarka and Kang, 2009, for studies of test score gains; and Angrist et al., 2016, and Dobbie and Fryer, 2014, for evidence of students' longer-run outcomes, including college attendance).

An exception to the above is Gleason et al. (2010) who use lottery estimates from 36 charters across 15 states and find little evidence of improvements in pupil performance on average. However, they do find performance improvements for disadvantaged children (defined as those on free school meals). Similarly, Angrist, Pathak and Walters (2013) find that when splitting their Massachusetts sample between urban and non-urban charters, gains are positive in urban schools but negative for non-urban charters. As the majority of the lottery studies are based upon charters serving disadvantaged children in urban areas - such as New York and Boston - these latter studies shed light on seemingly disparate findings between lottery and non-lottery based studies.

Charters differ from academies in two important dimensions; firstly, charters are often newly built or set up schools; and secondly, applications to charters tend not to be co-ordinated with applications to other local schools.[50] Some recent studies appear pertinent to the case of English academies in these two dimensions. A small number of US studies have looked at conversions of already existing public schools to charters (as in the study of school takeovers in Boston and New Orleans by Abdulkadiroglu et al., 2016), as well as the introduction of practices used in charters to US public schools (as in Houston schools studied by Fryer, 2014). These report substantial improvements in test scores in those setting due to the use of methods of "best practice". Alongside these, Abdulkadiroglu et al. (2017) report lottery estimates for charters in Denver where, contrary to usual practice, places at charters and public schools are allocated using a common assignment scheme. In a school choice setting similar to the one studied here, they find positive effects of charters on performance.

On academies themselves, there remains little rigorous work. Very early work by Machin and Wilson (2008) looked at differences in pupil performance between a small sample of the first academy schools and a matched group of schools, finding modest, statistically insignificant, relative improvements. A PwC Report (2008) reported higher percentage point increases in the results of academies compared to the national average (which is not a good comparison since academies are well below average performers in their predecessor state), while a National Audit Office (2010) report on the Labour academies looked at their performance compared to a selected group of maintained schools, with similar pupil intakes and performance to the academies pre-treatment, finding a significant improvement in pupil performance in the academies. There is also some largely descriptive, non-causal school-level empirical work in the education field.[51]

---

[50] While academies can set their own admissions criteria so long as it accords with legislative guidelines, applications to state schools are co-ordinated at the local authority level. Compliance with local authority co-ordination of admission arrangements is part of an academy's funding agreement.

[51] See, for example, Gorard (2014) or West and Bailey (2013).

3.3. Data and Research Designs

3.3.1 Data

The main data source is the National Pupil Database (NPD).[52] The NPD is centrally collected census data containing pupil and school characteristics combined with the annual National Curriculum key stage attainment data at the pupil level. The Pupil Level Annual Census data (PLASC) contains information on characteristics of all pupils in the English maintained sector. This has been collected three times per year (January, May and September) from the 2001/02 school year onwards (though pupils can be traced back to earlier years of the key stage attainment data via their unique id). For this paper, only use the year-on-year January collection is used because this collection is the most available and consistent over time.

In England, compulsory education is organised around four key stages for years of schooling from ages 5 to 16. These are key stage 1 (in grades 1 and 2) and key stage 2 (grades 3 to 6) in primary school; and key stage 3 (grades 7 to 9) and key stage 4 (grades 10 and 11) in secondary school. In studying academy conversion impacts, the two outcomes of interest are pupil intake and pupil performance. To study intake for pupils enrolling in secondary school in grade 7, the first grade of secondary school, the focus is on the key stage test exams (KS2) that pupils take at the end of primary school (aged 10/11 at the end of grade 6) before they make the transition to secondary school. To study performance in grade 11, the final year of compulsory secondary schooling, the key stage 4 (KS4) examinations that pupils take at the end of compulsory schooling

[52] The use of pupil-level data throughout and a heavily refined research design are the key innovations compared to the version of this paper circulated earlier (Machin and Vernoit, 2011). Of course, use of pupil-level data (which the earlier version did not have full access to) makes the analysis more appropriate in that the right level of treatment is the effect of schools on the pupils that attend them compared to schools they would otherwise have attended. Put another way, changing pupil composition due to academy conversion because the demand for places alters compared to the predecessor school can render school-level estimates biased.

(aged 15/16 at the end of grade 11) are studied. These school leaving exams are known as GCSEs (standing for the General Certificate of Secondary Education).

The impact of academy conversion needs to be analysed at the pupil-level. This is because the underlying composition of students attending schools may change over time and, indeed, it turns out that pupil intake does change post-conversion. It is important to devise an empirical strategy that is not contaminated by the changing quality of post-conversion enrolees. A causal effect of academy attendance on pupil performance is therefore identified by focusing on pupils who were already enrolled in an academy pre-conversion. These pupils are referred to as being legacy enrolled. Because they had been enrolled in the school prior to conversion this avoids the endogeneity of the post-conversion enrolment decision that would contaminate estimates obtained from also looking at newly enrolled students.[53]

3.3.2 Comparison Schools

The research design combines difference-in-differences with instrumental variables. Before going into specific details, first the comparison schools are defined. Table 3.2 compares pre-treatment characteristics of academy schools and other types of maintained English secondary schools. The 106 academies (who have both a grade 7 intake and grade 11 exam takers) very clearly have significantly different pupil characteristics and levels of pupil performance than other state maintained secondary schools.

The fact that these schools show higher signs of disadvantage and record lower achievement in school leaving tests is not surprising as Labour's academy programme was aimed

[53] One further practical issue concerns the definition of schools that convert to academies. There are a small number of examples where multiple predecessor schools combine to create a single academy school. Where this occurs, create one hypothetical pre-academy school is created (see a fuller discussion in the Data Appendix). This adopts hypothetical characteristics that are a weighted-average of the characteristics of the merged schools.

at poorly performing schools. Thus, a naive comparison between academy schools and all other state-maintained schools is likely to suffer from significant selection bias. There is one exception here, as the 12 conversions from City Technology Colleges (CTCs) were already highly autonomous schools that were performing well. These are therefore omitted and the treatment group defined as the 94 new academies that converted from the four groups of state maintained schools: community, voluntary controlled, foundation and voluntary aided schools.[54] In addition to selection on observables, a related issue is that schools that go on to become academies may have common unobservable characteristics (e.g. they have a type of school ethos that is more in line with the academy model). Finally, there is also scope for mean reversion, as academies were badly performing schools in their predecessor state.

Panel 2 of Table 3.2 shows the pre-treatment characteristics of both the 94 schools that become academies in the sample period and 114 schools that become academies later after the study sample period ends. In contrast with the top panel, the characteristics of these two sets of schools appear balanced in the pre-treatment period; that is, for most of the variables considered (the exceptions being the proportion white for a 5 percent significance level and free school meal eligibility at the 10 percent level), one cannot reject the null hypothesis that the 94 academies that convert in the sample period and the 114 academies that convert in the following two school years have, on average, the same sets of characteristics.[55] This partially legitimises the use of pupils

[54] In fact, some commentators have identified CTCs as the precursors of academies (see West and Bailey, 2013). Almost all CTCs took up the opportunity to become academies when it arose with the introduction of academy schools. They were already highly autonomous schools already, being able to not fully follow the national curriculum, to run their own admissions, and not being maintained by the local authority. One can argue that the autonomy gains they experienced from academy conversion were negligible, unlike for the state-maintained schools that converted studied in this paper. The working version of this paper (Eyles and Machin, 2015) contains results that include city technology colleges – the results are largely unchanged by their omission.

[55] A test of joint significance was carried out by collapsing the data to school level and running a probit model of treatment status on all the variables. The null hypothesis that the variables were jointly insignificantly in predicting treatment status could not be rejected (the p-value testing joint insignificance was 0.153).

attending future converters as a control group in the D-i-D setting. It is further legitimised in the empirical findings described below where there is no sign of differential pre-conversion trends between treatment and control schools in test scores, thus allaying concerns of mean reversion.

3.3.3 Modelling Approach

To first study the issue of changing pupil composition post-conversion, a brief analysis of the impact of academy school conversion on pupil intake is first presented. For some of this analysis, intake is measured in terms of ability composition by the end of primary school standardised KS2 average points score[56] of pupils who enrol into grade 7, the first year of secondary school. Alongside this ethnicity, free school meal status, and gender of the incoming cohort are also considered.

For the main analysis – the impact of academies on pupil performance – the outcome of interest is the KS4 performance of pupils, measured for much of the analysis as the standardised best 8 exams points score of individual grade 11 students.[57] How robust the findings are to the use of different measures of pupil performance is also considered.

In terms of the timing of academy conversion, an academy is designated as starting for the first whole school year when it has academy status. One can define c as the number of academic years before or after conversion. In the intake analysis, the first treated individuals ($c = 0$) are those entering the academy in the September that it opens for business. For the performance sample, $c = 0$ refers to those sitting their GCSEs in the school in the following May/June i.e. those

[56] This is calculated by totaling (for each pupil) their raw scores in English, Maths and Science, then averaging across the three before standardising to have mean zero and standard deviation one.
[57] The precise measures used for KS2 and KS4 are described in detail in the Data Appendix, together with additional performance results for a range of different KS4 measures.

that sit their exams in the school's first academic year as an academy. Limiting the sample to pupils in schools that either convert or are set to convert after the sample period enables implementation of the treatment-control comparison across conversion cohorts that is described below.

3.3.4 Research Design

The main empirical question of interest is the impact of becoming an academy on end of secondary school examination performance. To clarify the research design it is useful to first introduce some notation. Let t denote academic year, which runs from 2001/02-2008/09, $g(i,t)$ denote the grade in which pupil i is enrolled in year t (this takes values 7-11 for secondary school pupils), and let $j(i,t)$ denote the school in which pupil i is enrolled in year t. The year of conversion for school j can be defined as CY_j. Finally, conversion cohorts are sets of schools – S_t – that convert in the same academic year t.

The focus on legacy enrolled pupils in the main analysis is initially justified by showing that the pupil composition of academy schools changed post conversion. This involved looking at treatment-control differences in several intake measures for the population of grade 7 pupils entering schools over the 2001/02 – 2008/09 period. In the following equation, for an intake measure for pupil i, who enrolled in grade 7 at school j in year t, the key parameter of interest is the differences-in-differences coefficient δ_1:

$$\text{Intake}_{it} = \alpha_j + \delta_1 \text{Academy}_{it} + T_t + u_{1it} \tag{1}$$

In (1), α_j and T_t respectively denote school and year fixed effects and u_{1it} is an error term. The key independent variable in the regression is defined as follows:

$$\text{Academy}_{it} = \begin{cases} 1 \text{ if } t \geq CY_{j(i,t)} \\ 0 \text{ if } t < CY_{j(i,t)} \end{cases} \tag{2}$$

The estimates of (1) reported below show that pupil intake did change after academisation. This change in composition means it would be misleading to study pupil performance effects for the children newly enrolling post-conversion. This is dealt with by using legacy enrolment as an instrument for academy attendance. The approach has similarities to that taken in Abdulkadiroglu et al. (2016), who study school takeovers in New Orleans, referring to pupils who stay in a converting school as "grand-fathered" pupils.

In the main pupil performance analysis, the focus is placed on grade 11 pupils in the school years 2001/02-2008/09. For these pupils the legacy enrolment instrument – Z_{it} – is defined as:

$$Z_{it} = \begin{cases} 1 \text{ if } j(i,t\text{-}1) \in S_t \text{ and } 11 > g(i,t\text{-}1) \geq 7 \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

In other words, pupils are legacy enrolled if they are enrolled in an academy in the year prior to conversion and are not in their final year of compulsory schooling (grade 11).

The empirical research design to study performance effects can now be operationalised by means of the following value-added equation, where *KS4* is end of secondary school examinations performance and *KS2* is end of primary school performance for pupil i in year t:

$$KS4_{it} = \beta_j + \theta_1 \text{Academy}_{it} + \pi_1 X_{it} + \varphi_1 KS2_{it} + T_t + v_{1it} \tag{4}$$

In (2), X is a vector of control variables and v_1 is an error term. Importantly, β_j is a school fixed effect measured at the date of legacy enrolment, rather than the date when KS4 exams are sat. As there are multiple conversion cohorts – from 2002/03-2008/09 – equation (4) is estimated separately for each conversion cohort, each time using pupils in schools that convert out of sample

– in 2009/10 and 2010/11 – as a control group. Estimates are pooled together across conversion cohorts in the presentation of the results.[58]

Ordinary least squares estimates of θ_1 from (4) may not reflect a causal estimate if individuals sort into academies post-conversion in a non-random way. Selection into and out of treatment is accounted for as follows:

i) From the point of conversion onwards, focus is placed on a fixed set of pupils who are legacy enrolled in the school; therefore, for conversion cohort S_t, the focus is on grade 11 pupils sitting their exams in schools $j \in S_t$ before academic year t (who form the 'before' group of pupils in the difference-in-differences analysis) alongside pupils that are legacy enrolled in these schools in t-1 (who form the 'after' group of pupils in the difference-in-differences analysis). Pupils enrolling in schools $j \in S_t$ after, or in, year t are removed from the sample.

ii) To make the treatment and control groups as comparable as possible, the same restrictions applied to the treatment group are applied to the control group; accordingly, when considering conversion cohort S_t, the control group comprises grade 11 pupils who sit their exams in control schools prior to t, and those enrolled in a control school in t-1, but who are not in their final year of compulsory schooling.[59] The composition of the treatment and control groups is shown in Appendix Table A3.2.

---

[58] As the estimates are pooled, the same pupils are used multiple times as controls. In almost all cases, multiple observations of control pupils occur within the same school. Standard errors are therefore clustered at school level.

[59] The rationale for restricting the control group in a similar fashion to the treatment group is to avoid conflating estimates with the effect of mobility on test scores. If we were to take as a control group all grade 11 students in 2009/10 and 2010/11 converters then, for conversion cohort S_t, control group pupils observed after t would be more mobile than legacy enrolled pupils observed in the same academic years. This restriction also allows harmonisation of the school fixed effects - β_j - across treatment and control schools. For conversion cohort S_t, β_j corresponds to the t-1 school for those sitting their grade 11 exams after, or in, t. For those sitting their KS4 exams in, or prior to, t-1, β_j, corresponds to the grade 11 school.

iii) The legacy enrolment variable - Z_{it} - is used as an instrument for $Academy_{it}$, to estimate a local average treatment effect (LATE). Because pupils selecting into schools after the point of conversion are removed from the sample, the LATE estimate corrects for the fact that not all legacy enrolled pupils remain in the academy until grade 11.

Formally, the local average treatment effect is obtained from estimates of the following two reduced forms:

$$Academy_{it} = \beta_j + \theta_2 Z_{it} + \pi_2 X_{it} + \varphi_2 KS2_{it} + T_t + v_{2it} \tag{5}$$

$$KS4_{it} = \beta_j + \theta_3 Z_{it} + \pi_3 X_{it} + \varphi_3 KS2_{it} + T_t + v_{3it} \tag{6}$$

In the first stage, equation (5), estimates of θ_2 measure the proportion of the legacy enrolled pupils that stay in the academy and take KS4 exams there. Equation (6) is the reduced form regression of KS4 on the instrument. The instrumental variable (IV) LATE estimate is then the ratio of the reduced form coefficient to the first stage coefficient, θ_3 / θ_2.

The specifications adopted so far impose an average post-conversion effect across all post-conversion years. A more flexible specification estimates separate treatment effects for pre- and post-conversion years in an event study setting. The IV setting already described can be extended to an event study framework where separate estimates are obtained for each of the four post-conversion years ($c = 0$ to $c = 3$) using four separate instruments, which are equivalent to dummies for the enrolment grade of a pupil (7-10) in the pre-conversion year. The event study structural form comparable to equation (4) becomes:

$$KS4_{it} = \beta_j + \sum_{c=-4}^{3} \theta_{4c}\, Academy_{it} \times 1\left[CY_{j(i,t)} = t-c\right] + \varphi_4 KS2_{it} + \pi_4 X_{it} + T_t + v_{4it} \tag{7}$$

In (5), $1[CY_{j(i,t)}=t\text{-}c]$ is an indicator for whether pupil *i* takes their KS4 exams in a treatment school *c* years before/after conversion, with conversion taking place at $c = 0$. Therefore, event study estimates of four pre-conversion θ_4's (from $c = -4$ to $c = -1$) and four conversion year and post-conversion θ_4*'s* (from $c = 0$ to $c = 3$) can be obtained. The former are informative about differential pre-conversion trends. All time periods 5 or more years before conversion comprise the omitted category which is set to zero.

3.4. Empirical Results

3.4.1 Academies and Pupil Intake

Table 3.3 shows results from the analysis of changing composition for grade 7 enrolees at the start of secondary school. The Table reports differences-in-differences estimates based on equation (1), with the following four dependent variables: the end of primary school KS2 test score, and dummy variable indicators of free school meal status, being of white ethnic origin, and being male. In each case, and in all that follows, standard errors are clustered at the school level.

The estimated coefficients in Table 3.4 show that academies, post-conversion, did alter their intake in a number of dimensions. They were less likely to admit free school meal eligible pupils, and they admitted pupils with significantly higher KS2 scores. Column (1) shows that, on average, pupils enrolling in an academy at grade 7 have a KS2 mean points score that is 0.09σ higher than those attending schools yet to attain academy status. Column (2) shows a 3.3 percentage point fall in the number of free school meal pupils entering academies post conversion.

By contrast, the gender and ethnic composition of their intake appear unchanged by a school becoming an academy.[60]

These results are shown simply to make clear the need to study the legacy enrolment cohorts in the main pupil performance analysis. They indicate that the composition of newly enrolled children, beginning their secondary school years, did change differentially in treatment versus control schools before and after academisation. Hence, for the pupil performance analysis that comes next, to avoid biases from changing composition it is necessary to focus on legacy enrolled pupils.

3.4.2 Academies and Pupil Performance

The first set of results from the analysis of the main question of interest – the impact of academies on pupil performance - are reported in Table 3.4. It shows OLS, reduced form and IV estimates of the impact of academy conversion on end of secondary school Key Stage 4 pupil performance for grade 11 children. Columns (1) to (3) show estimates of the impact of academy conversion on pupil performance from specifications without control variables. Columns (4) to (6) show estimates from value added specifications that net out end of primary school pupil performance and other pupil characteristics. Columns (1) to (3) show that being in an academy school increases pupil's KS4 test scores by a statistically significant 0.12σ to 0.20σ. Adding the prior achievement measure (KS2) and control variables in columns (4) to (6) reduces this by a very small amount to a range of 0.12σ to 0.18σ, with all estimates remaining strongly significant.

[60] It is noteworthy that, while academies gain freedom to handle their own admissions, they remain subject to the same statutory rules as other state schools and operate under a common admissions regime. In particular, unless they are oversubscribed they must admit all children who apply and – in the case of oversubscription – cannot discriminate on any of the outcomes studied. These intake results are therefore indicative of a post-conversion change in preferences of the local community, rather than a change in recruitment practices of the schools.

The interpretation of the legacy enrolment estimate in column (5) is of a 0.11σ higher KS4 score for children enrolled in a pre-conversion school as compared to children enrolled in control schools in the same school years. The IV estimate in column (6) corrects for the fact that not all legacy enrolled children sat their KS4 examinations in the school. In fact, the vast majority - 93.8%[61] - did as the highly significant first stage at the bottom of the Table shows. Because of the high rate of compliance, the IV estimate rises only a touch compared to the reduced form estimate, increasing to 0.12σ. This is the preferred baseline average causal estimate of academy conversion.

Aside from the fact that pupil achievement is significantly higher on average for pupils attending schools that converted to an academy, a further point stands out from the results shown in the Table - the estimates are similar regardless of estimation method and the set of control variables used. This reflects two aspects of the data; first, the treatment and control pupils are well balanced in terms of covariates, including end of primary school KS2 test scores; and second, there is a high rate of compliance for legacy enrolled students. Because of this, the reduced form and IV estimates broadly align with each other.

Figure 3.2 plots the event study D-i-D estimates from the IV specification including control variables.[62] Importantly, there are seen to be no discernible pre-treatment trends in the outcome variable. However, there is a significant positive, and rising over time, impact after conversion. Conversion year test scores are 0.06σ higher (though statistically insignificant) at c = 0, the conversion year, and rise to (a statistically significant) 0.28σ four years post-conversion.

[61] The implied degree of pupil mobility in the secondary school years from this 93.8% (or 6.2% moving) lines up well with pupil mobility numbers for English schools described in Machin, Telhaj and Wilson (2006).
[62] The full set of event study estimates are shown in Table A3.3 of the Appendix.

The results reported previously are pooled versions of difference-in-difference estimates from different cohorts of academy conversions occurring in the school years 2002/03 through 2008/09. Figure 3.3 plots IV estimates from the models separately by conversion cohort.[63] It is very clear that a null hypothesis of the same average effects across cohorts is not rejected by the data. The gradually rising positive performance effects are seen across the three cohort groups of conversions shown in the Figure. Furthermore, the lack of differential pre-treatment trends for all cohorts is highly supportive of the research design that is implemented.

For pupils that attend academies four years after conversion, these impacts of academisation are quite large. To contextualise this, it is worth comparing them with some of the results found in the US charter school literature. Dobbie and Fryer (2013) exploit lottery admission in New York City charters and find gains of 0.13σ/0.05σ in math/ELA tests scores for middle school aged students. The research designs most similar to our own (Fryer, 2014, and Abdulkadiroglu et al, 2016) - in that they either inject charter practices into pre-existing schools or focus on pupils 'grandfathered' into takeover schools – report estimates of between 0.15σ and 0.32σ for math middle and high school students and insignificant to 0.39σ for ELA students. Although these correspond to a slightly different age range, and a broader measure of achievement – to reflect to the nature of KS4 exams – is used, the results fall in line with the high achieving charter school findings.

3.4.3 Heterogeneous effects

[63] The breakdown by cohort is: 11 conversions in school years 2002/03 to 2004/05 (3 from 2002/03, 6 from 2003/04 and 2 from 2004/05), 21 conversions from school years 2005/06 and 2006/07 (7 from 2005/06 and 14 from 2006/07) and 62 conversions from 2007/08 and 2008/09 (25 from 2007/08 and 37 from 2008/09). In each case they are compared to the control group of 114 schools that convert after the study sample period ends.

While there are positive estimates of performance effects for pupils attending academies, it is possible that substantial heterogeneity is obscured by a focus on average effects. For instance, the charter school literature tends to find positive effects for disadvantaged pupils in urban charters, but little (and sometimes even negative) effects in non-urban settings. Similarly, as noted earlier in the paper, treatment intensity varies in the setting studied in this paper: schools that become academies from community schools gain much greater freedom than those converting from religious schools or foundation schools.

Table 3.5 shows IV estimates which allow the treatment effect to vary by predecessor school type and whether the school is in an urban area. The estimates in columns (1) to (4) show that while the average effect for academy attendance is positive, effects appear larger for pupils who are pre-enrolled in community schools and those who attend urban academies. The sizeable effect of 0.14σ for community converters and 0.11σ for urban academies contrast with the insignificant coefficients for pupils in non-urban schools and in non-community predecessors.

3.4.4 Robustness: Pre-trends, choice of control group and outcome measures

Various robustness checks were also undertaken with an intention of testing the sensitivity of the main results to various assumptions. These checks are motivated by the following possible concerns; there might be school specific pre-trends in outcomes; results may depend on the choice of control group; and results may be sensitive to the KS4 outcome used to measure end of secondary school performance.

To address the first concern, specifications that control for a set of school specific pre-conversion linear trends were also estimated. To aid in the precision of these estimates, data going back to 1997/98 for which there are data on KS4 scores, but no pupil level covariates, was added.

For brevity's sake, only the result of the main estimate is reported here: the equivalent of the IV estimate in column (6) of Table 3.4. Adding school specific trends shifts the main estimate by a small amount up to 0.13σ (0.03), which remains statistically significant. Thus, it does not appear to be the case that school specific pre-trends in test scores explain the results.

Secondly, it is worth noting that while consideration of the control group adds power to the estimates, it is possible to dispense with the control group altogether and rely only on time variation in the receipt of treatment to identify performance effects. Once again, if the main pooled IV estimate is obtained with only the 94 treatment schools results are very similar – the average effect of academy conversion in this case is 0.09σ (0.03).

In terms of different performance measures two sets of additional tests were undertaken. The first looked at other measures of KS4 performance. These results are shown in Appendix Table A3.4, where the following alternative measures of pupil performance were considered: GCSE Math, GCSE English, and 5 A*-C GCSEs including English and Math. Their use as dependent variable reveals a very similar pattern of results to those using the main KS4 points score.

Finally, a different measure of whether academisation under the Labour programme resulted in improved school performance was considered. This looked at Ofsted inspections of schools before and after conversion, again relative to control schools.[64] The probability of transitioning between Ofsted grades was set up as a function of becoming an academy. Transitioning constitutes movements in inspection rankings (of outstanding, good, satisfactory, or

[64] Ofsted is the Office for Standards in Education, Children's Services and Skills which is a government department of Her Majesty's Chief Inspector of Schools in England which undertakes inspections of schools as part of the strongly enforced school accountability system that operates in England.

inadequate) before and after academy conversion for academies in the early and late 2000s and the same for comparison schools. Not all schools were inspected twice in this period, so the analysis is confined to the sub-set of schools that were inspected twice. Ordered probit estimates reported in Table 3.6 show a statistically significant improvement in inspection rankings of academies. This act as complementary and corroborative evidence in line with the KS4 performance gains already reported.

3.5 Mechanisms

The above results uncovered evidence of significant performance improvements for pupils treated by academy conversion. They also showed these improvements to be more pronounced for those attending schools that gained the greatest autonomy. We now address the question - what use of academy freedoms can account for these findings? It must be acknowledged that the analysis here is somewhat limited in what can be done with available data, but it is possible to offer three main sources of evidence on the question of mechanisms that may be at play. The first from comes from survey data on academies, the second on data on changes in head teachers and teaching staff before and after conversion, the third considering whether peer effects may have played a role.

3.5.1 Academies survey

To begin this discussion of mechanisms, we first draw on the Department for Education's (2014) survey of academy schools 'Do Academies Make Use of Their Autonomy? This survey collected information on a wide array of changes that may have occurred following conversion.[65]

[65] In May 2013 the Department for Education sent a questionnaire to all 2919 open academies. Of the 720 respondents, 148 were sponsored academies, with 74 of these being secondary schools. Of the 74, 23 converted pre-May 2010 and thus were academies at some point in the sample period.

These are summarised in Table 3.7 for 20 (and 3 comparable schools on which there is not full data) of the Labour academies analysed in this paper, and for 148 academies (including the 23 Labour academies) overall.

Table 3.7 ranks the responses in order of the percent making the particular change considered in the survey. The three most prominent changes, amongst the 23 converters in the sample, were ‘changed school leadership’, ‘procured services that were previously provided by the local authority’ and ‘changed the curriculum you offer’. Over 75% of the schools said they made these changes pursuant to gaining the new academy freedoms. This ranking is broadly consistent with that of the 148 sponsored academies overall.

When asked what the most important change was, two answers dominate - ‘changed school leadership’ (at 56%) and ‘changed the curriculum you offer’ (at 26%). Furthermore, both of these were reported to be linked to improved outcomes (in 73% and 77% of cases respectively). Other changes that were notably linked to improved outcomes were ‘Increased the length of the school day’ (63%) and ‘Collaborated with other schools in more formalised partnerships’ (45%).

3.5.2 Changes in Head teacher and Teaching Staff

It is also possible to look at statistical difference-in-differences estimates at school-level for three of the important factors identified in Table 3.8: whether a new head teacher is taken on upon conversion; whether more pupils are enrolled; and whether more teachers are taken on. This is facilitated by the availability of school level data over time on each of these.

Column (1) of Table 3.8 reports results for head teacher change. There is significant head teacher turnover before and after conversion to an academy. Over all post conversion years this averages out to 21% more head teacher turnover in academies. Further probing makes it clear that this substantial degree of turnover is very much concentrated in the conversion year. In treatment schools, 45% more head teacher turnover occurred in the year of conversion c as compared to the control schools. This is shown in event study estimates equivalent to Table 3.8 in Appendix Table A3.5. It seems to be a one-off change that occurred as the subsequent year treatment effects from (c=1) to (c=3) were all insignificantly different from zero.

Thus, a strong feature of academy conversions is to replace the head teacher. There is a more modest turnaround of the rank-and-file teaching staff and, if there is an increase, it seems to be due to a need to take on more teachers if more pupils enrol in academies post conversion. This can be gleaned from looking at columns (2) through to (4) in Table 3.8. The Table shows D-i-D estimates of the effect of academy conversion on the number of teachers, number of pupils and the teacher-pupil ratio. Looking first at column (2), there is evidence that the number of teachers rose once the school gained academy status. This is because, as shown in column (3), more pupils were enrolled once the academies were up and running, although this change itself is statistically insignificant. Finally, column (4) shows that the increased number of teachers went hand in hand with increases in pupil enrolments as the teacher-pupil ratio did not rise significantly post-conversion. Overall, the Table shows far less clear evidence of post-conversion teacher turnover, certainly when compared to the very significant evidence of head teacher turnover that occurred upon academy conversion.[66]

[66] Moving beyond this, we also re-estimated the main IV pupil performance specification with an interaction of treatment with the extent of teacher turnover. We found no evidence that schools who exhibited large changes in the numbers of pupils, teachers, or the ratio of the two generate larger gains for their pupils. The same exercise, but with an interaction for head teacher change, shows that those who do change head teacher generate slightly higher effects, but that the difference was not statistically significant. Of course, we cannot estimate the contribution of each mechanism separately without having separate "mechanism experiments".

3.5.3 Peer Effects

So far changes in staff, and changes enacted by management, have been considered as potential sources of performance improvements. A final potential mechanism is whether an increase in peer quality resulted from academy conversion. Earlier analysis has already shown that academy conversion resulted in an upward shift in the test scores of grade 7 entrants into academies. It also resulted in a downward shift in the proportion of free school meal eligible pupils entering the schools.

As well as entrants into grade 7, schools can also enrol students in other grades; therefore, new students do enrol in the same grades as legacy enrolled pupils in the years following conversion. There are not that many such students, but if a comparison of the attributes of those joining the *same* grades as the legacy enrolled pupils reveals them to be similar in terms of gender and FSM eligibility and that they are more likely to be non-white and, if anything, have lower prior test scores.[67]

Table 3.9 presents a more formal test of whether peer effects that could result from this post-conversion entry display a connection with the observed performance improvements. It does so by estimating a specification comparable to column (6) of Table 3.4, but with inclusion of an additional interaction between the treatment variable and the average standardised KS2 score of incoming pupils.[68] These averages can be separately calculated for two sets of peers, namely for incoming pupils into the same grade as the legacy enrolled pupils (as one would expect peer

[67] It is also worth noting that inflows of pupils in grades other than grade 7 are small. In the sample, of all those with a KS4 record in an academy post conversion, fewer than 10% were not legacy enrolled.
[68] To be precise, treatment status is interacted with measures of compositional changes in the treatment school and instrumented by legacy enrolment status interacted with the same measure of compositional change measured at the legacy enrolment school.

effects to operate primarily within grade), and for all pupils entering the schools post conversion (which would require cross-grade peer effects to operate).

Peer effects do not seem to account for the main pattern of results. As can be seen from columns (2) and (4) of Table 3.9, the interactions between treatment status and KS2 scores are small and insignificant. This is irrespective of whether KS2 scores are averaged over all post-conversion enrolees or only those that sit their grade 11 exams with the legacy enrolled pupils. These empirical tests imply that the estimated performance improvements do not seem to come about because performance effects and changes in peer composition are related.

3.6 Conclusions

Whether new school types can potentially alleviate poor education standards has become a question of significant interest to educators, policymakers and parents. This paper focuses on a school reform that has become a high profile in this regard – the introduction of academy schools to the English secondary school sector. The impact of academy school conversion on pupil performance is studied, using a legacy enrolment methodology free of bias from changing pupil composition. Academy conversion is seen to generate significant improvements in pupil performance for those who attended schools treated by academy conversion.

Transformation to an academy raised end of secondary school educational outcomes by 0.12σ on average, and by more for children receiving more years of treatment, rising to 0.28σ three years post-conversion. These findings complement existing work from different settings like that on US charter schools (both newly set up and more closely to takeovers of public schools) on whether different school types can affect pupil performance. They also add significantly to this literature as many of the best identified studies of US charters are often focussed on a single city or state setting. The national scope of the effort studied in this paper makes the findings of the paper less likely to be driven by context-specific factors than some of that research. As well as finding larger gains for pupils who spend more time in an academy, the paper also reports suggestive evidence that schools in urban areas, and those that gain the most autonomy from conversion, are the most likely to benefit from the program. Finally, there is not any evidence that improvements in peer composition drive the results, suggesting that the programme effects can, at least to some extent, be scaled up.

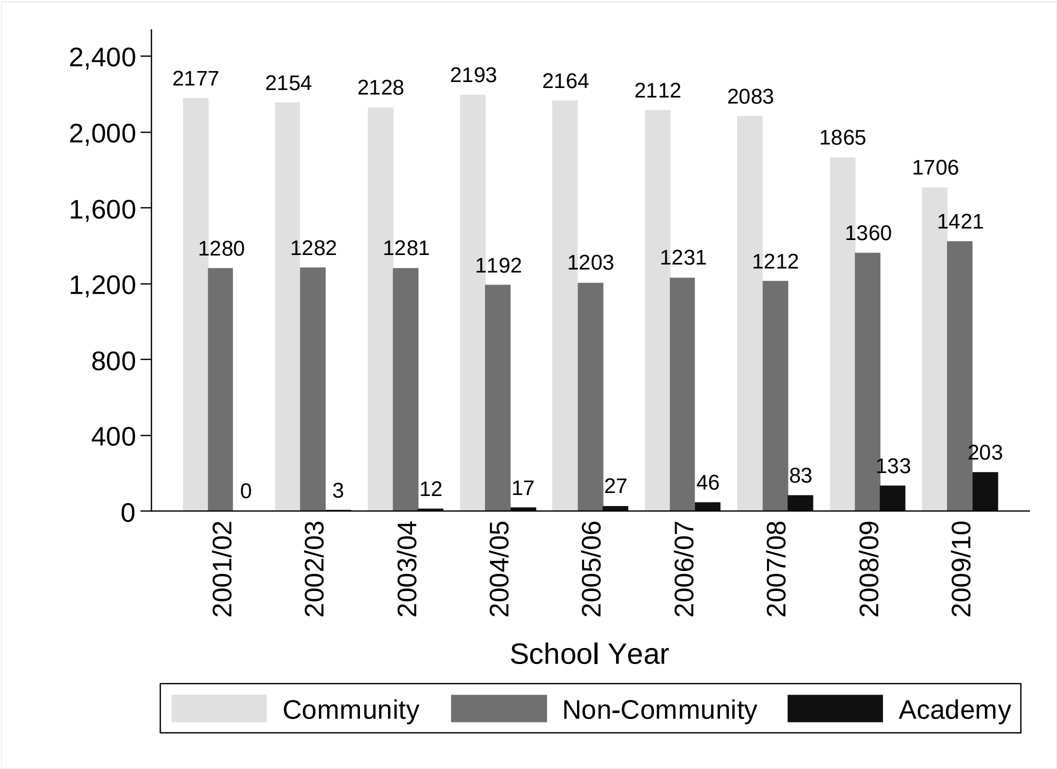
Before finishing, it is appropriate to place these findings into their policy context, especially given the large and rapid education reforms that have occurred recently in England. This paper studies the sponsored academies set up under the Labour government's programme, which had 203 academies up and running in May 2010 when a new coalition government was voted in. Since then, the academies programme has been massively expanded and taken on a new direction, with the number of conversions skyrocketing and with new convertors not only being in the secondary sector, but also covering primary schools, and even reaching outside the state sector to some private schools. Moreover, the new coalition academies need not have a sponsor when they are converted. In the 2010s, mass academisation has become the order of the day in English education.

A key feature distinguishing these new coalition academies is that, on average, they are not characterised by poor performance and disadvantage in their predecessor state like the sponsored academies introduced and approved under the previous Labour government analysed in this paper.[69] The way some of them are run is also different with, for example, some of the post

[69] See Eyles, Machin and Silva (2017) for an empirical analysis of the different nature of pre- and post-May 2010 academies.

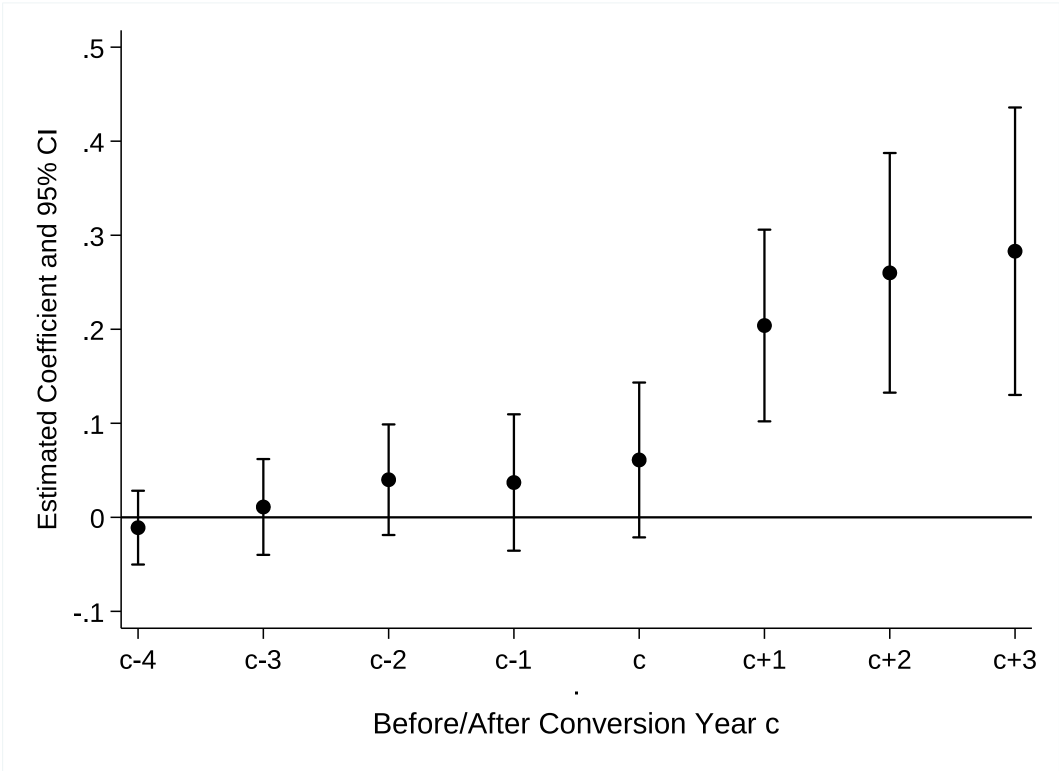
May 2010 academies being run as chains of schools by major sponsors. It will be an important future research challenge to determine whether these new convertor and chain run academies are able to deliver the kinds of performance improvements for students enrolling in them that the Labour programme analysed here seemed to deliver.

Figure 3.1: Secondary School Types in England, 2001/02 to 2009/10



Notes: Author's calculations using Department for Education data. Includes middle schools. Excludes special schools. Non-Community schools are city technology colleges, foundation schools, voluntary aided schools, and voluntary controlled schools. See Table 3.1 for more detail on these school types.

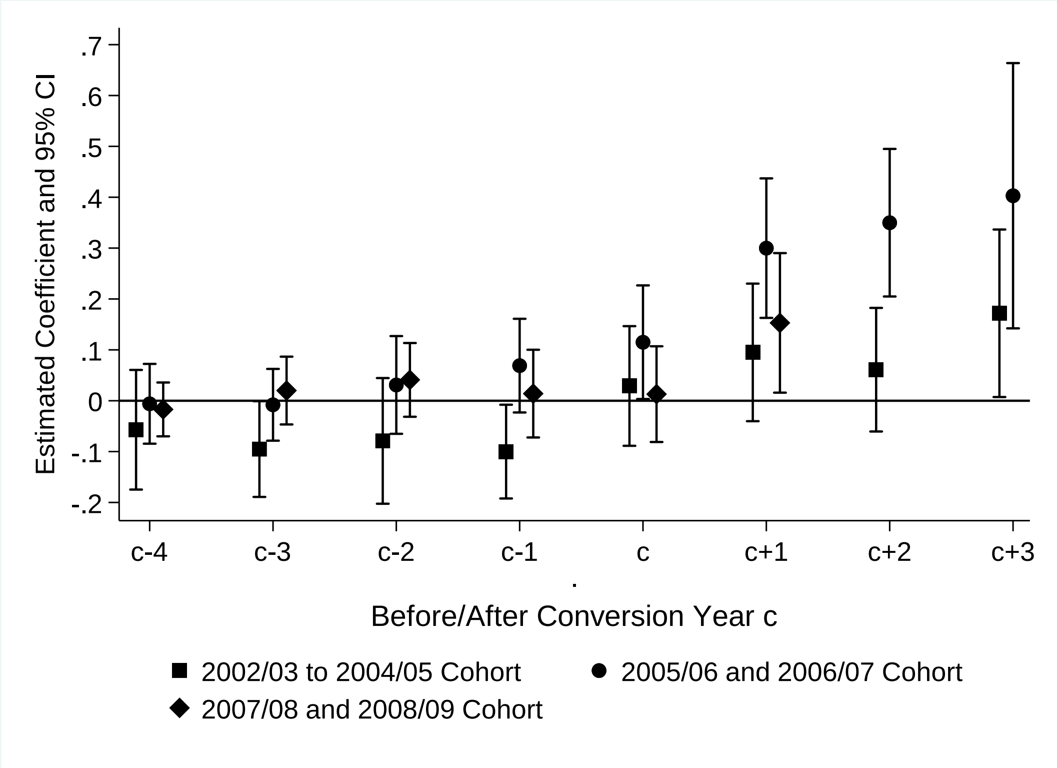
Figure 3.2: Event Study Instrumental Variable Estimates of Pupil Performance and Academy Conversion, End of Secondary School Test Scores, Grade 11 Pupils



Notes: Event study estimates, from specification in column (1) of Table A3 in the Appendix. The outcome measure is the best 8 capped point score.

Figure 3.3: Event Study Instrumental Variable Estimates of Pupil Performance and Academy Conversion,

End of Secondary School Test Scores, Grade 11 Pupils, By Groups of Academy Conversion Cohorts



Notes: Event study estimates, from cohort specific specifications comparable to column (1) of Table A3.3. The outcome measure is the best 8 capped point score.

Table 3.1: Characteristics of Autonomy and Governance in English Secondary Schools

	Non-LA Admission Authority	Maintained by Non-LA body	Not obliged to follow National Curriculum	Fee Charging
Registered independent school[a]	✓	✓	✓	✓
Academy[b]	✓	✓	✓	×
City technology college[c]	✓	✓	✓	×
Voluntary-aided[d]	✓	×	×	×
Foundation[e]	✓	×	×	×
Voluntary-controlled[f]	×	×	×	×
Community[g]	×	×	×	×

Notes:

a - Registered independent schools are independent of the local authority (LA), and are fee-charging.

b - Academy schools (prior to 2010/11): all ability independent specialist schools, which do not charge fees, and are not maintained by the local authority; established by sponsors from business, faith, HE institutions or voluntary groups, working in partnership with central government. Sponsors and the DfE provide the capital costs for the Academy. Running costs are met by the DfE in accordance with the number of pupils, at a similar level to that provided by local authorities for maintained schools serving similar catchment areas.

c - City Technology Colleges: all ability independent schools, which do not charge fees, and are not maintained by the local education authority. Their curriculum has a particular focus on science and technology education (see West and Bailey, 2013). They were established by sponsors from business, faith or voluntary groups. Sponsors and the DfE provided the capital costs for the CTC. Running costs are met by the DfE in accordance with the number of pupils, at a similar level to that provided by local authorities for maintained schools serving similar catchment areas.

d – Voluntary-aided schools are maintained by the local authority. The foundation (generally religious) appoints most of the governing body. The governing body is responsible for admissions and employing the school staff. Land at voluntary-aided schools is usually owned by trustees, although the local authority often owns any playing field land (Department for Schools, Children and Families, 2008).

e - Foundation (formerly grant-maintained) schools are maintained by the local authority. The governing body is responsible for admissions, employing the school staff, and either the foundation or the governing body owns the school's land and buildings (Department for Schools, Children and Families, 2008).

f – Voluntary-controlled schools are maintained by the local authority. These are mostly religious schools where the local authority continues to be the admission authority. Land at voluntary-controlled schools is usually owned by trustees (Department for Schools, Children and Families, 2008, 2013).

g - Community schools are maintained by the local authority. The local authority is responsible for admissions, employing the school staff, and it also owns the school's land and buildings.

Table 3.2: Pre-Conversion Characteristics and Tests of Balancing

	End of Primary Test Score (KS2) (mean)	End of Secondary Test Score (KS4) (mean)	Proportion getting 5 or more A*-C GCSEs or equivalents (mean)	Proportion male	Proportion white	Proportion eligible for free school meals	Number of Schools
A. All Schools							
City technology college	74.786	57.804	0.934	0.487	0.968	0.095	2
Voluntary aided	66.763	43.323	0.578	0.505	0.798	0.126	502
Foundation	65.516	43.340	0.573	0.522	0.85	0.092	470
Voluntary controlled	66.827	43.515	0.579	0.510	0.876	0.077	96
Community	61.983	38.312	0.460	0.503	0.828	0.153	1933
Academies (Pre-conversion)	57.230	31.689	0.316	0.536	0.725	0.250	106
B. Academy Schools							
Current academies (treatment group)	55.408	29.619	0.267	0.536	0.730	0.264	94
Future academies (control group)	56.476	30.912	0.285	0.515	0.812	0.232	114
Difference	-1.068 (0.796)	-1.293 (0.834)	-0.018 (0.018)	0.021 (0.015)	-0.082** (0.040)	0.032* (0.019)	

Notes: Standard errors clustered at school level reported in parentheses. Both panels refer to characteristics in the 2001/02 school year. Panel A is maintained schools in England, which do not convert to academies prior to, or in, the academic year 2008/09. All variables with the exception of KS2 point score, which refers to characteristics of the incoming 2001/02 cohort i.e. incoming pupils in the school year 2001/02, before any academies had opened, refer to those in their final school grade. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 3.3: Pupil Intake, Various Measures, Enrolled in Grade 7, 2001/02 to 2008/09

	End of Primary KS2 Test Score	Free School Meals	White Ethnicity	Male
	(1)	(2)	(3)	(4)
Enrols in Academy in Grade 7	0.093***	-0.033***	-0.015	-0.008
	(0.023)	(0.009)	(0.014)	(0.010)
School Fixed Effects	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
Sample Size	1321157	1321157	1321157	1321157
Number of Treatment and Control Schools	208	208	208	208

Notes: Robust standard errors (clustered at the school level) are reported in parentheses. The outcome in column 1 (the End of Primary KS2 Test Score) is calculated by totalling (for each pupil) their raw scores in English, Math and Science. It is then averaged across the three before standardising to have mean zero and standard deviation one – see the Data Appendix for precise definitions. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 3.4: Pupil Performance, End of Secondary School Test Scores, Grade 11, 2001/02 to 2008/09

	Standardised End of Secondary KS4 Test Scores					
	OLS	Legacy Enrolment	IV	OLS	Legacy Enrolment	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Takes KS4 in Academy	0.195***	0.111***	0.118***	0.182***	0.108***	0.115***
	(0.029)	(0.028)	(0.030)	(0.029)	(0.029)	(0.031)
Standardised End of Primary KS2 Test Score				0.599***	0.599***	0.599***
				(0.005)	(0.005)	(0.005)
Control Variables	No	No	No	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	1263751	1263751	1263751	1263751	1263751	1263751
Number of Treatment and Control Schools	208	208	208	208	208	208
First Stage Coefficient on Legacy Enrolment			0.938***			0.938***
			(0.003)			(0.003)

Notes: Robust standard errors (clustered at the school level) are reported in parentheses. Control variables included are dummies for whether the pupil is male, the pupil's ethnicity group, and whether they are eligible for free school meals, entered together with end of primary school KS2 test scores and a dummy variable for pupils for whom KS2 data is unavailable. The dependent variable is the standardised best 8 examinations point score of the pupil – see the Data Appendix for precise definitions. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 3.5: Pupil Performance, End of Secondary School Test Scores, Grade 11, 2001/02 to 2008/09, Heterogeneous Effects

	Standardised End of Secondary KS4 Test Scores			
	Pupils in Community Predecessor School	Pupils in Non-Community Predecessor School	Pupils in Urban Schools	Pupils in Non-Urban Schools
	IV	IV	IV	IV
	(1)	(2)	(3)	(4)
Takes KS4 in Academy	0.140***	0.069	0.113***	0.061
	(0.034)	(0.064)	(0.034)	(0.048)
Standardised End of Primary KS2 Test Score	0.603***	0.582**	0.591**	0.625***
	(0.005)	(0.012)	(0.006)	(0.009)
Control Variables	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
Sample Size	981249	282502	974434	289317
Number of Treatment and Control Schools	158	50	170	38
First Stage Coefficient on Legacy Enrolment	0.935***	0.945***	0.934***	0.963***
	(0.004)	(0.006)	(0.003)	(0.007)

Notes: As for Table 3.4.

Table 3.6: Ordered Probit Estimates of Change in Ofsted Ranking, School Level

	Pr[Change in OfSted Ranking]	
	(1)	(2)
Academy	0.865***	0.825***
	(0.241)	(0.241)
Control Variables	No	Yes
Number of Treatment and Control Schools	155	155
Marginal Effects:		
Pr[Change = 2||Treatment=1] – Pr[Change=2|Treatment=0]	0.329***	0.314**
	(0.091)	(0.102)
Pr[Change = 1||Treatment=1] – Pr[Change=1|Treatment=0]	-0.098**	-0.092*
	(0.048)	(0.047)
Pr[Change = 0||Treatment=1] – Pr[Change=0|Treatment=0]	-0.231***	-0.223***
	(0.049)	(0.050)

Notes: Ofsted is a non-ministerial department of the government that inspects English schools and gives them an overall effectiveness rating (on a four point scale) based upon, amongst other things, teaching quality, leadership effectiveness, pupil outcomes, and personal development. A more thorough discussion of Ofsted is given in Section 5 of the appendix. The dependent variable is coded as 0 for a reduction in Ofsted rating, 1 for no change and 2 for an improvement. Robust standard errors in parentheses. The control variables included in specification (2) are proportion male, proportion white, and proportion of pupils eligible for free school meals – all of which are measured in the year of first inspection. The above is estimated on a subsample of treated schools for whom an Ofsted rating is observed before and after conversion; for control schools, all Ofsted inspections over the period 2000/01 to 2009/10 are used. Year of inspection dummies are included in all specifications. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 3.7: Department of Education Survey of Changes After Academy Conversion,

23 Labour Academies and 148 Sponsored Academies

	23 Labour Academies	148 Sponsored Academies Including the 23 Labour Academies	148 Sponsored Academies Including the 23 Labour Academies	148 Sponsored Academies Including the 23 Labour Academies
	% Making Change	% Making Change	% Say Most Important Change	% Making Change Say Linked to Improved Attainment
Changed school leadership	87	72	56	73
Procured services that were previously provided by the LA	78	83	5	17
Changed the curriculum you offer	74	61	26	77
Changed the performance management system for teachers	74	70	3	39
Collaborated with other schools in more formalised partnerships	70	68	8	45
Introduced savings in back-office functions	70	55	0	12
Added non-teaching positions	70	50	3	31
Reconstituted your governing body	65	76	0	26
Changed your pattern of capital expenditure	65	54	1	19
Increased the number of pupils on roll	61	41	0	12
Hired teachers without qualified teacher status (QTS)	48	24	0	14
Introduced or increased revenue-generating activities	48	34	0	8
Changed your admission criteria	43	20	0	7
Increased the length of the school day	39	18	0	63
Changed staff pay structures	30	24	0	9
Sought to attract pupils from a different geographical area	13	12	0	11
Changed the length of school terms	9	6	0	22
Reduced the number of pupils on roll	4	3	0	0

Notes: Taken from Department for Education (2014). The 23 labour academies comprise of 20 schools in the sample and 3 schools that are excluded due to incomplete data.

Table 3.8: Change in Staff and Pupils Before and After Academy Conversion

	Change in Head teacher	Log(Number of Teachers)	Log(Number of Pupils)	Log(Teachers Per Pupil)
	(1)	(2)	(3)	(4)
Academy	0.214***	0.067*	0.045	0.022
	(0.047)	(0.039)	(0.035)	(0.018)
School Fixed Effects	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
Sample Size	1641	1641	1641	1641
Number of Treatment and Control Schools	208	208	208	208

Notes: Robust standard errors (clustered at the school level) are reported in parentheses. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. Columns (2), (3), and (4) control for whether the schools, in each year, enrol pupils prior to grade 7 and post grade 11.

Table 3.9: Change in Peer Quality After Academy Conversion

	Entrants into Same Grade	Entrants into All Grades
	(1)	(2)
Academy	0.126*** (0.035)	0.124* (0.064)
Academy* Average KS2 of New Entrants Post Conversion	0.030 (0.072)	0.027 (0.134)
School Fixed Effects	Yes	Yes
Year Dummies	Yes	Yes
Sample Size	1263751	1263751
Number of Treatment and Control Schools	208	208

Notes: Robust standard errors (clustered at the school level) are reported in parentheses. The estimates here are equivalent to those in column 6 of Table 3.4. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Appendix

1. Data on Academy Schools

We first identified all schools that became academies over the school years 2002/03 to 2010/11. Sources for this are Department for Education extracts that give information on all academies that have opened or are in the process of doing so. The extract gives the opening date of the academy, its URN (a unique identifier for the school allowing us to identify it in various governmental data sources such as the National Pupil Database and the Pupil Level Annual Census data), DFE number (a second unique identifier combining school specific and local authority specific numbers) and the URN number of the predecessor school.

Using performance tables data from the Department for Education (DfE) we match in predecessor school types. This gives 244 schools that became academies between the first 3 academy openings in 2002/03 and those that gained academy status by September 2010 (the beginning of the academic school year). Previously independent schools were omitted due to pupils in these schools not having exam information at KS4. Similarly, brand new academy schools were omitted as they have no predecessor school.

In order to have a balanced panel we focus on academies that have some form of predecessor school open from at least 1997 onwards. Any later and the school will not have KS4 results for 2002. In order for the sample to be balanced for intake we exclude academies who do not enrol pupils in grade 7. The final sample contains 106 (of which we use the 94 who were not CTCs in their predecessor state) treatment schools (those that opened as academies prior to, or in, September 2008) and 114 control schools with observations ranging over the years 2001/02-2008/09. None of the control schools become academies during these sample years, but convert by September 2010.

2. Pupil Level Data

We use data from PLASC (pupil level annual schools census) and the NPD (national pupil database). The NPD contains information on all key stage 2 (KS2) and key stage 4 (KS4) exams sat at the end of primary and secondary school respectively. Each pupil is identified by a unique reference number and the data gives the unique URN of the school in which they sat the exam. While the NPD reports on pupils in examination years PLASC has a record for every pupil for each year that they are in the maintained school sector. PLASC data gives the pupil, grade and school as well as demographic variables such as ethnicity, gender, and free school meal eligibility. We can track pupils through secondary school using the unique pupil identifier. This identifier is common to the NPD enabling us to merge NPD and PLASC data. This gives a panel of pupils with their demographic information, their KS2 and KS4 test results and the school(s) that they attended from grade 7 (first grade of compulsory secondary education) through to grade 11 (final grade of compulsory education). We then extract those pupils who attended the 208 treatment and control schools at some point over the sample period. We can now see which schools pupils attended in every secondary compulsory year of schooling[70], their demographic information and their exams results at KS4 and KS2. The intake analysis focuses on those who enter as a grade 7 student in 2001/02 – 2008/09 while the results analysis focuses on those who sit exams, are legacy in one of the 94 treatment schools, or sit exams in one of the 114 control schools over the same period.

Finally it is worth noting that PLASC does not cover years prior to 2002. For the observations before then we do still have NPD data on KS2 and KS4 performance (we have these going back to

[70] Strictly speaking this is not true. Some pupils enter the schooling system either from another country or from independent schools. We observe when the pupils enter but not precisely where they came from. These pupils are retained in the analysis.

1997 for KS4 and 1996 for KS2). Therefore, in the cohort and school specific trend estimates we include school fixed effects and time effects, but no covariates.

3. Clustering

A final note relates to how we define ‘school’. For each of the treatment and control schools we assign a unique number. It is possible that two pupils from different schools are given the same number should the two differing schools later become the same academy. We identify when schools merge by looking at linked schools in edubase (this is a Department For Education database of all open and closed maintained schools in England). In one case a single school becomes two separate academies (North Westminster Community School splits into Paddington Academy and Westminster Academy in 2006). Pupils attending the predecessor school are randomly assigned one of the two numbers given to the two academies that open later. In estimated specifications, standard errors are clustered on this unique school number resulting in 208 clusters.

4. Attainment Measures

The main variable in the analysis of intake is an average score across three subject specific key stage 2 tests: English, Maths and Science. Test scores are reported in two ways: firstly, a level from 2-5 is awarded in each subject and secondly as a raw test score. The raw test score is graded out of 80 for science and is the sum of two separate science papers each marked out of 40 while the English test score is marked out of 100 and is composed of the sum of two separate test scores, each marked out of 50, in reading and writing. Finally, Maths is composed of two marks out of 50 with one of the tests being in mental arithmetic. The levels are based upon these underlying test scores but are not always consistent. For instance, after an initial level is assigned after grading the test there may be a review of the pupil’s test score resulting in a higher or lower level being awarded even if the underlying raw test mark is not altered. Similarly, the mark required for any one level varies both between subjects and within subjects across years. For these reasons we use standardised raw test scores as the main dependent variable in KS2 regressions.

When pupils are not awarded a test mark or are performing at a level below the level of the test, we award pupils a mark of 0. Those who miss the tests are excluded from the sample for the purposes of the KS2 regressions but are included in the KS4 regressions where we include a dummy for those who do not have a KS2 record or who miss KS2 exams. The KS4 results are robust to re-running the regressions omitting those without a KS2 record and those whose scores are below test levels.

The main KS4 qualification in the UK is the GCSE (General Certificate of Secondary Education). GCSEs are graded from A*-G. The current points score calculations give an A* a score of 58 and a G a score of 16 with grades in between going up in increments of 6 between adjacent grades as follows:

Grade	Points	Grade	Points
A*	58	D	34
A	52	E	28
B	46	F	22
C	40	G	16

Prior to this an A* was given a score of 8 and G a score of 1 with scores rising in unit increments.

Grade	Points	Grade	Points
A*	8	D	4
A	7	E	3
B	6	F	2
C	5	G	1

As well as GCSEs there are a wide range of equivalent qualifications focusing on more vocational subjects. These include GNVQs and BTecs. Depending upon the type of equivalent these are often worth multiple GCSEs and are often graded as a combination of GCSE grades i.e. a distinction in an intermediate GNVQ is equivalent to gaining two GCSEs with one at grade A and the other at grade A*.[71] The points score given to the qualification reflects the underlying GCSE grades that it is based upon so that under the new scoring system the aforementioned qualification would be given a score of 110. The points system we use is as follows:

Grade	Points	Grade	Points
A*	10	D	4
A	8	E	3
B	7	F	2
C	6	G	1

The points system we use addresses some of the concerns expressed pertaining to the 16-58 and 1-8 scales used over the course of the sample.[72] The non-linearity reflects the fact that it appears hardest to jump from grades D to C and from A to A*.

We cap points scores at best 8 qualifications. To do this we normalize raw point scores by their GCSE equivalent i.e. a qualification worth 4 GSCEs and 208 points (under the 16-58 scale) is normalized to be worth 52 points. We then convert these points to the new measure and rank them highest to lowest. We then add up the grade weightings (in terms of GCSEs), taking fractions of qualifications if need be, until we reach 8. All those in the top 8 are then multiplied through by their weight and summed to give the points score.

The decision to cap at 8 is motivated by two concerns. Total points scores have the problem that pupils can appear to do well by entering many exams and performing poorly in them. Similarly

[71] Most equivalents are graded as pass, merit or distinction but the Department for Education equates these categories, combinations of, A*-G grades.
[72] We are grateful to Tim Leunig and Mike Treadaway for very helpful correspondence on this.

using, for instance, 5 best means that those who focus very narrowly on a small set of exams may appear better than those who perform well over a larger selection of subjects/qualifications. The decision to cap at 8 balances these two concerns.

Finally, it is worth noting that the point measures create some notable discrepancies with the official method. For instance, an equivalent qualification worth two GCSEs graded CD is worth 74 points under the 16-58 scale meaning that it is worth more than a A* at GCSE. Using the system such a qualification is worth 10 points (the sum of the points scores for grades of C and D) – the equivalent of a GCSE at grade A*. A further example is a BTEC that is worth 76 points on the old scale and equivalent to 4 GCSEs. This is the same as achieving grades of 2 Fs and 2 Gs. In the system this is equivalent to a point score of 6. Thus, the points mean the qualification is the same as getting a C at GCSE whereas the old measure means that the qualification is again worth more than an A*. In general, the system reduces the relative points scores of equivalent qualifications compared to the official method. Despite this the results remain unchanged when using the (standardized) old (1-8) and new (16-58) points systems and when using total rather than capped scores. We present results in Table A4 using different dependent variables.

5. Ofsted Reports[73]

Ofsted is the government department that carries out inspections of maintained schools in England and Wales and reports to Parliament. Inspectors give schools minimal warning of inspection and proceed to inspect the school based upon a pre-set criteria before awarding the school and overall

[73] Throughout this and Section 5 of the main text school refers to the variable school that we cluster on as described in Section 3 of the appendix – all mechanism regressions are performed at this level.

effectiveness rating.[74] Overall effectiveness is based upon many criteria like pupil achievement, the effectiveness of management and the level of well-being and personal development of the pupils.

Post 2005 there are 4 possible inspection ratings – outstanding, good, satisfactory, and inadequate. Prior to 2005 the possible ratings given were excellent, very good, good, satisfactory, unsatisfactory, poor, and very poor. To measure whether academies improve over time we equate the 7 ratings given prior to 2005 into the 4 categories given post 2005 in the following manner:

Prior to 2005	Post 2005
Excellent, very good	Outstanding
Good	Good
Satisfactory	Satisfactory
Unsatisfactory, poor, very poor.	Inadequate

The main interest is whether schools converting to academies are more likely to improve their rating relative to the control schools. To study this question, we use Ofsted ratings for the years 2000-2010. We limit the sample to the years 2000-2010 as post 2010 all the schools in the sample have converted to academies making any comparisons between converters and those yet to convert impossible.

For the estimates, we use all inspection outcomes available for control schools. For treatment schools, we use the latest pre-conversion inspection and the earliest post conversion one. These restrictions results in the sample of treatment schools falling to 45 with the first three cohorts not represented in the sample at all. For controls schools we omit those that only have a single inspection over the period thus reducing the sample of control schools to 110. For this sample we define a variable equal to 0 if the school's first inspection is worse than its last, 1 if the inspections are the same and 2 if

[74] Overall effectiveness ratings have been awarded since 2000.

the latter inspection is an improvement on the first. As a robustness check we replicate the results using the following two conversions for Ofsted scores:

Conversion 1

Prior to 2005	Post 2005
Excellent	Outstanding
Very good, good	Good
Satisfactory	Satisfactory
Unsatisfactory, poor, very poor.	Inadequate

Conversion 2

Prior to 2005	Post 2005	New Scale
Excellent, very good	Outstanding	Good
Good	Good	Good
Satisfactory	Satisfactory	Good
Unsatisfactory, poor, very poor.	Inadequate	Bad

The results prove robust to these changes.

6. Data on Mechanisms

As well as considering Ofsted reports we study mechanisms by looking at survey results from the Department for Education (2014), head teacher change and teacher turnover.

We collect data on head teachers using the School Workforce Census and match a head teacher to each of the schools for each year in the sample. For each year we define a binary variable equal to 1 if this year's head teacher is different from last years. When two schools merge we set this variable to 1 only if the head is not the head of either of the predecessors. When two separate schools are defined as being the same (with respect to the clustering variable) we set this variable to 1 if both schools change their head teacher in that year. We have also defined change as either one of the two schools changing their head – the results are unchanged by this.

For the teacher and pupil analysis we use data from the annual schools census. The data gives us the number of qualified and unqualified teachers in all maintained secondary schools for the years 2002-2009. The measure of teachers is the full time equivalent of both qualified and unqualified teachers, while the measure of pupils is full time equivalent pupils. Because schools often open sixth forms upon conversion to academy status, and a few schools merge with schools that enrol children pre-grade 7, we include dummies in the regressions for schools that have an attached sixth form, and those that enrol children in grades lower than 7. These latter variables come from maximum and minimum age group variables in the school performance tables.

Table A3.1: The Nature of Academy Conversions

	All Schools							
	Pre-Academy School Type							
	All	New	Independent	City technology college	Voluntary aided	Foundation	Voluntary controlled	Community
All academies	244	12	5	12	18	34	2	161
	All Schools With Full Data (Pre- and Post-Academy Conversion)							
	Pre-Academy School Type							
	All	New	Independent	City technology college	Voluntary aided	Foundation	Voluntary controlled	Community
All academies	220	0	0	12	15	33	2	158
Become academies, up to 2008/09	106	0	0	12	10	15	1	68
Future academies, after 2008/09	114	0	0	0	5	18	1	90

Notes: Source for upper panel, same as Table 2. Source for lower panel, own calculations from Edubase, School Performance Tables and Annual Schools Census.

Table A3.2: Sample Composition in Difference-in-Differences Analysis, Conversion Cohort S_t.

	Before	After
Treatment (conversion cohort S_t)	Grade 11 pupils in t-1 and prior in schools $j \in S_t$.	Pupils legacy enrolled in grades 7-10 in t-1 in schools $j \in S_t$.
Control (for conversion cohort S_t)	Grade 11 pupils in t-1 and prior in 2009/10 and 2010/11 converters.	Pupils enrolled in grades 7-10 in t-1 in 2009/10 and 2010/11 converters.

Table A3.3: Event Study Estimates, Pupil Performance, End of Secondary School Test Scores, Year 11, 2001/02 to 2008/09

	Standardised End of Secondary KS4 Test Scores				
	All	Pupils in Community Predecessor School	Pupils in Non-Community Predecessor School	Pupils in Urban Schools	Pupils in Non-Urban Schools
	IV	IV	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)
Takes KS4 in Academy x (c = – 4)	-0.011 (0.020)	-0.023 (0.022)	0.024 (0.037)	-0.012 (0.023)	0.008 (0.027)
Takes KS4 in Academy x (c = – 3)	0.011 (0.026)	0.034 (0.028)	-0.026 (0.052)	0.015 (0.029)	0.005 (0.049)
Takes KS4 in Academy x (c = – 2)	0.040 (0.030)	0.062* (0.033)	0.011 (0.059)	0.049 (0.034)	-0.018 (0.063)
Takes KS4 in Academy x (c = – 1)	0.037 (0.037)	0.066 (0.040)	0.002 (0.076)	0.044 (0.041)	-0.016 (0.090)
Takes KS4 in Academy x (c = 0)	0.061 (0.042)	0.099* (0.045)	0.003 (0.086)	0.060 (0.048)	0.028 (0.074)
Takes KS4 in Academy x (c = 1)	0.204** (0.052)	0.250** (0.060)	0.135 (0.098)	0.207** (0.058)	0.114 (0.102)
Takes KS4 in Academy x (c = 2)	0.260** (0.065)	0.281** (0.072)	0.264* (0.144)	0.255** (0.070)	-0.024 (0.091)
Takes KS4 in Academy x (c = 3)	0.283** (0.078)	0.317** (0.085)	0.235 (0.165)	0.266** (0.086)	0.239* (0.101)
Standardised End of Primary KS2 Test Score	0.599*** (0.005)	0.603*** (0.005)	0.582*** (0.012)	0.591*** (0.006)	0.625*** (0.009)
Control Variables	Yes	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes
Sample Size	1263751	981249	282502	974434	289317
Number of Treatment and Control Schools	208	158	50	170	38
First Stage Coefficient on Legacy Enrolment x (c = 0)	0.964*** (0.003)	0.962*** (0.003)	0.967*** (0.005)	0.961*** (0.003)	0.980*** (0.003)
First Stage Coefficient on Legacy Enrolment x (c = 1)	0.926*** (0.004)	0.926*** (0.004)	0.927*** (0.009)	0.924*** (0.004)	0.933*** (0.014)
First Stage Coefficient on Legacy Enrolment x (c = 2)	0.877*** (0.007)	0.879*** (0.007)	0.870*** (0.020)	0.876*** (0.007)	0.907*** (0.000)
First Stage Coefficient on Legacy Enrolment x (c = 3)	0.840*** (0.015)	0.837*** (0.017)	0.851*** (0.031)	0.836*** (0.015)	0.887*** (0.000)

Notes: Robust standard errors (clustered at the school level) are reported in parentheses. Control variables included are dummies for whether the pupil is male, the pupil's ethnicity group, and whether they are eligible for free school meals, entered together with end of primary school KS2 test scores and a dummy variable for pupils for whom KS2 data is unavailable. The dependent variable is the standardised best 8 examinations point score of the pupil – see the Data Appendix for precise definitions. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A3.4: Alternative End of Secondary School Outcomes

	Alternative End of Secondary KS4 Outcomes					
	Five A*-C, with English and Math	Five A*-C, with English and Math	English GCSE	English GCSE	Math GCSE	Math GCSE
	IV	IV Event Study	IV	IV Event Study	IV	IV Event Study
	(1)	(2)	(3)	(4)	(5)	(6)
Takes KS4 in Academy	0.022* (0.011)		0.100** (0.031)		0.074* (0.033)	
Takes KS4 in Academy x (c = – 4)		-0.003 (0.007)		0.018 (0.019)		0.012 (0.021)
Takes KS4 in Academy x (c = – 3)		0.001 (0.008)		0.031 (0.021)		0.001 (0.023)
Takes KS4 in Academy x (c = – 2)		0.007 (0.010)		0.029 (0.033)		0.032 (0.026)
Takes KS4 in Academy x (c = – 1)		0.016 (0.011)		0.074* (0.034)		0.039 (0.031)
Takes KS4 in Academy x (c = 0)		0.014 (0.014)		0.111*** (0.039)		0.051 (0.043)
Takes KS4 in Academy x (c = 1)		0.039* (0.017)		0.165*** (0.053)		0.136*** (0.044)
Takes KS4 in Academy x (c = 2)		0.040 (0.021)		0.137* (0.064)		0.152*** (0.052)
Takes KS4 in Academy x (c = 3)		0.085** (0.023)		0.192* (0.082)		0.220*** (0.055)
Standardised End of Primary KS2 Test Score	0.212*** (0.005)	0.212*** (0.005)	0.581*** (0.006)	0.581*** (0.006)	0.656*** (0.006)	0.656*** (0.006)
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	1263751	1263751	1263751	1263751	1263751	1263751
Number of Treatment and Control Schools	208	208	208	208	208	208
First Stage Coefficient on Legacy Enrolment	0.938*** (0.003)		0.938*** (0.003)		0.938*** (0.003)	
First Stage Coefficient on Legacy Enrolment x (c = 0)		0.964*** (0.003)		0.964*** (0.003)		0.964*** (0.003)
First Stage Coefficient on Legacy Enrolment x (c = 1)		0.926*** (0.004)		0.926*** (0.004)		0.926*** (0.004)
First Stage Coefficient on Legacy Enrolment x (c = 2)		0.877*** (0.007)		0.877*** (0.007)		0.877*** (0.007)
First Stage Coefficient on Legacy Enrolment x (c = 3)		0.840*** (0.015)		0.840*** (0.015)		0.840*** (0.015)

Notes: Robust standard errors (clustered at the school level) are reported in parentheses. Control variables included are dummies for whether the pupil is male, the pupil's ethnicity group, and whether they are eligible for free school meals, entered together with end of primary school KS2 test scores and a dummy variable for pupils for whom KS2 data is unavailable. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A3.5: Event Study Estimates, Change in Staff and Pupils Before and After Academy Conversion

	Change in Head teacher		Log(Number of Teachers)		Log(Number of Pupils)		Log(Teachers Per Pupil)	
	(1)		(2)		(3)		(4)	
Academy x (c = – 4)	0.044	(0.043)	0.014	(0.012)	0.007	(0.011)	0.008	(0.009)
Academy x (c = – 3)	0.064	(0.057)	0.012	(0.019)	-0.006	(0.016)	0.019	(0.016)
Academy x (c = – 2)	0.032	(0.066)	0.004	(0.025)	-0.025	(0.024)	0.028	(0.019)
Academy x (c = – 1)	0.076	(0.085)	0.002	(0.034)	-0.048	(0.033)	0.050**	(0.025)
Academy x (c = 0)	0.453***	(0.106)	0.036	(0.046)	-0.035	(0.047)	0.071**	(0.031)
Academy x (c = 1)	0.056	(0.113)	0.112*	(0.060)	0.040	(0.058)	0.073*	(0.038)
Academy x (c = 2)	0.102	(0.141)	0.147*	(0.078)	0.099	(0.073)	0.048	(0.043)
Academy x (c = 3)	0.143	(0.163)	0.277***	(0.077)	0.229***	(0.075)	0.048	(0.060)
School Fixed Effects	Yes		Yes		Yes		Yes	
Year Dummies	Yes		Yes		Yes		Yes	
Sample Size	1641		1641		1641		1641	
Number of Treatment and Control Schools	208		208		208		208	

Notes: Robust standard errors (clustered at the school level) are reported in parentheses. Columns (2), (3), and (4) control for whether the schools, in each year, enrol pupils prior to grade 7 and post grade 11. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

References

Abdulkadiroglu, Atila, Joshua Angrist, Susan Dynarski, Thomas Kane and Parag Pathak (2011) Accountability and Flexibility in Public Schools: Evidence From Boston's Charters and Pilots, *Quarterly Journal of Economics*, 126, 699-748.

Abdulkadiroglu, Atila, Joshua Angrist, Yusuke Narita and Parag Pathak (2017) Research Design Meets Market Design: Using Centralised Assignments for Impact Evaluation, *Econometrica*, 85, 1373-1432.

Abdulkadiroglu, Atila, Joshua Angrist, Peter Hull and Parag Pathak (2016) Charters Without Lotteries: Testing Takeovers in New Orleans and Boston, *American Economic Review*, 106, 1878-1920.

Adonis, Andrew (2012) Education, Education, Education: Reforming England's Schools, Biteback Publishing.

Altonji, Joseph, Todd Elder and Christopher Taber (2005) Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools, *Journal of Political Economy*, 113, 151-84.

Angrist, Joshua, Sarah Cohodes, Susan Dynarski, Parag Pathak and Christopher Walters (2016) Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry and Choice, *Journal of Labor Economics*, 34, 275-318.

Angrist, Joshua, Susan Dynarski, Thomas Kane, Parag Pathak and Christopher Walters (2010) "Inputs and Impacts in Charter Schools. KIPP Lynn." *American Economic Review*, 100, 239-43.

Angrist, Joshua, Parag Pathak and Christopher Walters (2013) Explaining Charter School Effectiveness, *American Economic Journal: Applied Economics*, 5, 1-27.

Betts, Julian, Lorien Rice, Andrew Zau, Emily Tang and Cory Koedel (2006) Does School Choice Work? Effects on Student Integration and Achievement, Public Policy Institute of California.

Clark, Damon, (2009) The Performance and Competitive Effects of School Autonomy, *Journal of Political Economy*, 117, 745-83.

Center for Research on Education Outcomes (2009) Multiple Choice: Charter Performance in Sixteen States, Stanford University, CREDO, Stanford University.

Center for Research on Education Outcomes (2013), National Charter School Study, Stanford University, CREDO, Stanford University.

Department for Children, Schools and Families (2007) Academies and Independent Schools: Prospectus, https://core.ac.uk/download/pdf/4156163.pdf

Department for Children, Schools and Families (2008) The Composition of Schools in England, http://webarchive.nationalarchives.gov.uk/20130321145625/https://media.education.gov.uk/assets/files/pdf/b022008pdf.pdf

Department for Education (2014) Do Academies Make Use of Their Autonomy?, https://www.gov.uk/government/publications/do-academies-make-use-of-their-autonomy

Dobbie, Will, and Roland Fryer (2011) Are High Quality Schools Enough to Close the Achievement Gap? Evidence From a Social Experiment in Harlem, *American Economic Journal: Applied*, 3, 158–87.

Dobbie, Will and Roland Fryer (2013) Getting Beneath the Veil of Effective Schools: Evidence from New York City, *American Economic Journal: Applied*, 5, 58–75.

Dobbie, Will and Roland Fryer (2014) The Medium Term Impacts of High Achieving Charter Schools, *Journal of Political Economy*, 5, 985–1037.

Epple, Dennis, Richard Romano and Miguel Urquiola (2017) School Vouchers: A Survey of the Economics Literature, *Journal of Economic Literature*, 55, 441-492.

Epple, Dennis, Richard Romano and Ron Zimmer (2016) Charter Schools: A Survey of Research on their Characteristics and Effectiveness, in Hanushek E., S. Machin and L. Woessmann (eds.) *Handbook of the Economics of Education*, Volume 5, Amsterdam: Elsevier.

Eyles, Andrew, Claudia Hupkau and Stephen Machin (2016) Academies, Charter and Free Schools: Do New School Types Deliver Better Outcomes? *Economic Policy*, 31, 453-501.

Eyles, Andrew, and Stephen Machin (2015) The Introduction of Academy Schools to England's Education, Centre for Economic Performance Discussion Paper 1368.

Eyles, Andrew, Stephen Machin and Sandra McNally (2017) Unexpected School Reform: Academisation of Primary Schools in England, *Journal of Public Economics*, 155, 108-212.

Eyles, Andrew, Stephen Machin and Olmo Silva (2018) Academies 2: The New Batch, *Fiscal Studies,* 39(1), 121-158.

Evans, William and Robert Schwab (1995) Finishing High School and Starting College: Do Catholic Schools Make a Difference? *Quarterly Journal of Economics*, 100, 941-74.

Fryer, Roland (2014) Injecting Charter School Best Practices into Traditional Public Schools: Evidence From Field Experiments, *Quarterly Journal of Economics*, 129, 1355-1407.

Gleason, Philip, Melissa Clark, Christina Clark Tuttle, Emily Dwoyer and Marsha Silverberg (2010) The Evaluation of Charter School Impacts, (NCES 2010-4029), U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Gorard, Stephen (2014) The Link Between Academies in England, Pupil Outcomes and Local Patterns of Socio-Economic Segregation Between Schools, *Research Papers in Education*, 29, 268-284.

Hanushek, Eric, and Ludger Woessmann (2011) The Economics of International Differences in Educational Achievementm in Hanushek Eric, Stephen Machin and Ludger Woessmann (eds.) *Handbook of the Economics of Education*, Volume 3, Amsterdam: Elsevier.

Hanushek, Eric, and Ludger Woessmann (2015) The Knowledge Capital of Nations: Education and the Economics of Growth, MIT Press.

Hoxby, Caroline, Sonali Murarka and Jenny Kang (2009) How New York City's Charter Schools Affect Achievement, New York City Charter Schools Evaluation Project.

Machin, Stephen, Shqiponja Telhaj and Joan Wilson (2006) The Mobility of English School Children, *Fiscal Studies*, 27, 253-80.

Machin, Stephen and James Vernoit (2011) Changing School Autonomy: Academy Schools and their Introduction to England's Education, Centre for the Economics of Education Discussion Paper 123.

Machin, Stephen and Joan Wilson (2008) Public and Private Schooling Initiatives in England: The Case of City Academies, in Chakrabarti R. and P. Peterson (eds.) *School Choice International*. MIT Press, Cambridge MA.

National Audit Office (2010) Department for Education: The Academies Programme, http://www.nao.org.uk/publications/1011/academies.aspx..

Neal, Derek. (1997) The Effects of Catholic Secondary Schooling on Educational Achievement, *Journal of Labor Economics*, 15, 98-123.

OECD (2011) School Autonomy and Accountability: Are They Related to Student Performance?, PISA Focus 2011/9.

Price Waterhouse Coopers (2008) Academies Evaluation Fifth Annual Report. Annesley, DCSF Publications

West, Anne and Elizabeth Bailey (2013) The Development of the Academies Programme: 'Privatising' School-Based Education in England 1986-2013, *British Journal of Educational Studies* 61, 137-59.

Chapter 4: Unexpected School Reform: Academisation of Primary Schools in England

Abstract

The change of government in 2010 provoked a large structural change in the English education landscape. Unexpectedly, the new government offered primary schools the chance to have 'the freedom and the power to take control of their own destiny', with better performing schools given a green light to convert to become an academy school on a fast track. In England, schools that become academies have more freedom over many ways in which they operate, including curriculum design, budgets, staffing issues and the shape of the academic year. However, the change to allow primary school academisation has been controversial. This paper reports estimates of the effect of academisation on primary school pupils. While the international literature provides growing evidence on the effects of school autonomy in a variety of contexts, little is known about the effects of autonomy on primary schools (which are typically much smaller than secondary schools) and in contexts where the school is not deemed to be failing or disadvantaged. The key findings are that English primary schools did change their modes of operation after the exogenous policy change, utilising more autonomy and changing spending behaviour, but at this primary phase of schooling academisation did not lead to improved pupil performance.

4.1 Introduction

Since 2010, the educational landscape in England has radically altered. By 2017, nearly two-thirds of secondary schools and over a fifth of primary schools are academies, which are schools that have been granted considerable operational autonomy by government. As Michael Gove, the Minister then responsible for education, put it schools have been 'given the freedom and the power to take control of their own destiny'.[75]

Although academies were present before then – principally as a school improvement policy for underperforming secondary schools since 2002 - the programme was radically altered and expanded following the election of the new UK government in May 2010. It became a school structure to which all schools were invited to aspire.

Enabling legislation - Academies Act of 2010 - was rapidly put in place two months after the election of the new government[76]. For the first time, and through a completely unexpected policy change[77], primary schools were invited to become academies, with better performing schools given priority to convert. The first batch of such schools converted in the school year beginning in September 2010. This paper reports estimates of the impact of primary school conversion on their operation and on the performance of enrolled pupils.

This introduction of primary academies took place in an international context where publicly-funded, autonomous schools have become a familiar school improvement policy, most notably through charter schools in the US and free schools in Sweden. Research on the former

[75] Department for Education (2013). Forward by Michael Gove MP.

[76] Most new academies since 2010 are 'converters'. However, some academies are sponsored (i.e. managed by a private team of independent co-sponsors) and these are schools that have been underperforming. The effect of academisation on these schools (which are closer to the original New Labour academies, studied by Eyles and Machin, 2015, and comprise about 30 percent of primary academies) is not considered in this paper, because we want to explore the unexpected dimension of academisation that applied to converters, and especially those rated outstanding prior to the 2010 change in policy.

[77] The introduction of 'Free schools' and education reform were issues raised in the manifesto of the new government prior to their election; however, there was no mention of large-scale expansion of the academies programme. Free schools are completely new schools that can be set up by interested parties (e.g. parents, or community members). By 2016/17, there were 139 free primary schools open or approved.

tends to find achievement gains associated with charter status and with the 'injection' of charter school features to public schools, particularly in urban settings where the schools typically enrol disadvantaged students (Epple, Romano, and Zimmer, 2015 provide an overview of the literature).[78] Studies of the Swedish free schools report some positive short-term effects, working through competition, but no evidence of medium to long-term effects (see Bolhmark and Lindahl, 2015).

The policy studied here is different from most others in the literature in three important respects. Firstly, it involves conversion of existing schools rather than the creation of new schools.[79] Secondly, it is about the voluntary conversion of better performing schools and not the forced conversion of failing schools. These better performing schools tend to have lower proportions of children from disadvantaged backgrounds. Thirdly, the focus is on young children (aged 7-11) who attend primary schools, which are much smaller than secondary schools[80]. Although there have been studies of elementary schools in the charter school context, these are less prevalent than studies of middle and high schools. Similarly, studies of autonomy in the context of the English education system have focused on a particular subset of secondary schools; specifically, advantaged secondary schools voluntarily gaining greater autonomy (Clark, 2009), disadvantaged secondary schools (Eyles and Machin, 2015), and secondary

[78] While something of a consensus has emerged, there is some controversy within the literature. Recent experimental studies of charters in or near particular US cities (Boston and New York) find positive impacts on educational achievement (see Abdulkadiroglu et al. 2011, 2014; Angrist et al. 2013, 2016; Dobbie and Fryer 2011; Hoxby and Murarka 2009). Wider coverage evaluations have produced more mixed results (Betts et al. 2006; Center for Research on Education Outcomes, 2009, 2013; Gleason et al. 2010). Similarly, there is no consensus on the longer term effects of charters. Angrist et al. (2016) and Dobbie and Fryer (2014) fine that charter attendance improves longer run outcomes such as college attendance. In later work, Dobbie and Fryer (2016) find negative earnings returns for those attending charters that are ineffective at raising test scores and no returns for charters that are successful at raising test scores.

[79] While the majority of school autonomy studies focus on newly set up autonomous schools (e.g. the majority of US charters are new schools), there are some examples of studies where existing schools become more autonomous. Clark (2009) and Eyles and Machin (2015) study English secondary schools gaining more autonomy, while Abdulkadiroglu et al. (2016) study the conversion of traditional public schools in New Orleans to (in-district) charters. Alongside these Steinberg (2014) studies the granting of greater operational freedom to a subset of principals in already existing Chicago Public Schools.

[80] While the majority of charter papers focus on middle and high schools, some papers do include results for elementary schools (Dobbie and Fryer 2011, 2013; and Hoxby et al. 2009).

schools in relatively disadvantaged local authorities (e.g. Birmingham in the case of Bertoni et al., 2017).

Upon conversion, academy schools fall outside of local authority control and gain autonomy over many process and personnel decisions. This greater freedom may have positive effects on student outcomes because of superior information held by local decision makers (Hanushek and Woessmann, 2011). Indeed, the first secondary schools in England to become academies (in the early 2000s) did seem to deliver positive effects on student outcomes (Eyles and Machin, 2015; and Eyles at al. 2016a, 2016b). However, the context was one in which a couple of hundred (previously underperforming) secondary schools became academies. It is not necessarily the case that these positive effects carry through to better performing schools and/or to (much smaller) primary schools.

If the autonomy offered within the academies model was unambiguously advantageous for schools, one would imagine that all schools would want to become academies. However, recently the UK government has had to back out of a policy to force all schools in England to become academies by the end of 2022 because of fierce hostility to this by the educational establishment (although the current government vision is still to encourage all schools to become academies).

Whether such radical upheaval is in the interests of students is an empirical question. Most schools yet to convert are primary schools, which represent most schools in England. One might hypothesise that schools which volunteered to convert to academy status early-on are those that were most amenable to academy status, anticipating positive benefits. If effects are not found for such schools, one might question whether it is such a good idea to extend it to schools that are less enthusiastic.

An important feature of the policy being studied here is that it was in no way anticipated by schools or parents. This gives us leverage to identify causal effects since the conversion was exogenous to pupils already enrolled in the school. Thus, the sample is restricted to these "legacy enrolled" pupils who can be observed before and after academisation takes place. The importance of estimating effects for pupils who were already enrolled in the school prior to conversion is that student mobility post-conversion is potentially endogenous to the policy itself. For example, parents may be attracted by the idea of academy status and be more likely to enrol their students to newly converted primary schools. Exit from the school post-conversion might also be non-random (for example, if schools change policies in a way that is less attractive to certain students or their parents). However, a very strong first stage estimate (of the effect of pre-conversion enrolment on the probability of attending an academy) suggests that the effect is estimated for the majority of eligible pupils in the school.

In practical terms the empirical strategy first involves selection of a treatment and control groups of schools. The treatment groups consist of primary schools that converted to academy status between 2010/11 and 2014/15. In each case, the control groups are those that converted in later academic years, but before 2016/17. These treatment and control groups are shown to have similar pre-trends in outcome variables. Further, enrolment in the primary school prior to conversion can be used as an instrument for actual attendance in the academy in grade 6 when national tests in reading and maths take place. The legacy enrolment strategy mirrors that used in Eyles and Machin (2015) in their study of the first underperforming English secondary schools to become academies in the early 2000s. It also draws on Fryer (2014) who looks at the effect of injecting charter school practices into traditional public schools and Abdulkadiroglu et al. (2016), who study school takeovers in New Orleans, referring to pupils who stay in converting schools as 'grand-fathered' pupils

The rest of the paper is structured as follows. Section 2 describes primary education in England and how academies have been introduced. Section 3 describes the data and research strategy. Section 4 reports results from the first part of the empirical analysis, looking at whether primary schools that became academies did in fact change their modes of operation upon conversion. Section 5 report the legacy enrolment results looking at causal effects of academy conversion on pupil performance. Conclusions are given in Section 6.

4.2 Primary Education in England and Academies Policy

In England, children start school in the September after they reach the age of 4. Most children attend a primary school up to age 11, after which they go to secondary school.[81] Schooling in England is organised into Key Stages. At the end of Key Stage 1 (age 7), pupils are assessed by their teachers in English, Science and Maths according to national guidelines. At the end of Key Stage 2 (age 11), they undertake national tests in English and Maths.[82] These tests are used to construct Performance Tables for primary schools, which are publicly available. There is very little grade repetition within the system.

Up until the introduction of academies in 2010, schooling had been organised at the local level into Local Education Authorities (LEAs). There are 152 LEAs in England and around 15,000 primary schools. The LEA's main functions in relation to primary schools are in building and maintaining schools, providing support services (e.g. for children with special needs), and acting in an advisory role to the head teacher regarding school performance and implementation of government initiatives. LEAs also have an important role in the funding allocations of schools. The bulk of schools funding comes from the dedicated schools grant

[81] There is a small number of infant schools and middle schools in parts of the country. They are not included in this analysis unless they are 'linked', meaning that students at an infant school are prioritised for places at the junior school; in these cases, the proportion of infant school attendees switching to the linked junior school is very high and the two linked schools are treated as though they were one single school.

[82] Prior to 2010, students were also assessed in science.

which is given to LEAs and then distributed according to each LEAs own funding formula. The funding allocated to the LEA is based on a historically determined formula which is mainly driven by the numbers of pupils, ‘additional educational needs’ and local conditions. These local conditions include population sparsity, measures of deprivation, and wage costs in the area (Roberts and Bolton, 2017). As well as allocating funding, the LEA also appoints one or two representatives on to a school’s governing body – a group of parents, teachers and community representatives that provides governance to the school. LEAs typically also offer several administrative and management functions including training, personnel, and financial services. Up until the 2010/11 school year, most primary aged pupils (67%) attended community schools in which LEAs are also the statutory employer of school staff, owner of the buildings and the authority that manages student admissions.[83] Most other state primary schools are faith schools (which have greater autonomy from the LEA). Although parents can apply to send their child to any primary school (i.e. there are no strict catchment areas), popular schools are often oversubscribed and places are rationed according to a Schools Admissions Code.[84]

When a school becomes an academy, it is governed outside the Local Authority and is overseen and funded directly by central government. A school is run in many ways like a company, where governors are classed as trustees or directors and the principal/head teacher is the chief executive. Strong financial management and governance at the level of the individual academy are very important (National Audit Office, 2012), especially given that oversight is no longer provided by the Local Authority. Unlike Community Schools (i.e. most state primary

[83] For more information about the operation of primary schools and local government prior to 2010, see Gibbons et al. (2011).

[84] The School Admissions code applies to all state-maintained schools and academies. In practice, schools have very little scope to employ differing admission criteria (all schools aside from community schools and a limited amount of religious schools, where the LEA determines the criteria, set their own). All schools have to accept applications unless they are oversubscribed. In the case of oversubscription the criteria that schools can use is limited (distance, presence of sibling at school) and the adopted criteria tends to vary little across schools.

schools), academies manage their own admissions. While they still have to adhere to the Schools Admissions Code, they may choose to run their admissions policy differently than in the past. Although academies are required to teach a broad and balanced curriculum, including English, maths, science and religious education, they are not legally required to use the national curriculum. They have the ability to set their own pay and conditions for staff and more freedom in their hiring decisions (e.g. they may hire unqualified teachers).[85] Although academies should be funded on an equal basis with non-academies, they do get extra funds to cover the services that the LEA provides freely to other state maintained schools[86]; therefore, they have greater freedom on how to use the budget allocation. They also have the responsibility of organising payroll functions, insurance, and accountancy functions in-house or by contracting this out. Academies also can change the length of the school day and the shape of the academic year (through term times).

In the interests of minimising risk, the Department of Education adopted a phased approach to the criteria for schools wishing to convert (National Audit Office, 2012), prioritising better performing schools. A key component of this decision is the report by the Schools Inspectorate (Ofsted) that visits schools every 3-5 years and rates schools on a four-point scale ranging from ‘outstanding’ to ‘unsatisfactory’. At the time, about 20% of schools were rated as ‘outstanding’ and 50% as ‘good’.

The government initially prioritised schools rated as outstanding and fast-tracked their applications for conversion. The first such schools were converted to academy status in September 2010. In November 2010, this fast-track route was extended to all good schools with outstanding features. At the same time, recognising the potential for economies of scale,

[85] Those with qualified teacher status typically have an undergraduate degree and have completed a one-year postgraduate teacher training course.
[86] Schools are given a £25,000 grant to support the conversion process.

academies were also encouraged to convert in chains or undertake some post-conversion collaborative arrangement with other schools. This option was made available for any school (irrespective of Ofsted grade) if it joined an academy trust with an outstanding school or an education partner with a strong record of improvement. In April 2011, the criteria were further widened to include schools that were ‘performing well’, which included consideration of the last three years’ exam results, the latest Ofsted inspections, and financial management.

As shown in Figure 4.1, the initial take-up rate for primary academies in the first possible academic year (2010/11) was modest. This is unsurprising given the unexpected nature of the announcement, with legislation being rapidly passed by receiving royal assent in June 2010 and the fact that schools are likely to take time before making the decision to take on extra responsibilities (especially given the small size of primary schools in England). However, after that, there was a huge rise in the number of primary school academies in England between 2010/11 and 2016/17, with nearly a quarter of the sector being academy schools by 2016/17.

The number of schools in the sample of converter academies studied in this paper, by year of academy conversion, is given in Table 4.1. There are a number of reasons for the discrepancy in numbers between the Figure and the Table; firstly, because of our research design we only include schools in the sample that have students in grades 2 and grades 6 in each academic year between 2006/07 and 2014/15; secondly, we only focus on academies that voluntarily convert to academy status (around 30 percent of primary academies are sponsored academies that typically convert to academies as a result of government intervention); and finally, we remove schools that participate in the KS2 strike in 2009/10 and therefore have missing outcome data in that year.

Following the way in which numbers were constructed for Figure 4.1, schools are said to convert in a given academic year if they are running as an academy by December of that academic year; for example, a school is classed as converting to academy status in 2014/15 if it converts between January and December (inclusive) of 2014.

Schools have also been encouraged to convert in a chain or partnership because ‘this can enable schools to support one another once they are academies, share resources, experience, and ideas. Such an approach is particularly valuable to small primary schools where working together allows economies of scale to be achieved’ (Department for Education, 2013). The most prevalent model of collaboration is the multi-academy trust (MAT) wherein all schools are governed by one trust and board of directors. MATs perform a role like that which would otherwise be played by the LEA in that they hire/fire teachers and are responsible for negotiating every aspect of teacher contracts - the disciplinary process and redundancy pay amongst other things - with the exception of pensions. MATs can also substitute for local educational authorities (LEAs) in that they top-slice funds allocated to schools under their trust and use this to supply central services previously provided by the LEA. In 2016/17, about 80% of primary academies were in a multi-academy trust. In the sample studied here, there are slightly fewer schools in MATS –70%.

4.3 Data and Methodology

4.3.1 Data

The National Pupil Database (NPD) is a census of all pupils in the state system in England. During the primary phase of education, this accounts for the vast majority of children. The NPD includes basic demographic details of pupils – such as ethnicity, deprivation (measured by whether they are eligible to receive free school meals), gender, and whether or not English is their first language. The school attended by pupils can be linked to other school-level information such as the date of conversion to an academy school and the date and grade of Ofsted inspections (which are publicly available data). The data is longitudinal and tracks students as they progress through the school system.

As discussed in Section 4.2, the national curriculum in England is organised around 'Key Stages', the first two undertaken in primary school (at ages 5-11) and the second two in secondary school (at ages 11 to 16). At the end of Key Stage, head teachers have a statutory duty to ensure that their teachers comply with all aspects of the Key Stage 1 assessment and reporting arrangements. The nationally set assessments are in reading, writing, speaking, and listening, mathematics and science. Local Authorities (and other recognised bodies) are responsible for moderation of schools. Thus, although teachers make their own assessments of students (and therefore are susceptible to potential bias), there is a process in place to ensure that there is a meaningful assessment that is standardised over all of England. At age 7, students are given a 'level' (i.e. there is no test score as such). However, following standard practice, National Curriculum levels achieved in Key Stage 1 assessments are transformed into point scores using Department for Education point scales and these scores are included in the regressions that are estimated.[87]

At the end of primary school (or the Key Stage 2 phase of education), pupils take national tests in reading and maths, which are externally set and marked on a scale of 1-100.

Our final dataset contains cross sections of grade 6 pupils linked to their school, demographic information, and test scores for the academic year 2006/07 to 2014/15. Test scores – both baseline KS1 and the outcome KS2 - are standardised, within the sample, at the grade/year/subject level.

To supplement our main analysis, we also include some other data sources. The School Workforce Census is school level data that is available from the 2010/11 school year and provides a snapshot of each maintained school's workforce composition. We also utilise publicly available information on the income and expenditure of maintained schools and

[87] The point score can take on 7 values ranging from 3-27. The main results, given in Table 8, are unchanged when KS1 is removed as a control.

academies that is available from the 2009/10 school year. Finally, some results from a survey conducted by the Department for Education regarding the use of academy freedoms is presented (Cirin; 2014). This survey, which covers 25% of the 2919 academies open by 1st May 2013, pertains to the freedoms exercised by schools once they gain academy status.

4.3.2 Methodology

The main question of interest is to identify the effect of academy conversion on pupil achievement in national tests taken by pupils at the end of primary school, i.e. Key Stage 2 (KS2) when they are aged 11. In order to do this, instrumental variables are combined with difference-in-differences. The outcomes of individuals in academies are compared with those who attend schools that later become academies, but do so after they sit their KS2 exams. To allow for students to (potentially) sort into schools non-randomly as a result of the school obtaining academy status a binary variable – whether the pupil was already enrolled in the school in the year prior to conversion in grades 2-5 – is used as an instrument for academy attendance. Those for whom this variable takes a value of one are referred to as being intention-to-treat (ITT).

Attention is focussed on those pre-enrolled in grades 2 to 5 as grade 5 is the penultimate year of primary education and grade 2 is when the KS1 assessment takes place, thus ensuring that KS1 assessment does not take place in an academy for the ITT pupils[88]. It is important to note that pupils enrolling in the school after conversion are not included in the analysis. To ensure that the control group and treatment group are selected in the same way, a slightly different control group is used for each cohort of academy converters. For those converting in

[88] As schools typically enrol students 3 years prior to grade 2, we could estimate the effects for students pre-enrolled in academies in earlier grades. In our case, we would be able to estimate the effect of one extra year for those in grade 1 in the first cohort of converters. However, doing so would either entail dropping KS1 scores from our estimates or assuming that KS1 performance is unaffected by academy attendance. Given that we do not gain many observations by adding extra pre-enrolled pupils we focus on those pre-enrolled in grade 2-5.

2010/11 for example, the control group consists of pupils who are in grades 2-5 in 2009/10 at schools that convert between 2011/12 and 2016/17, but are not expected to sit their exams in an academy[89]. Control groups are defined similarly for all schools converting up to and including 2014/15[90].

Because of these restrictions the event study on pupil performance must be limited to a maximum of four years post conversion, including the year of conversion itself. This is because there are 4 remaining years of primary school after the Key Stage 1 assessment. Thus, pupils affected by conversion in grade 2 of primary school (when KS1 assessments are taken) could have up to four post-conversion years of education in the academy. Similarly, children affected by conversion when enrolled in the predecessor school in grade 3 could have up to three conversion years, and so on for children in grades 4 and 5 in the predecessor school.

Incorporating these features into a research design enables us to estimate the causal impact of being in an academy. Administrative data that follows pupils through their school careers is used to estimate the impact on Key Stage 2 performance by means of the following equations:

$$\mathrm{Academy_i} = \alpha_s + \alpha_t + \theta_1 \mathrm{ITT}_i + \sum_{j=1}^{J} \pi_{1j} X_{ji} + \varphi_1 \mathrm{KS1_i} + v_{1i} \quad (1)$$

$$\mathrm{KS2_i} = \alpha_s + \alpha_t + \theta_2 \mathrm{ITT_i} + \sum_{j=1}^{J} \pi_{2j} X_{ji} + \varphi_2 \mathrm{KS1_i} + v_{2i} \quad (2)$$

Where i denotes pupil, α_s is a legacy enrolment school fixed effect[91] and α_t is a time effect for the academic year in which the pupil is in grade 6. The vector X is a set of control variables, and $\mathrm{Academy_i}$ takes value 1 if pupil i sits their end of primary school KS2 examination in an

[89] For example, in the case of the first cohort this necessitates removing those who, in 2009/10, are grades 2-4 in 2011/12 converters, grades 2-3 in 2012/13 converters and grade 2 in 2013/14 converters.
[90] The exact control group/treatment group structure is given in the appendix
[91] These are identified by our inclusion of grade 6 students who sit exams in treatment schools prior to conversion.

academy school. ITT_i is the legacy enrolment status of pupil i as defined earlier. In the first stage, (1), estimates of θ_1 show the proportion of the ITT group that stay in the academy and take their KS2 tests there. Equation (2) is the reduced form regression of KS2 on the instrument. A two stage least squares estimate (2SLS) estimate can then be obtained as the ratio of the reduced form coefficient to the first stage coefficient, θ_2 / θ_1. Our main specifications are pooled estimates of equations (1) and (2) for each of the five cohorts of academy conversions.

Extending this to an event study framework enables separate estimates for the number of years a pupil is exposed to being in an academy post conversion (up to a maximum of four including conversion year) to be obtained. In this case, there are four instruments for whether a pupil is expected to sit their exams in the year of conversion (those in grade 5 in the year prior to conversion to instrument one year of exposure), the next year (those in grade 4 in the year prior to conversion to instrument two years of exposure) and so on up to the maximum of four years exposure (for those legacy enrolled in grade 2). It should be noted that, because of the data that we have, not all cohorts of converters contribute to the exposure estimates for later years. For instance, we can only identify the effect of four years exposure for those who are pre-enrolled in grade 2 in the first two cohorts of conversions.

4.3.3 Comparison Schools

A naive comparison between primary academies and all other state-maintained schools is likely to suffer from significant selection bias, since (as discussed above) conversion to an academy was done on a voluntary basis and better-performing schools were prioritised and actively encouraged to convert.[92] One might expect schools seeking to become academies to have common unobservable characteristics such as having a school ethos more in line with the academy model. To account for this, pupils attending future converters - schools that convert

[92] In other words, the instrument is only assumed to satisfy the exclusion restriction conditional on pupils being in a well-defined sub-sample of the population.

in the 2015/16 and 2016/17 academic years – are used as a control group in a difference-in-differences setting. Thus, the data structure that is utilised is a balanced panel of schools for the school years 2006/07 to 2014/15 with repeated cross-sections of grade 6 pupils.

4.3.4 Balancing Tests

This approach can be legitimised first through covariate balancing tests between treatment and controls in the baseline academic year (2006/07) are shown. Second, and probably more importantly, the empirical analysis shows there to be no evidence on differential pre-conversion trends in outcomes between pupils in treatment and control schools.

On the former of these, Table 4.2 shows the extent to which treatment and control groups are balanced at baseline (2006/07) for the full sample of treatment and control schools, and separately for outstanding and non-outstanding schools. In terms of the full sample, there is a significant difference with respect to KS2 scores prior to the policy, with treatment schools being better performing in maths. The workforce of treatment schools also appears to be both larger and, on average, younger.

The above differences are not surprising as the government prioritised better performing schools for conversion to academy status. For instance, within our sample of schools, over 80% of the first cohort of conversions were deemed outstanding by Ofsted. This proportion declines monotonically to 11% for the 2016/17 cohort of conversions[93]. For this reason, we look within Ofsted grades (as defined by the latest Ofsted grade awarded prior to 2010/11) when comparing treatment and control schools. When we do this, the schools look much more balanced on observables. In fact, as Table 4.2 shows, within outstanding schools and non-outstanding there are few statistically significant differences at baseline between treatment and control schools. In fact, for outstanding schools there are no statistically

[93] The exact proportions for the years 2010/11 to 2016/17 are 83%, 43%, 20%, 17%, 14%, 14%, and 11%.

significant differences for any baseline characteristics including KS2 and KS1 scores. Thus, regressions are estimated for schools within each Ofsted grade, as well as for the pooled sample.[94]

4.4 Did Primary Academies Change Their Modes of Operation?

Before looking at the effect of primary academies on pupil performance, evidence is presented on whether changes in the mode of operation occurred at primary schools that became academies prior to or during the 2014/15 academic year. Four aspects of this are considered. First, whether primary schools took up the option to exercise the many academy freedoms that became available from increased autonomy. Second, whether patterns of expenditure changed. Third, whether there were changes in workforce composition. Fourth, whether academies altered their pupil intake.

4.4.1 Use of Academy Freedoms

There have been various investigations into whether schools use their academy freedoms upon conversion (e.g. Academies Commission, 2013; Cirin, 2014). The existing descriptive evidence confirms that they mostly do, but with some degree of variation. The Academies Commission (2013) conclude that take-up of freedoms had been 'piecemeal rather than comprehensive', in part because changes can take time to implement and sometimes require consultation. Surveys of recent converters by Bassett et al. (2012) and Cirin (2014) found financial motives to be important in the decision to convert. In the former study, over 75% of respondents cited it as one of their reasons for converting and two-fifths as their primary reason. Cirin (2014) found that the desire 'to gain greater freedom to use funding as you see

[94] While there are four possible grades awarded by Ofsted, two groups are considered, outstanding and non-outstanding, the latter comprising good, satisfactory, and unsatisfactory. The direct focus on the outstanding group is because they were the ones that were earmarked for the fast track to become an academy. The non-outstanding group are also amalgamated on account of the relatively small number of satisfactory and unsatisfactory schools in the sample. When estimating the pooled regressions over all schools, all variables are interacted with Ofsted grade.

fit' was the most cited reason for conversion (cited by 83% of respondents). The vast majority (almost 9 in 10) also moved to procure services themselves.

Importantly, Cirin (2014) breaks down results by primary and secondary status. This shows that many academies do exercise freedoms, but this is more common in secondary than in primary schools. This is shown in Table 4.3, taken from his survey of 720 academies which were open on 1 May 2013. The numbers in the Table show most schools report a use of academy freedoms, but that the percentage of primary schools making a particular change is smaller than it is for secondary schools. Furthermore, Cirin (2014) reports that almost all schools surveyed made at least one change (702 out of 720), implying that at least 95% of primary converters (262 primary converters were surveyed) exercised at least one freedom, with two-thirds believing that the changes improved attainment.

4.4.2 Changes in Expenditure Patterns

Studies cited above on the use of academy freedoms suggest that the financial motive to convert was important. Table 4.4 shows numbers on income and expenditure before and after conversion in treatment and control schools using administrative data on school income and expenditure. Changes between the 2009/10 and the 2014/15 school years are reported. There are some data issues that need to be highlighted upfront before discussing these numbers. First, the timing of reporting changed after conversion, with academies reporting in the September-August school year as opposed to the April-March financial year.[95] The latter is in line with local authority financial statements and was the practice in schools before they converted to become an academy and in control schools throughout the period of the analysis. Secondly, accounts for schools that do not convert in the period (the control schools) do not include the value of LEA provided services; however, information is available on how much

[95] For a small number of schools (35), the accounts cover a period exceeding 12 months. In these cases, it is known how long the accounts cover and numbers for them are weighted accordingly (i.e. proportionately scaling down all items by the fraction 12/period covered).

extra income is given to academies to cover the value of these services (the Education Services Grant - ESG). To make the numbers comparable this is removed from both the grant income and expenditure for academies in column (2) of Table 4.4.

Columns (1) and (4) of Table 4.4 show per pupil income and expenditure was similar for the treatment and control schools before conversion. For example, as shown in Panel A, total income in all treatment and control schools was £3,974 and £4,156 per pupil respectively. Total expenditure was £3,966 and £4,154 per pupil in treatment and control schools respectively. As shown in Panels B to C of the Table, these pre-conversion numbers are also closely aligned for the comparisons undertaken within the outstanding and non-outstanding groups of schools.

It is evident, however, that converting primary schools both received more money and spent more money post-conversion, even once the extra money given for LEA provided services is accounted for. The Table also shows the income and expenditure per pupil after conversion and a difference-in-difference estimate in the final column. This shows significant income and expenditure gaps arising after conversion relative to what happened in the control schools. The differences in total income and expenditure are estimated as £296 and £522 per pupil per year. The increases are clearly driven by the relative increase in grant income. A similar qualitative pattern is shown for schools classified as outstanding and non-outstanding, but with higher income and expenditure shown for the latter schools, most likely reflecting a higher proportion of disadvantaged students in this group.

Table 4.5 shows the change in categories of expenditure per pupil before and after conversion.[96] There are three Panels, which differ according to assumptions made about which services the academies procure post-conversion given that they are no longer provided for them

[96] The detailed expenditure categories that have been aggregated to the four categories in Table 5 are reported in Appendix Table A1.

by the local authority. The numbers in the upper Panel A are changes inclusive of the extra money delegated to them. The numbers in the middle Panel B subtract an equal share of the ESG money from each category of expenditure. Finally, those in the lower Panel C remove all of the extra ESG money from the expenditure on non-staff related running costs.

In each case it is very clear that, even though primary academies spent more on teaching staff, non-teaching staff, and other running costs after conversion (relative to control schools), the increase was greater for administrative costs (i.e. non-teaching staff and other running costs). This is true for schools in all Ofsted categories. Because the amount of money earmarked for services previously provided by LEAs from expenditure is removed, these shifts cannot be attributed solely to the mechanical shift caused by the school having to take on more administrative tasks post-conversion. It seems that the primary academies studied in this paper did receive more income, but that they spent it disproportionately on day to day running operations rather than on 'frontline services' such as teaching staff.

4.4.3 Changes in Workforce

Table 4.6 reports evidence on changes in the composition of the school workforce between 2010/11 and 2014/15 for schools that became academies in that period relative to schools that became academies in 2015/16 and 2016/17. Changes are shown for all schools and stratified by Ofsted rating. The Table reports difference-in-differences estimates for the total number of teachers employed, the pupil/teacher ratio, the mean teacher salary, the proportion of teachers who are in the leadership group or whether the school changes its head teacher.

In general, the results reported in the Table show little evidence of workforce changes resulting from academisation. The one exception is headteacher turnover. For the full sample, there is a statistically significant 6.3 percentage point reduction in headteacher turnover in primaries that became academies. When broken down by Ofsted category, this occurs only in non-outstanding schools, which are 7.2 percentage points less likely to take on a new headteacher. This stands in direct contrast to the finding of Eyles and Machin (2015) who found that the vast majority of the first phase of academy conversions in the 2000s were characterised by new headteachers coming into academies and therefore that changes in managerial structure were a key feature of academy conversion that facilitated increased autonomy. This mechanism appears to be completely absent in the case of primary schools.

4.4.4 Changes in Intake

Alongside performance effects we look at whether pupil composition changed once a school gained academy status. As data is available prior to 2010/11, the analysis considers year-on-year changes between 2007/08 and 2014/15 in the characteristics of those entering the earliest grade in which the schools enrol pupils. We also include observations of pupils over this period that enter schools that convert out of sample (in 2015/16 and 2016/17). The three outcomes considered are the fraction of the pupil intake who are eligible for free school meals, the fraction with English as a native language, and the total size of the entry year intake (in logs). In each case, school and year effects are included. The results, presented in Table 4.7, show no evidence that schools alter their intake along these dimensions.[97]

4.4.5 Summary

Taken together, these findings suggest that primary schools did change some aspects of their operations after becoming academies. Most primary academies began to use freedoms made available to them because of conversion. They also received more income and altered how their expenditure was allocated across functions. Regarding the latter, the spending changes made were mainly to affect administrative functioning and day to day operations,

[97] This contrasts with the findings on the first batch secondary school academies reported in Eyles and Machin (2015), where intake changed significantly. In addition to the outcomes used here we also created an index by regressing 2006/07 KS2 scores on first language, gender, ethnicity, and free school meals status. Using these coefficients, we then predicted a KS2 score for the incoming pupils in later years, standardised this, and used it as an outcome. The results using this measure are 0.001 (0.011), -0.003 (0.021), and 0.002 (0.013) for all schools, outstanding schools, and non-outstanding school respectively.

because of the removal of such provision from the local authority. At the same time, there was not much change in the school personnel or in composition of the pupil intake.

4.5 Pupil Performance Results

This section reports the results on pupil performance, starting with the main baseline set of results showing the causal impact of academy conversion on pupil performance. Then, in the light of the previous section's results showing that most, but not all, primary schools altered their modes of operation post-conversion, heterogeneous estimates along several dimensions are reported.

4.5.1 Main Results

Table 4.8 shows estimates of the 2SLS specifications studying the impact of academisation on pupil performance in reading and maths in tests at the end of primary school. Separate coefficients are shown for each subject, both for the pooled sample and by whether the predecessor school's Ofsted grade was outstanding or not. Columns (1) to (3) show estimates when the treatment is whether the school converts to academy status. Columns (4) to (6) show estimates for years of exposure.

As most legacy enrolled pupils stay in the school to take their KS2 exams - first stage estimates range from 0.92 to 0.95 - we have only presented 2SLS estimates.[98] In all cases, there is no evidence of any performance boost from academisation. The estimates are small in magnitude, sometimes negative, and almost all statistically insignificant. In terms of magnitude, the largest positive estimate is 0.02σ (with standard error 0.03) for reading in outstanding schools as reported in specification (2) of the Table. All the other 2SLS estimates

[98] We also looked at mobility between grades 2 and 6 and how it differs between treatment and control schools. Using pupil mobility as an outcome variable there is no differential transfer between the two sets of schools. Running this on all schools, and on the Ofsted groupings, detected no significant differences.

are lower than this, and nine of the estimates (including all six for maths) have negative signs. When considering the average of reading and maths scores it seems that primary age pupils did not benefit from attending an academy school in terms of their performance at the end of primary school. As results are similar whether we consider reading or math marks as the outcome, we focus on average points from this point onwards[99].

One might be concerned about the research design being potentially contaminated by differential pre-policy trends[100]. Figure 4.2 therefore shows estimates from an event study, for the pooled sample, for pupils attending academies four years prior to academy conversion to three years after. The effects of being in an academy remain numerically small and insignificant (as the c to c+3 coefficients all overlap with the zero line on the Figure). Moreover, there is no sign of pre-policy trends, nor any gradual improvement in results post-conversion. Figure 4.2.

Table 4.9 also further generalises the Table 4.8 baseline results by reporting estimates for legacy enrolled pupils by discrete years of exposure, ranging from one to a maximum of four. Again, there is neither any sign of a positive effect nor any suggestion that benefits might be increasing with years of exposure. If anything, the opposite is the case, as the absolute values of the negative coefficients mostly get larger with more years of exposure.

4.5.2 Heterogeneity

While there is no evidence of performance effects on average, nor in the event study and years of exposure analysis, it may still be the case that academisation has scope to benefit some subsets of students and not others. It is also possible that certain school characteristics may be associated with differential academy effects on pupil performance.

---

[99] Results for tables 8-10 for maths and reading are available in the appendix.
[100] A second concern flagged by referees was the potential for spillovers between treated and control schools. To deal with this we re-estimated our main regressions, but removed control schools that were within 3km of any treated school. Our results were unaffected by this change.

Table 4.10 therefore shows results from investigations of whether the effect size differs in several ways: i) with whether the pupil is eligible for free school meals or not; ii) with an indicator for whether the school is in an urban area or not (given that the charter school literature finds positive effects to be concentrated amongst urban schools); iii) whether it differs with pre-conversion school size (as larger schools may be more adept at managing their extra freedoms); and iv) whether the school joins a multi-academy trust (MAT).

The results reported in the Table do little to alter the prior analysis. First, there is little evidence that the effect of academy attendance differs depending on whether one is eligible for free school meals or attends an urban academy. Panel C of Table 4.10, shows that the same can be said for pupils attending schools of differing sizes. Although performance effects appear to decline with school size, none of these interactions reach statistical significance.

The final aspect of heterogeneity considered – whether pupils attend an academy that becomes part of a (MAT) or not – does uncover some differences. The most noteworthy is that some of the estimates for not being in a MAT are significantly negative. This is the case for all schools where there is a 0.06σ (0.02) fall – closer investigation shows that this is confined to the non-outstanding schools. Although this suggests that conversion in stand-alone (non-MAT) schools, which are not able to benefit from the economies of scale that a MAT brings, may have proven detrimental to pupils enrolled in previously non-outstanding academies, this result should be taken with caution. About 60% of the primary academies considered here are part of a MAT, but it should be acknowledged that whether a school is able to join a trust is endogenous to KS2 performance; results showing performance drops could be due to negative selection of non-outstanding schools that are not part of a MAT.

4.6 Conclusion

The English government has radically restructured its school system under an assumption that academisation delivers benefits to schools and students. This paper reports results from investigations studying the totally unexpected policy change that occurred in 2010 that enabled (and encouraged) primary schools to become academies. It looks at the first primary schools that have become academies in England (between 2010/11 and 2014/15) and finds no evidence of pupil performance improvements resulting from conversion.

How should an overall zero effect be interpreted in the light of some evidence showing positive effects of autonomy in other contexts? One reason is that schools that converted were already doing well within the system and simply did not require additional autonomy to thrive and therefore did not make substantive changes. Indeed, the limited changes that are seen – increasing expenditure on non-instructional tasks – do not correspond to the kinds of changes, such as effective discipline and higher quality teaching that have been found to increase test scores in other contexts such as charter schools (Fryer, 2014).

In existing research, much of the positive effects of autonomous schools have been shown for disadvantaged students and not so much for advantaged students. While there was scope to improve achievement within these schools, it may be that changes introduced because of school autonomy simply do not benefit such students at the margin. However, given the survey evidence reported above and the research into how additional income was used by schools, many of these schools did not make changes that affect 'frontline services' (as opposed to administrative roles).

Another possible reason is that effects are estimated in the short run. It may be that the programme will bear fruit once more schools convert and facilitate greater economies of scale by entering or deepening collaborative arrangements with each other. In the heterogeneity analysis, we found some evidence of variation by whether schools are in a multi-academy trust.

Although we do not take the effect to be causal due to the endogenous decision to join a MAT, it is still a worrying finding that performance dips for the non-outstanding primary schools (around 40 percent of converters) that do not gain scale economies from being part of a multi-academy trust.

Finally, one of the key models for some successful urban charters in the US and some secondary schools in England[101] – an effective discipline approach for academies and the No Excuses model of charters – is of less relevance to the age range of children enrolled in English primary schools than for secondary age children (since behavioural problems that may lead pupils to be suspended or excluded from school are much more prevalent in the latter).[102] In the light of all these factors, it is not surprising that there has been no overall effect on pupil performance.

One might argue that if academisation has no average effect on pupil performance, this could still be a reasonable public policy if there are other reasons for why this might be beneficial – for example, if school leaders can more easily make changes that might benefit students (or their parents) and staff. However, the process of restructuring individual schools has been shown to be financially costly and restructuring on a system wide basis would likely prove to be too costly in the long run if it fails to generate gains for students in terms of test scores. Furthermore, risks are also posed by an increasing number of schools becoming academies.[103] For example, they are no longer regularly monitored at the local level. Problems might not therefore come to light unless they are flagged up by an Ofsted inspection, which are not regular events. There are potential negative spillovers on other schools if opting out of

[101] A well-known, and highly publicised, example of the latter is Hackney's Haggerston School which is a secondary school has fully utilised an effective discipline and good behaviour approach in its successful rise up the KS4 achievement distribution, despite having a relatively disadvantaged pupil intake.
[102] For instance, exclusions and fixed term suspensions are extremely rare in the age range that we study. In English schools, 3.88% of pupils received a fixed period exclusion in 2014/15, and 0.07% were permanently excluded. For primary schools the numbers were much lower - 1.1% of pupils received a fixed term exclusion and 0.02% were permanently excluded in the school year 2014/15 (Department for Education, 2016).
[103] See Ladd and Fiske (2016).

Local Authority control undermines services that the Local Authority is able to provide to other schools in the same geographic area (e.g. child psychologists to support children with special needs in many schools). Studying the operational aspects of academies, and the institutional structures in which they function, is an important subject for future research.

References

Abdulkadiroglu A., J. Angrist, S. Dynarski, T. Kane and P. Pathak (2011) Accountability and Flexibility in Public Schools: Evidence From Boston's Charters and Pilots, *Quarterly Journal of Economics*, 126, 699-748.

Abdulkadiroglu A., J. Angrist, P. Hull and P. Pathak (2016) Charters without Lotteries: Testing Takeovers in New Orleans and Boston, *American Economic Review*, 106, 1878-1920.

Academies Commission (2013) Unleashing Greatness: Getting the base from an academised system.
http://www.educationengland.org.uk/documents/pdfs/2013-academies-commission.pdf

Angrist J., S. Cohodes, S. Dynarski, P. Pathak and C. Walters (2016) Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry and Choice, *Journal of Labor Economics*, 34, 275-318.

Angrist J., S. Dynarski, T. Kane, P. Pathak and C. Walters (2010) Inputs and Impacts in Charter Schools. KIPP Lynn, *American Economic Review*, 100, 239-43.

Angrist J., P. Pathak and C. Walters (2013) Explaining Charter School Effectiveness, *American Economic Journal: Applied Economics*, 5, 1-27.

Bassett, D., G. Lyon, W. Tanner, B. Watkin (2012) Plan A+ Unleashing the Potential of Academies. London: Reform and Schools Network.
http://www.reform.uk/publication/plan-a-unleashing-the-potential-of-academies/

Bertoni, M., S. Gibbons and O. Silva (2017) What's in a Name? Expectations, Heuristics and Choice During a Period of Radical School Reform, Centre for Economic Performance Discussion Paper 1477.

Betts J., L. Rice, A. Zau, Y. Tang and C. Koedel (2006) Does School Choice Work? Effects on Student Integration and Achievement, Public Policy Institute of California.

Bolhmark A. and M. Lindahl (2015) Independent Schools and Long-run Educational Outcomes: Evidence from Sweden's Large-scale Voucher Reform, *Economica*, 82, 508–551.

Clark, D. (2009) The Performance and Competitive Effects of School Autonomy, *Journal of Political Economy*, 117, 745-83.

Center for Research on Education Outcomes (2009) Multiple Choice: Charter Performance in Sixteen States, Stanford University, CREDO.

Center for Research on Education Outcomes (2013) National Charter School Study, Stanford University, CREDO.

Cirin, R. (2014) Do Academies Use Their Autonomy? Department for Education Research Report.

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/401455/RR366_-_research_report_academy_autonomy.pdf

Department for Education (2013) Academies Annual Report. Academic Year: 2011/12. London: The Stationary Office.

Department for Education (2016) Permanent and Fixed Period Exclusions in England: 2014 to 2015, National Statistics, SFR 26/2016, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/539704/SFR_26_2016_text.pdf

Dobbie W. and R. Fryer (2011) Are High Quality Schools Enough to Close the Achievement Gap? Evidence From a Social Experiment in Harlem, *American Economic Journal: Applied Economics*, 3, 158–87.

Dobbie W. and R. Fryer (2013) Getting Beneath the Veil of Effective Schools: Evidence from New York City, *American Economic Journal: Applied Economics*, 5, 58–75.

Dobbie W. and R. Fryer (2014) The Medium Term Impacts of High Achieving Charter Schools, *Journal of Political Economy*, 5, 985–1037.

Dobbie W. and R. Fryer (2016) Charter Schools and Labor Market Outcomes, National Bureau of Economic Research Working Paper 22502.

Epple D ., R. Romano and R. Zimmer (2015) Charter Schools: A Survey of Research on Their Characteristics and Effectiveness, National Bureau of Economic Research Working Paper 21256.

Eyles, A., C. Hupkau and S. Machin (2016a) Academies, Charter and Free Schools: Do New School Types Deliver Better Outcomes?, *Economic Policy*, 31, 453-501.

Eyles, A., C. Hupkau and S. Machin (2016b) School Reforms and Pupil Performance*, Labour Economics*, 41, 9-16.

Eyles, A., and S. Machin (2015) The Introduction of Academy Schools to England's Education, CEP Discussion Paper. No. 1368.

Fryer, R. (2014) Injecting Charter School Best Practices into Traditional Public Schools: Evidence From Field Experiments, *Quarterly Journal of Economics*, 129, 1355-1407.

Gibbons, S., S. McNally and M. Viarengo (2011) Does Additional Spending Help Urban Schools? An Evaluation Using Boundary Discontinuities. SERC Discussion Paper No. 90; CEE Discussion Paper No. 128. London School of Economics.

Gleason P., M. Clark, C. Clark Tuttle, E. Dwyer and M. Silverberg (2010) The Evaluation of Charter School Impacts (NCES 2010-4029), U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Hanushek E. and L. Woessmann (2011) The Economics of International Differences in Educational Achievement, in Hanushek E., S. Machin and L. Woessmann (eds.) Handbook of the Economics of Education, Volume 3, Amsterdam: Elsevier.

Holmlund, H., S. McNally and M. Viarengo. (2010) Does Money Matter for Schools? *Economics of Education Review*. 29, 1154-1164.

Hoxby C. and S. Murarka (2009) Charter Schools in New York City: Who Enrolls and How They Affect Student Achievement, National Bureau of Economic Research Working Paper 14852.

Hoxby C., S. Murarka and J. Kang (2009) How New York City's Charter Schools Affect Achievement, New York City Charter Schools Evaluation Project

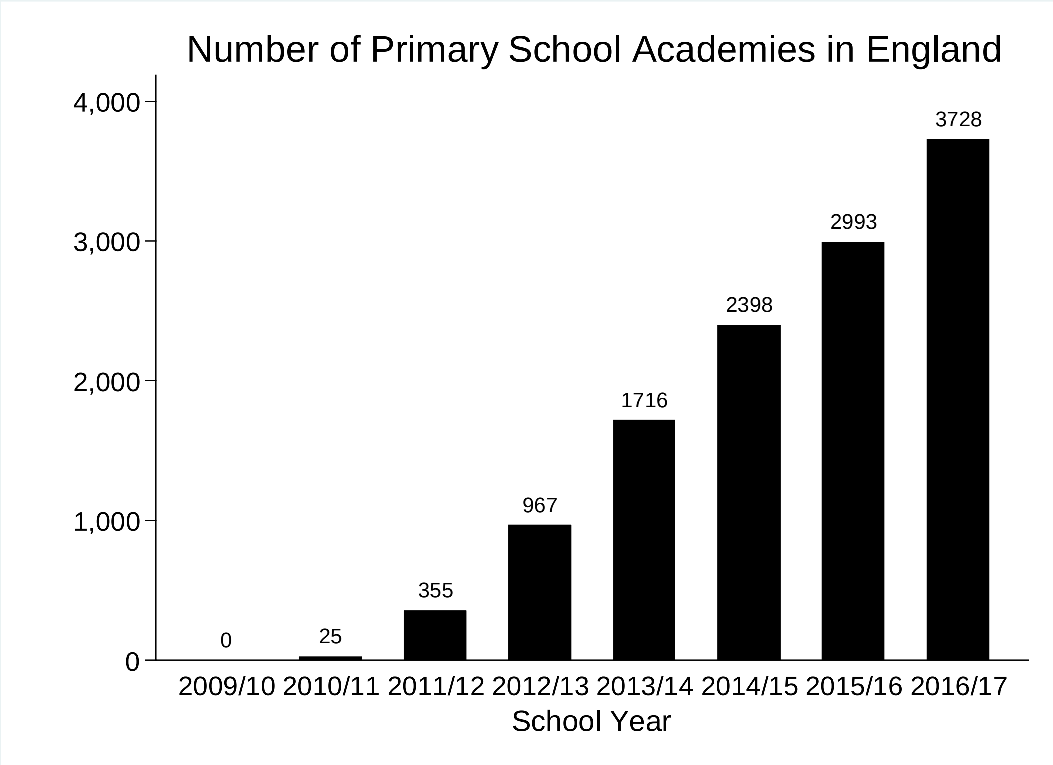
Ladd H. and E. Fiske (2016) England Confronts the Limits of School Autonomy, Working Paper 232 National Center for the Study of Privatization in Education Teachers College, Columbia University

National Audit Office (2012). Managing the Expansion of the Academies Programme. London: The Stationary Office.

Roberts N. and P. Bolton (2017) School Funding in England. Current System and Proposals for 'Fairer School Funding, House of Commons Briefing Paper Number 06702

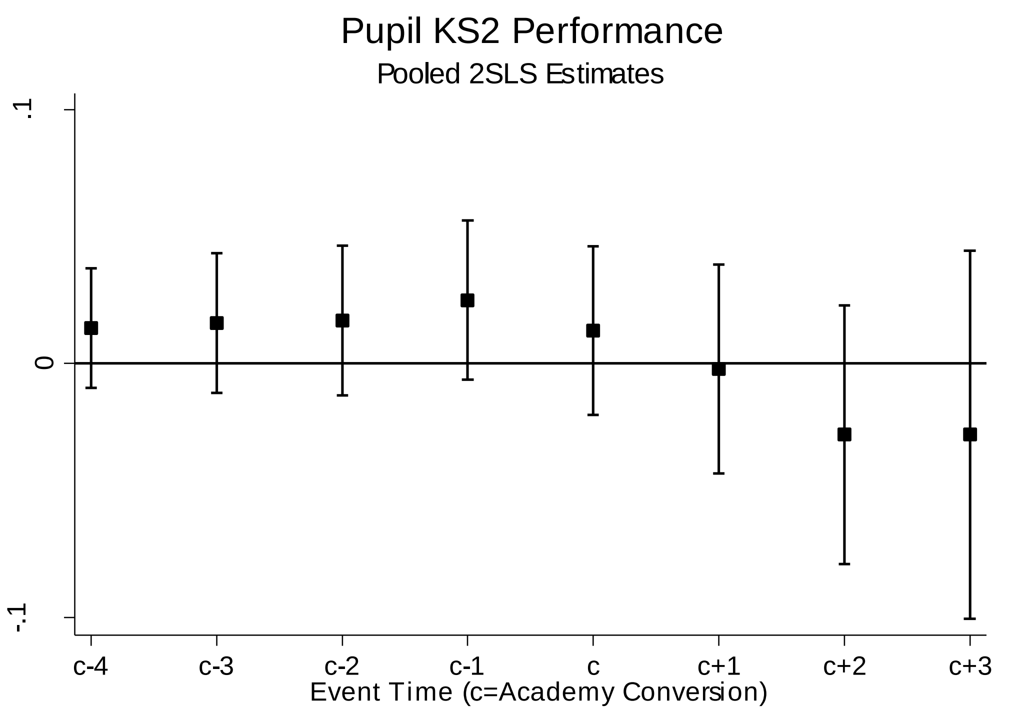
Steinberg M. (2014) Does Greater Autonomy Improve School Performance? Evidence from a Regression Discontinuity Analysis in Chicago. *Education Finance and Policy,* 9(1), 1-35.

Figure 4.1: Number of Primary School Academies in England



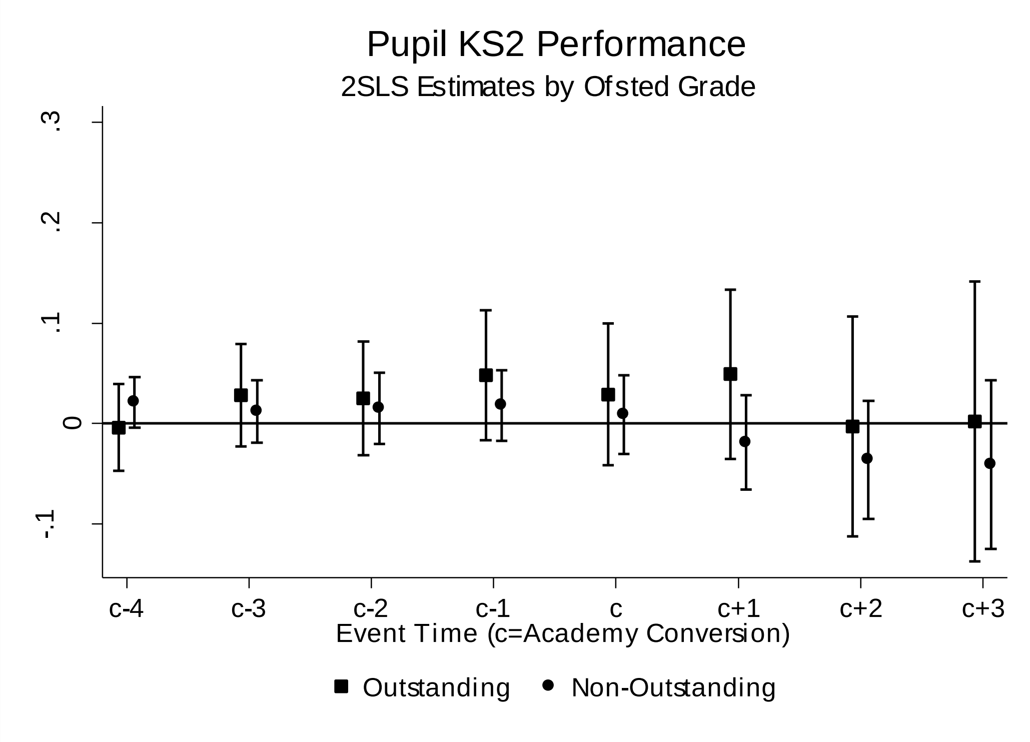
Notes: The Figure shows the number of primary/ middle deemed primary schools open in each school year as academies, and/or free schools. Source data Edubase, available at http://www.education.gov.uk/edubase/home.xhtml

Figure 4.2: Event Study Estimates, Pre- and Post-Academy Conversion



Notes: c refers to academy conversion year. KS2 performance is measured by standardised average point score. The coefficients come from the same 2SLS estimates as reported in column (1) of Table 4.8, but with dummies for the number of years before or after conversion the exam is sat in an academy. The four post conversion dummies (c to c+3) are instrumented for with four ITT/ITT grade interactions. A joint test for the significant of the pre-conversion dummies gives a chi square statistic of 0.71 (p-value = 0.59).

Figure 4.3: Event Study Estimates by (Pre-Intervention) Ofsted Grade



Notes: c refers to academy conversion year. KS2 performance is measured by the standardised average point score. The coefficients come from the same 2SLS estimates as reported in columns (2) and (3) of Table 4.8, but with dummies for the number of years before or after conversion the exam is sat in an academy. The four post conversion dummies (c to c+3) are instrumented for with four ITT/ITT grade interactions. A joint test for the significant of the pre-conversion dummies gives a chi square statistic of 1.26 (p-value = 0.28) in the case of outstanding schools and 0.79 (p-value = 0.60) in the case of non-outstanding.

Table 4.1: Number of New Primary Converter Academies in the Study Sample

	Academic Year						
	Treatment Schools, 2010/11 to 2014/15					Control Schools, 2015/16 and 2016/17	
	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17
Pooled	12	174	252	234	243	210	309
Outstanding	10	74	51	40	35	30	35
Non-Outstanding	2	100	201	194	208	180	274

Notes: In order to implement the research design, only schools in the sample that have students in grades 2 and grades 6 in each academic year between 2006/07 and 2014/15 are included. The main discrepancy between the numbers in this Table and the total number of primary academies given in Figure 1 arise because of: a) the removal of infant and junior schools (the latter do not to enrol children in grade 2, while the former do not do so in grade 6) and b) because Figure 1 also includes sponsored academies (comprising around 30 percent of primary academies) and a small number of free schools (139 by 2016/17), which are not studied in this paper.

Table 4.2: Baseline Characteristics: Pooled Sample and by Ofsted Grade

	All Schools			Outstanding Schools			Non-Outstanding Schools		
	Treatment	Control	Treatment – Control p-value	Treatment	Control	Treatment – Control p-value	Treatment	Control	Treatment – Control p-value
English is first language	0.924	0.926	0.858	0.908	0.911	0.808	0.929	0.929	0.964
White British	0.877	0.881	0.609	0.855	0.854	0.989	0.883	0.886	0.543
Eligible to receive free school meals	0.126	0.133	0.336	0.108	0.109	0.889	0.131	0.138	0.309
Male	0.512	0.509	0.073	0.509	0.506	0.279	0.512	0.509	0.141
KS2 reading	0.049	0.003	0.418	0.298	0.28	0.647	-0.025	-0.048	0.496
KS2 maths	0.06	-0.010	0.032	0.334	0.296	0.300	-0.021	-0.066	0.06
KS1 reading	0.03	0.003	0.721	0.186	0.173	0.732	-0.016	-0.028	0.817
KS1 maths	0.024	-0.002	0.727	0.172	0.163	0.826	-0.020	-0.032	0.777
Number of teachers	14.958	14.294	0.020	16.443	15.419	0.087	14.516	14.09	0.093
Proportion unqualified teachers	0.033	0.03	0.484	0.039	0.036	0.587	0.031	0.029	0.624
Number of pupils	278.709	264.89	0.018	313.448	289.825	0.052	268.362	260.357	0.109
Pupil teacher ratio	18.555	18.322	0.231	19.179	18.843	0.402	18.369	18.228	0.372
Mean teacher age	41.077	41.483	0.034	40.276	40.413	0.63	41.322	41.683	0.032
Mean teacher salary	32412	32545	0.034	32568	32545	0.863	32364	32545	0.046
Number of schools	1434			275			1159		

Notes: All variables are measured in the school year 2006/07. All KS1 and KS2 scores are standardized to have mean zero and standard deviation of 1 (within the year and overall sample). Ofsted grades are measured prior to the policy. Since Ofsted inspect schools every 3-5 years (see Section 3), the grades here are the most recent grade between 2006/07 and 2009/10 (i.e. prior to the policy change).

Table 4.3: Percentage Using Freedoms Since Becoming an Academy: Primary and Secondary Schools

	Secondary Schools	Primary Schools
Changed your pattern of capital expenditure	63	54
Introduced savings in back-office functions	62	54
Changed the performance management system for teachers	63	49
Changed the curriculum you offer	60	49
Changed school leadership	51	43
Introduced or increased revenue-generating activities	41	28
Hired teachers without qualified teacher status (ATS)	23	8
Sought to attach pupils from a different geographical area	14	5
Increased the length of the school day	10	5
Changed the length of school terms	6	2
Number of schools	360	334

Source: Cirin (2014). Online survey of 720 academies that were open on 1 May 2013.

Table 4.4: Changes in School Income per Pupil and Expenditure per Pupil Before and After Academy Conversion

	Treatment Schools			Control Schools			Treatment – Control
	Before	After	Change	Before	After	Change	Difference-in-Difference
	(1)	(2)	(3) = (2) – (1)	(4)	(5)	(6) = (5) – (4)	(7) = (3) – (6)
A. All Schools (843 Treatment, 466 Control)							
Total income	3974	4997	1023 (26)	4156	4883	727 (41)	296 (48)
Grant income	3810	4771	961 (24)	4019	4704	685 (40)	276 (47)
Other income	164	226	63 (10)	137	179	42 (7)	20 (12)
Total expenditure	3966	5121	1155 (33)	4154	4788	633 (43)	522 (54)
B. Outstanding (200 Treatment, 59 Control)							
Total income	3755	4807	1052 (47)	3851	4819	967 (199)	85 (203)
Grant income	3580	4559	979 (43)	3706	4598	892 (199)	88 (202)
Other income	175	248	73 (23)	145	221	76 (27)	-2 (35)
Total expenditure	3754	4890	1135 (59)	3834	4754	920 (201)	215 (208)
C. Non-Outstanding (643 Treatment, 407 Control)							
Total income	4042	5056	1014 (30)	4200	4892	692 (37)	322 (48)
Grant income	3882	4837	955 (29)	4064	4719	655 (36)	300 (46)
Other income	160	220	59 (11)	136	173	37 (6)	22 (13)
Total expenditure	4032	5193	1161 (39)	4201	4792	592 (39)	569 (55)

Notes: The sources for expenditure data are publicly available consistent financial reporting records for all state-maintained schools and academies financial benchmarking data for academy schools. The former are available at https://www.compare-school-performance.service.gov.uk/ and the latter can be accessed at https://www.gov.uk/government/collections/statistics-local-authority-school-finance-data. For academies opening in April to August of the school year, incomes and expenditures in the first full year of conversion are appropriately scaled. In columns (3), (6) and (7), long changes are considered between 2009/10 (Before) and 2014/15 (After). Standard errors in parentheses.

Table 4.5: Changes in Expenditure Category per Pupil Before and After Academy Conversion

	All Schools		Outstanding		Non-Outstanding	
	Pre-Change Mean	Difference-in-Difference	Pre-Change Mean	Difference-in-Difference	Pre-Change Mean	Difference-in-Difference
	(1)	(2)	(3)	(4)	(5)	(6)
A. Includes ESG						
Total teaching staff	2063	58 (27)	1969	-43 (104)	2086	65 (27)
Total non-teaching staff	1236	168 (24)	1113	10 (66)	1266	197 (26)
Learning and ICT resources	212	-8 (8)	216	-3 (22)	211	-10 (9)
Other running costs	523	305 (22)	475	251 (47)	535	318 (25)
B. ESG Equally Deducted						
Total teaching staff	2063	32 (27)	1969	-75 (104)	2086	40 (27)
Total non-teaching staff	1236	148 (24)	1113	-12 (66)	1266	178 (26)
Learning and ICT resources	212	-12 (8)	216	-8 (22)	211	-13 (9)
Other running costs	523	294 (22)	475	239 (47)	535	308 (25)
C. ESG Deducted From Other Running Costs						
Total teaching staff	2063	58 (27)	1969	-43 (104)	2086	65 (27)
Total non-teaching staff	1236	168 (24)	1113	10 (66)	1266	197 (26)
Learning and ICT resources	212	-8 (8)	216	-3 (22)	211	-10 (9)
Other running costs	523	245 (22)	475	181 (49)	535	262 (26)
Number of treatment schools	843		200		643	
Number of control schools	466		59		407	

Notes: As for Table 4.4. The top panel includes extra money given to academies to cover services previously provided by the LEA. The middle panel removes this expenditure equally from each expenditure category. The bottom panel removes all the extra money from the other running costs category.

Table 4.6: Changes in Workforce, School-Level Difference-in-Differences Estimates

	All Schools	Outstanding	Non-Outstanding
	(1)	(2)	(3)
	0.007	0.015	0.007
Log(Number of teachers)	(0.012)	(0.025)	(0.014)
	-0.018	-0.007	-0.016
Log(Pupil teacher ratio)	(0.011)	(0.023)	(0.012)
	-0.003	0.001	-0.006
Log(Mean teacher salary)	(0.007)	(0.013)	(0.008)
Proportion of teachers in	0.003	0.005	0.003
leadership group	(0.005)	(0.010)	(0.005)
	-0.063	-0.002	-0.072
Change in headteacher	(0.028)	(0.070)	(0.031)

Notes: Based on data from the schools' workforce census for the academic years 2010/11 and 2014/15. All variables are long changes between these two academic years. The subsample is the sample of schools who are observed in each of the two years. We exclude schools converting in 2010/11 as we do not observe a pre-treatment observation. Robust standard errors in parentheses are reported in each case. The sample sizes for the first three rows (Number of teachers; Pupil teacher ratio; proportion of teachers in leadership group) are 1326, 254, 1072 for all schools, outstanding schools, and non-outstanding schools respectively. For the headteacher regression the sample sizes are 1327, 257 and 1070 for all schools, outstanding schools, and non-outstanding schools respectively. Baseline means are: 15.321 teachers; 21.565 pupils per teacher; £36446 average salary; 0.173 of teachers are in the leadership group 0.445 of schools change headteacher over the course of the five years.

Table 4.7: Changes in Pupil Intake

	All Schools			Outstanding			Not-Outstanding		
	FSM	English Language	Log (No of Pupils)	FSM	English Language	Log (No of Pupils)	FSM	English Language	Log (No of Pupils)
Academy X Post-Conversion	0.000	-0.007	0.007	0.002	0.004	0.009	-0.001	-0.011	0.006
	(0.003)	(0.009)	(0.006)	(0.004)	(0.019)	(0.012)	(0.003)	(0.011)	(0.007)
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	467386	466635	12906	102565	102459	2475	364821	364176	10431
Number of schools	1434	1434	1434	275	275	275	1159	1159	1159

Notes: Variables refer to the pupils entering the lowest grade in the school in each year. Each cell is a coefficient estimated from a separate regression. Standard errors (in parentheses) are clustered at school level

Table 4.8: The Effect of Treatment on KS2 Test Scores (measured at age 11)

	(1) Pooled	(2) Outstanding	(3) Non-Outstanding	(4) Pooled	(5) Outstanding	(6) Non- Outstanding
Maths	-0.021	-0.002	-0.027	-0.012	-0.004	-0.016
	(0.014)	(0.030)	(0.016)	(0.007)	(0.014)	(0.008)
Reading	0.001	0.020	-0.005	-0.004	0.005	-0.007
	(0.013)	(0.026)	(0.014)	(0.006)	(0.012)	(0.007)
Average Point Score	-0.013	0.008	-0.021	-0.010	-0.001	-0.014
	(0.014)	(0.029)	(0.016)	(0.007)	(0.014)	(0.008)
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	1636948	296675	1340273	1636948	296675	1340273
Number of schools	1434	275	1159	1434	275	1159
First stage	0.937	0.950	0.932	0.928	0.942	0.922
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)

Notes: Each cell is a coefficient estimated from a separate regression. Full controls are included (for gender, ethnicity, speaks English as first language, eligible for free schools meals, prior attainment (Key Stage 1). Standard errors (in parentheses) are clustered at school level.

Table 4.9: Effects by Year of Exposure

	All Schools	Outstanding	Non-Outstanding
	Average Points	Average Points	Average Points
One year of exposure	0.001 (0.012)	0.011 (0.024)	-0.001 (0.014)
Two years of exposure	-0.015 (0.016)	0.030 (0.030)	-0.030 (0.019)
Three years of exposure	-0.042 (0.022)	-0.024 (0.045)	-0.049 (0.025)
Four years of exposure	-0.042 (0.033)	-0.020 (0.060)	-0.053 (0.039)
School fixed effects	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
Sample size	1636948	296675	1340273
Number of schools	1434	275	1159
First stage coefficient on IIT x one year of exposure	0.963 (0.001)	0.974 (0.002)	0.960 (0.002)
First stage coefficient on IIT x two years of exposure	0.931 (0.002)	0.946 (0.003)	0.926 (0.003)
First stage coefficient on IIT x three years of exposure	0.902 (0.004)	0.927 (0.005)	0.891 (0.005)
First stage coefficient on IIT x four years of exposure	0.869 (0.007)	0.889 (0.008)	0.857 (0.011)

Notes: 2SLS estimates reported. Full controls are included (for gender, ethnicity, speaks English as first language, eligible for free schools meals, prior attainment, primary school). Standard errors (in parentheses) are clustered at school level.

Table 4.10: Heterogeneity

	All Schools	Outstanding	Non-Outstanding
	Average Points	Average Points	Average Points
A. Free school meal eligibility			
Yes	0.008 (0.021)	0.025 (0.045)	0.004 (0.024)
No	-0.017 (0.014)	0.006 (0.029)	-0.025 (0.015)
B. Urban			
Yes	-0.012 (0.015)	0.019 (0.031)	-0.023 (0.018)
No	-0.021 (0.019)	-0.037 (0.038)	-0.013 (0.021)
C. Baseline school size			
Treatment	0.108 (0.103)	0.278 (0.166)	0.065 (0.128)
Treatment*Baseline school size	-0.021 (0.018)	-0.046 (0.029)	-0.015 (0.023)
D. Multi academy trust			
Yes	0.030 (0.016)	0.032 (0.033)	0.031 (0.019)
No	-0.057 (0.017)	-0.013 (0.034)	-0.075 (0.020)

Notes: 2SLS estimates comparable to columns (1), (2) and (3) of Table 8, but with mutually exclusive interactions included for panels A, B, and to D. Panel C reports estimates that interact treatment status with baseline (2006/07) school size in logs. In terms of free school meal status, 14% of pupils in the treated schools are eligible, 12% of pupils in outstanding treated schools are eligible, and 15% of pupils in non-outstanding treated. For all treated schools in the sample 71% are in urban areas and 57% are in multi-academy trusts. The same numbers for outstanding and non-outstanding schools are: urban 72%/71%; multi-academy trusts 50%/59% respectively.

Chapter 5: "Super Heads", School Autonomy and Leadership Pay

Abstract

This paper assesses the contribution of school autonomy – in particular, discretion in setting remuneration levels – to leadership salary differentials and inequality in English schools. It first shows there are head teacher salary differentials between academy schools (which have more autonomy) and traditional state schools, but that a large part is attributable to observable teacher and school characteristics. By contrast, the majority of the observed increase in head teacher salary inequality remains unexplained by the growth of academies. The emergence of "super heads", whose high levels of remuneration have driven the inequality rise, is not a direct consequence of the academisation of English schools. Instead, it reflects part of a more general shift towards the market determination of salaries in the English secondary school sector.

5.1 Introduction

A perennial research topic over the years has been the study of compensation received by individuals heading up organisations. This has been studied in depth for chief executives running private sector companies, often with a focus on whether high remuneration is justified by corporate performance.[104] At the same time, many questions have been asked about rising inequality of compensation of those at the top, especially in research on the top 1 percent, but also more generally.[105] However, because compensation structures have traditionally been more rigid in public sector organisations, and often tied to salary scales, much of this work has focused on the private sector, and often on the largest publicly quoted companies in various countries.

However, in some state sector settings, salary payments made to those running organisations have become more flexible. One reason for this is that the role of collectively bargained pay arrangements has diminished and market based compensation has become more commonplace. This is certainly the case in the context of this paper, where the determination of head teachers' salaries in secondary schools in England has been entirely removed from the national pay setting structures that used to operate. Salaries are now devolved to being set at the level of the school.

This decentralization of salary setting has gone hand-in-hand with rapidly rising inequality. Figure 5.1 plots (real) head teacher salaries over eighteen years (between school years 1995/96 and 2013/14), at three points of the overall distribution – the 10th, 50th and 90th percentiles. It shows a very clear fanning out of the distribution over time, with much faster growth seen at the 90th percentile (57% growth between 1995/96 and 2013/14) than at the 50th

---

[104] See, for example, the seminal Jensen and Murphy (1990) CEO pay for performance paper, with further developments linking to career concerns of CEOs by Gibbons and Murphy (1992) and the *Handbook of Labor Economics* chapter on executive compensation by Murphy (1999).

[105] On the top 1 percent see, inter alia, Piketty (2014) or Atkinson (2015). Discussion of more general trends in labour market inequality, and their drivers, are considered in the survey piece by Acemoglu and Autor (2010).

percentile (47%), which in turn grew significantly faster than the 10^{th} percentile (36%). The 90-10 salary ratio then rises by 15 percent over the full eighteen years, going from 1.35 to 1.55. Moreover, head teacher salary levels at the top end of the distribution are high in relative terms. In the 2013/14 school year, the 90^{th} percentile head teacher was paid an annual salary of £111,422 – or 4.1 times the median full time worker's annual salary of £27,195 in the UK's 2014 Annual Survey of Hours and Earnings (ASHE).[106] Back in 2000, the 90^{th} percentile head teacher was paid 3.2 times the median full time worker's salary.

Salaries at such levels have prompted media discussions about the emergence of "super heads". Indeed, in related work studying head teachers as public sector CEOs, Besley, Machin and Telhaj (2018) show empirical evidence that head teacher salaries in England have become more closely tied to school performance as salary setting has become considerably more flexible following the reforms that took place at the start of the 2000s. An important question that follows from this is whether the rise of the super head is a direct consequence of the increased autonomy some schools have to set salaries flexibly in the absence of collective bargaining.

This paper looks at this question and, in doing so, is the first to examine the link between school autonomy and leadership salaries. Greater autonomy has diffused into the English education system from a new type of school – the academy school – which first appeared on the English education landscape almost twenty years ago. Academies are autonomous schools that operate outside of local government control (see Eyles and Machin, 2019, or Eyles, Machin and Silva, 2018). Since the first academies were introduced in 2001/02 and, following the massive acceleration of academisation since 2009/10, the English secondary school sector has changed from one where the majority of schools were previously controlled by local education

[106] The Annual Survey of Hours and Earnings (ASHE) is a representative employer reported survey of one percent of the UK workforce carried out in April each year.

authorities (LEAs) to one in which the majority (68% in 2017)[107] of schools directly receive state funding and are run independently of local government.[108]

Academies have many freedoms related to the greater autonomy available to them – these include employing all academy staff, agreeing their levels of pay and conditions of service, changing the school curriculum, procuring services and managing budgets that were previously done by local authority, changing school leadership and management, lengthening the school day and many others. A focus has been placed on head teacher remuneration in academy schools, at least in part because academies have discretion in setting head teacher pay. Although policymakers have lionised some academy leaders – one particularly successful 'super head' became the Chief Inspector of Schools in England in 2012 – there have been questions from other quarters as to whether some academy heads are compensated excessively relative to heads in other state-maintained schools.

One might expect a salary premium for academy heads if granting schools greater autonomy improves school quality. International evidence suggests that, at least in developed countries, the degree of operational autonomy enjoyed by schools is positively correlated with PISA test scores (Hanushek, Link and Woessmann, 2013). US charter schools' evidence also points to greater school autonomy raising student achievement (Epple, Romano and Zimmer, 2016).[109] However, little has been said about how school autonomy impacts upon the labour market for teachers and school leaders. Jackson (2012) does look at teachers in his study of changes in the local teacher labour market as a result of opening of a nearby charter school in North Carolina. Examining changes in salaries at traditional public schools after nearby charter

[107] This number is computed from the January 2017 national tables available at https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2017.
[108] As will be explained in Section 2.2, academies are still subject to some statutory guidelines when it comes to curriculum and admissions.
[109] See, for instance, the sizable literature on charter schools surveyed in Epple, Romano and Zimmer (2016).

school openings, he finds evidence that public schools increased teacher compensation to retain better quality teachers.

This paper presents evidence of significant head teacher salary differentials between heads of academy schools and traditional state schools. There is also salary variation between the different types of academies that have been introduced. However, a large part of these differentials is accounted for by observable teacher and school characteristics. Moreover, the majority of the observed increase in head teacher salary inequality remains unexplained by the growth of autonomous schools. As such, the emergence of super heads, whose high levels of remuneration have contributed to the rise in inequality, is not due to the academisation of English schools. Instead, it reflects a more general shift towards the market determination of salaries in the English secondary school sector. Put differently, it is the decentralization of head teacher salary setting that has led to many head teachers being remunerated much more flexibly, and increasingly being paid above the salary spines that used to operate. However, because this is true of academy and non-academy head teachers alike, the general marketization of salary determination is the key factor underpinning increased salary inequalities. Finally, and somewhat more speculatively, the reported results may be indicative of a growing emphasis on the idea that effective head teachers – so called 'super heads' – have become critical inputs to the production of education.

The rest of the paper is structured as follows. Section 2 discusses the rise of academies in the English secondary schooling system and documents how the determination of head teacher salaries has changed through time. Section 3 describes the empirical strategies used to study the changing structure of head teacher salaries, while the data description is provided in Section 4. Section 5 presents the core results, together with several robustness checks undertaken to establish causality, and Section 6 concludes.

5.2 Context

5.2.1 *Academy Schools in the English School System*

Academy schools were first introduced in the early 2000s by the then Labour government. Initially, they entered via a remedial programme for failing schools, in which already-existing schools were given to government approved 'sponsors' – groups from the private or voluntary sector – with the intention that they would improve standards. These first cohorts of sponsored academies enjoyed significantly more operational autonomy than the majority of other state-maintained schools.[110]

In the 2001/02 school year, the year preceding the introduction of academy schools, the bulk of non-selective state-funded schools were maintained and run by their local authority. Unlike these schools, new academies were granted responsibility for their own admissions[111], employment of staff, and curriculum – although they had to offer a 'broad and balanced curriculum' that included English, math, and science. Furthermore, academies did not have to contract out services to local authorities and were given control over their own budgets. By May 2010, there was a modest number of academies – 203 out of around 3,000 secondary schools - up and running. Evidence suggests that these sponsored academies did indeed raise pupil attainment and improved post-secondary outcomes (Eyles and Machin, 2019; Eyles, Hupkau and Machin 2016a, 2016b).

The election of the (Conservative and Liberal Democrat) coalition government in 2010, and the pursuant passage of the Academies Act 2010, drastically altered the English educational landscape. The Act widened the remit of the programme by allowing schools to

[110] By the vast majority, we are referring to voluntary-controlled and community schools that comprised 69.3% of secondary schools in 2001/02 (Department of Education and Skills, 2002). The New Schools Network (2015), give a more comprehensive breakdown of the English secondary sector and how different school types have differing levels of freedom from local authority control.
[111] Despite overseeing their own admissions, academies are subject to the school admission code that limits the criteria by which schools can ration places when oversubscribed.

voluntarily apply for academy status and gain autonomy without a sponsor.[112] Whereas the first iteration of the academies programme was remedial in nature – with schools deemed to be failing involuntary converted to academies – the post-2010 iteration was dominated by voluntary conversions known as 'converter academies'. These schools were rated 'outstanding' by the Office for Standards in Education (OFSTED) and fast-tracked towards converter academy status. Alongside converters, academisation with a sponsor continued to be seen as the main way to improve standards at struggling schools. Not only were schools involuntarily turned into sponsored academies, but schools that were not deemed suitable for conversion also had the option of gaining academy status alongside high performing schools as part of an academy chain or multi-academy trust (MAT).[113][114]

The passage of the Academies Act facilitated a large shift in the number of academies in the English secondary sector. In the sample of schools that we study, the share of academy schools rose at an extraordinary rate, going from 4% to 55% between school years 2009/10 and 2013/14 – see Appendix Table A5.1. The lion's share of this growth came from voluntary conversions, which made up 79% of academies as of 2013/14, rather than schools that were compelled to become sponsored academies.

That the majority of academies are voluntary conversions of high performing schools is seen in their characteristics. Table 5.1 shows differences between state schools and the three distinct academy types – the first batch of pre-2010 sponsored academies, post-2010 sponsored academies and voluntary academy conversions (converters). Looking at the final three columns, one can see that converter schools are characterized by higher test scores and fewer free school meal (FSM) eligible pupils than their state counterparts. They also employ head

[112] A second way that the Act changed the nature of academies was to schools of all age ranges, rather than just secondary ones, to gain academy status.
[113] Hutchings and Francis (2017) look at the performance of academy chains – groups of schools that share a sponsor. They find mixed results on the extent to which sponsored academies within chains raise test scores.
[114] Eyles, Machin, and Silva (2015) provide a detailed overview of the differences between sponsored academies and converter academies and a discussion of how the process of conversion differs for each.

teachers who have more experience. On the other hand, sponsored academies, irrespective of when they opened, had a higher percentage of FSM pupils than the average state school and are characterized by lower test scores and younger less experienced head teachers.

Academies have greater freedom than local-authority controlled schools over how their staff are paid, and there is evidence that they use this greater freedom. In a survey carried out by the Department for Education, 36% of surveyed academies state that they had either made, or planned to make, changes to pay structure (Cirin, 2014).[115]

To give an initial idea of a possible contribution of academies to the changing salary pattern observed in Figure 5.1, Figure 5.2 plots mean head teacher salaries for state schools and academies opening before and after the Academies Act 2010. The Figure shows that academies opening prior to 2009/10 (Pre-2010 Sponsored) pay their head teachers much more than state schools. The salary differential is apparent in both: the early years when schools first became academies and afterwards once they had been in operation for at least four years.

Post-2010 academies also pay their head teachers more than the average state school, but the size of the premium is more modest. Figure 2 also shows that, amongst the post-2010 openings, salary premia are greater in converter academies compared to post-2010 sponsored academies. While post-2010 sponsored academies initially paid their head teachers a modest amount more than the average state school, this small premium had completely disappeared by 2013/14. Taken together, these raw salary numbers show that: early sponsored academies pay a large premium; converter academies a more modest premium; and later sponsored academy openings pay little, if any salary premium, relative to the average state school.

5.2.2 Head teacher salaries in England

[115] Unfortunately, the survey only documents changes in overall staff pay structure, and does not mention whether the schools altered their head teacher pay policy.

School governing bodies - typically comprised of parents, staff members, and, in some cases, local authority representatives - are responsible for holding head teachers accountable and ensuring that a performance management system is in place for head teachers (Earley et al., 2016). A sub-group of the governing body, along with an external advisor, monitors progress according to previously agreed-upon objectives, sets new objectives, and makes decisions about remuneration.

During the period studied in this paper nationally set pay scales determined both teacher and head teacher salaries, but as time progressed much more flexibility came into the system.[116] The national scales feature salary increases took the form of movements along spine points determined within schools, with the government set the salary scale range based upon recommendations by the School Teachers' Review Body (STRB). Established in 1991 as an independent body, every year the STRB offered recommendations to the Secretary State for Education and the Prime Minister on teachers' pay and conditions for England and Wales, after consultations with teacher unions and associations, organisations representing school governing' bodies and local authorities. Recommendations on changes in pay scales were meant to reflect the state of school teacher and head teacher supply and demand, and also the wider state of the labour market in England and Wales.[117] Each year, the government used these recommendations to update the School Teachers' Pay and Conditions Document (STPCD). This is the document that school governing bodies followed when setting salaries under this regime.

[116] Since September 2014, teacher pay in England has undergone drastic reform (see National Foundation for Educational Research et al. 2017 for details). The main change is that payment on suggested pay spine points is no longer mandatory. Under the current regime, state schools and academies are on equal footing with respect to head teacher pay. While pay spines are still published as a guide to remuneration, schools have the freedom to depart from the pay spines as they see fit (see Burgess, Greaves, and Murphy, 2019, for an analysis of the impacts of this reform).

[117] For instance, forecasted changes in the pupil population and the government's policy for public sector pay awards play a role in determining teacher pay policy. (STRB, 2018)

Depending on school year, the annual releases of the STPCD set minimum and maximum salaries for head teachers in six or eight different school groups, based on the STRB recommendations. The number of pupils enrolled in the school, their age composition, and special needs status determined the group. Within these groups, there are separate salary bounds determined by the school's location. Spine points then fall within these group specific bounds. Under this system, schools could pay more than the maximum for the relevant group, after seeking independent advice, if there were exceptional circumstances 'specific to the role of candidate'. The switch from six to eight salary bands occurred in the salary reforms at the start of the millennium in the 1999/00 school year, with additional reforms on spine points taking place in the 2000/01 and 2002/04 school years.

Figure 5.3 shows the percentage of schools, broken down by the typology introduced earlier, that pay their head teacher more than the maximum value for their group in even numbered years between 2001/02 and 2013/14.[118] Few schools – around 10% - in 2001/02, the year prior to the introduction of academies, paid head teachers above the group ranges relevant to their school. However, in the final year of the sample studied here, this fraction increased fourfold, up to 38%.[119]

Consistent with the descriptive evidence discussed earlier, academies are more likely to pay head teachers outside the ranges, though this varies by academy type. Around 70% of pre-2010 sponsored academies pay their head teachers a salary in excess of the maximum spine point by 2013/14. In the same school year, the percentage of converter and post-2010 sponsored openings that pay head teachers higher than the maximum spine point for their group does exceed the percentage for state schools, but the gap is far smaller. Indeed, state schools are not

[118] An alternative measure of how much schools depart from the STPCD is given by the percentage of schools not paying on the exact spine points recommended. We redo Figure 3 based upon this measure in Appendix Figure A1. As the Figure is almost unchanged, it suggests that all departures from the exact spine points are due to school paying above the maximum spine point for the group.
[119] The group ranges are set out in the yearly releases of the STPCD.

much less likely to pay outside of the spine points than the later academy conversions that make up most of the academy sector – 34% of state schools pay outside the spine points in 2013/14 as compared to 38% of converter academies.

As head teacher salaries rose in the relatively constrained non-academy sector prior to the large post-2010 expansion does raise questions on whether academisation was the prime mover behind rising head teacher salaries and increased inequality. Using the 2013 Global Teacher Status index – an index constructed to measure the level of respect for teachers and their social standing - Dolton (2013) notes the status of head teachers in the UK is amongst the highest in the world. In addition, he finds that school principal status is higher in the UK than in any of the other 21 countries surveyed and suggests that this may be due to the idea that head teachers are deemed to be pedagogical and managerial leaders rather than administrators. In the UK context, numerous Education Ministers have stressed the importance of effective head teachers, particularly at disadvantaged schools. As early as 1998, Tony Blair emphasised the role played by super heads in maintaining standards stating: "if a head teacher rises to the challenge of turning around a failing school, why should they not earn £60,000 or £70,000 a year?".[120]

This begs the question of whether academy/state school head teacher salary premia are driven by the selection of super heads into academies, rather than academies paying more per se. Figure 5.4 suggests that this may be happening. An individual is classified as a super head in year *t* if, in any year prior to *t,* that individual was at or above the 90th percentile of the salary distribution.[121] When looking at the salary distribution of previous years, we only focus on the distribution in the non-academy sector. The reason for doing so is that if academies confer a pay premium on head teachers, focusing on the total salary distribution will tautologically

[120] See http://news.bbc.co.uk/1/hi/education/210953.stm. A full transcript of the speech is available at http://www.britishpoliticalspeech.org/speech-archive.htm?speech=204.

[121] When calculating percentiles, we look at salary residuals. The residual salary nets out age and experience effects. Our results are very similar when raw salaries are used

assign super heads to academies. This is the reason for looking at how heads are paid in the non-academy sector as a proxy for super head status. Therefore, a super head, according to our definition, is a head teacher that has previously worked in the state sector for at least one academic year and, in at least one of those years, was at or above the 90th percentile of the earnings distribution for state school head teachers.

Figure 5.4 shows that both pre-2010 sponsored academies and converter academies employ a higher fraction of super heads compared to state schools. Interestingly, the proportion of super heads in academies is largest in the early years of the academies programme when more disadvantaged schools – sponsored academies – opened. This is consistent with the notion that super heads are paid a premium to turn around disadvantaged schools. In more recent years, converter academies are the most likely to employ a super head to run the school.

Thus, there is a need to exercise caution in interpreting academy/state school salary differentials. It appears that head teachers, who have previously been high performing in the state sector, are overrepresented in academy schools. Disentangling the rise of the super head from any causal effect of academisation on head teacher salaries, therefore becomes the challenge for the rest of the paper.

5.3 Empirical Strategy

5.3.1 Regression Estimates

The empirical plan is to study whether salary levels and their variance have been influenced by the growth of academies. The starting point is to look at academy/state school differentials through variants of the following equation:

$$w_{ist} = \sum_{j} \beta_1^j A_{st}^j + \delta_1 X_{it} + \gamma_1 S_{st} + a_t + e_{1ist} \tag{1}$$

where w_{ist} is the log salary of head teacher i in school s at time t; A^j_{st} takes value 1 if school s is an academy of type $j \in$ (*pre-2010 sponsored, post-2010 sponsored, converter*) at time t; X_{it} and S_{st} are vectors of head teacher and school characteristics, respectively; a_t denotes a set of year dummies; and e_{1ist} is an error term.

An ordinary least squares estimate of β^j_1 measures the average salary gap for an academy of type j head teacher versus a non-academy head teacher conditional upon head teacher and school characteristics. It is a useful descriptive point of departure, but for the estimand of interest, β^j_1, to be interpreted as causal, academy status needs to be uncorrelated with unobservable determinants of salaries that enter into e_{1ist}. Although a reasonably large set of covariates are available for inclusion in X_{it} and S_{st}, this condition seems unlikely to hold.

To move beyond this, one possibility is to exploit the matched firm-worker (i.e. school-head teacher) nature of the data. This enables controlling for time invariant individual and school level unobservables that may correlate with both A^j_{st} and w_{ist}. Specifically, the error term can be generalised to contain individual head teacher and school fixed effects, respectively a_i and a_s, to estimate the following more general equation:

$$w_{ist} = \sum_j \beta^j_2 A^j_{st} + \delta_2 X_{it} + \gamma_2 S_{st} + a_t + a_s + a_i + e_{2ist} \tag{2}$$

Unlike in the sizable literature that estimates wage equations in a variety of settings with worker and firm fixed effects[122], identification of the key parameters of interest does not rely on head teachers switching schools, but rather it comes from academy status changing within school/head teacher pairs.[123] Thus, the selection issue raised above, where super heads working in academies have stellar prior histories in state schools, can be dealt with.

[122] For instance, the seminal paper by Abowd, Kramarz, and Margolis (1999). For a comprehensive survey of such research see Card et al (2018).
[123] Of course, head teacher and school effects can only be identified separately if head teachers switch schools.

Equation (2) is estimated both with and without individual head teacher fixed effects. In the case of the latter, the model collapses to a regression model with two-way fixed effects (TWFE). A number of recent papers have shown that estimation of such models can be problematic when, as they are here, treatment effects are dynamic and differ across groups (Borusyak, Jaravel and Spiess, 2021; Callaway and Sant'Anna, 2021; De Chaisemartin and D'Haultfoeuille, 2020; Sun and Abraham, 2021; and Wooldridge, 2021 all provide detailed discussions regarding the interpretation of TWFE estimates).[124] To assuage concerns that the parameter of interest could be biased by negative weighting, estimates using the Callaway and Sant'Anna (2021) approach are shown alongside standard TWFE estimates in Appendix Tables A5.2 and A5.3. The estimates are similar to the conventional TWFE estimates.

One possible concern with (2), is that the academy effects aggregate over several distinct labour market transitions. For instance, A_{st}^{j} does not distinguish between those who have recently been hired into the academy and those employed by the school in previous academic years. This could be important if, for example, schools are more likely to hire new head teachers upon gaining academy status. Therefore, to allow pay premiums to differ by the type of transition in year *t,* (2) can be further generalised:

$$w_{ist} = \sum_{\substack{k=same,\\ switch}} \sum_{j} \beta_3^{jk}\, T_{it}^{j,k} + \beta_3^{state,switch} T_{it}^{state,switch} + \delta_3 X_{it} + \gamma_3 S_{st} + a_t + a_s + a_i + e_{3ist} \qquad (3)$$

The notation has now become a little complex, but $T_{it}^{j,same}$ takes the value of one if individual *i* remains in the same school, of type *j*, between years *t-1* and *t,* whereas $T_{it}^{j,switch}$

[124] Groups in our case refer to either conversion cohorts, defined by the year of academy conversion, or interactions between academy type and conversion cohorts. TWFE estimates tend to be interpreted as recovering weighted averages of time/group specific treatment effects. While this is the case, the recent literature has highlighted that the weights can be negative resulting in misleading inferences where, in the extreme case, TWFE can deliver an average effect that is negative even when all time/group average effects are positive.

takes value 1 if individual i is observed in a school of type j in t, but was not observed in that same school in the previous year. To set a baseline to be those who remain in the same state school between the years *t-1* and *t,* the specification includes $T_{it}^{state,switch}$– a dummy capturing the salary return to moving to a different state school in year t, rather than staying in the same state school across years *t-1* and *t.*

Finally, use of twenty years data means that time variations in estimated salary gaps can be considered. Thus, event study estimates are also presented. These are similar to the individual level estimates with school fixed effects (as each school only has a single head teacher) but have the advantage that tests for the existence of pre-trends in head teacher salaries at schools that gain academy status can be undertaken.[125]

5.3.2 Variance Decomposition

The previous section showed estimates of average head teacher salary differentials in models that do and do not deal with issues of selection. However, the salary distribution for head teachers has widened very dramatically through time. Assessing the extent to which this inequality increase is attributable to academisation offers a different challenge as it requires construction of a counterfactual salary distribution in the absence of academisation. One common method for computing such counterfactual distributions is the semi-parametric decomposition technique first developed in DiNardo, Fortin, and Lemieux (1996), hereafter referred to as DFL, and further refined over the years (see Fortin, Lemieux and Firpo, 2011).

The main interest is in computing a counterfactual distribution that would have prevailed in 2013/14 had academies not entered the English secondary sector. Borrowing notation from DFL, let $f^{2013/14}(w)$ denote the marginal density of log salaries in 2013/14:

[125] As with the TWFE estimates that pool together all post conversion years, we present event study estimates that use the Callaway and Sant'Anna (2021) methodology. These are in Appendix Figures A5.1 and A5.2.

$$f\left(w;\ t_x = 2013/14, t_{a|x} = 2013/14, t_w = 2013/14\right) \tag{4}$$

This is the distribution of log salaries when attributes *x*, the distribution of academies given attributes *a*|*x*, and the *w* schedule – the mapping between attributes, academy status, and prices - are all set to their 2013/14 level.[126] (4) can be rewritten as:

$$\iint f(w|\ x, a, t_w = 2013/14\,)\, dF(x|\ t_x = 2013/14\,) dF(a|x, t_{a|x} = 2013/14) \tag{5}$$

The following counterfactual can then be used to assess the role of academies in changing the salary distribution between the earliest pre-academy year (2001/02) and latest school years (2013/14)[127] in the sample:

$$f\left(w;\ t_x = 2013/14, t_{a|x} = 2001/02\,, t_w = 2013/14\right) \tag{6}$$

This is the head teacher salary distribution that would have occurred in 2013/14 had academies remained at their 2001/02 level and when attributes and the salary schedule remained unchanged at the 2013/14 level. Assuming the latter, so that the salary schedule and distribution of attributes themselves are unaffected by the rate of academisation, (6) can be written in a form similar to (5):

$$\iint f(w|\ x, a, t_w = 2013/14\,)\, dF(x|\ t_x = 2013/14\,) dF(a|x, t_{a|x} = 2001/02) \tag{7}$$

DFL show, by applying Bayes' rule, that we can go from (5) to (7) by reweighting 2013/14 observations by the following:

$$a * \frac{P\left[a = 1\middle|x, t_{a|x} = 2001/02\right]}{P\left[a = 1\middle|x, t_{a|x} = 2013/14\right]} + [1 - a] * \frac{P[a = 0|x, t_{a|x} = 2001/02]}{P[a = 0|x, t_{a|x} = 2013/14]} \tag{8}$$

---

[126] Attributes in this case are age, gender, tenure, attainment as measured by the percentage of pupils gaining 5 A*-C, the number of pupils at the school, and the percentage of pupils eligible for free school meals.

[127] Our choice of baseline year is irrelevant in this example - at least in terms of calculating counterfactuals - so long as the year predates the introduction of academies.

In our case, this simplifies further as we consider baseline years prior to the introduction of academies. In this instance, application of (8) means that we drop all academy observations in 2013/14 and weight the remaining attributes by the unconditional probability of not being in an academy in 2013/14, divided by the conditional probability of not being in an academy in 2013/14. The latter probability is estimated as a function of teacher and school attributes.[128] In essence, this amounts to adding more weight to non-academy head teachers who share similar characteristics to academy head teachers and less weight to those who do not.

Equation (8) is first computed for pre-2010 academies, then for post-2010 sponsored, and finally for converter academies, before considering the total effect of academies of the distributional statistics of interest. The counterfactual densities are used to construct counterfactual statistics, such as the variance that would have occurred in 2013/14 had the number of academies remained at zero. In addition to the main decomposition results, we also follow Firpo, Fortin, and Lemieux (2018) in performing a detailed Oaxaca-Blinder decomposition for the distributional changes considered.[129]

5.4 Data

Head teachers are studied in secondary schools in England between school years 1995/96 and 2013/14.[130] The main data sources used are the School Workforce Census (SWC) and the Database of Teacher Records (DTR). The SWC is an annual census covering the workforce of all maintained schools in England. Data are collected at the individual level and include characteristics such as place of employment, age, salary, tenure, and gender. Importantly for our purposes, each individual's role within the school is recorded, enabling us

---

[128] This weighting scheme is the same as that used by DiNardo and Lemiuex (1997), who consider how the absence of unions would affect the wage distributions of Canada and the US.

[129] Doing a detailed decomposition – that is sequentially isolating the role that individual covariates play in shaping distributional statistics over time – is not possible with the DFL reweighting. Firpo, Fortin, and Lemieux show how one can use recentred influence functions to extend the detailed decomposition technique of Oaxaca-Blinder, used to decompose difference in means, to general functionals of probability distributions. Our version of this is presented in Appendix Table A5.5.

[130] As well as excluding primary schools from our analysis, we also exclude non-mainstream schools, such as special schools and pupil referral units.

to identify head teachers. These data have been collected since the 2010/11 school year. Since the aim is to provide a picture of head teacher remuneration before and after the introduction of academy schools, we also match in the DTR to extend the data back to the 1995/96 school year. Like the SWC, the DTR collects individual-level information on teachers in maintained schools in England and allows us to identify both head teachers, their place of employment, and the aforementioned head teacher characteristics. The SWC and DTR data match well: the distribution of the key variables in both datasets is very similar, at both the individual and the school level.[131] The main difference between the two datasets is that the DTR only contains information for teachers who are in the Teacher's Pension Scheme. This restriction does not appear to matter in practice; for instance, in the raw data, we observe a drop in the number of school/head pairs from 3117 to 3095 between 2009/10 and 2010/11. This is almost identical in magnitude to the drop we observe between 2008/09 and 2009/10 - 3137 to 3117.[132]

Academy schools, and their precise year of conversion, are identified from an online register of all schools and colleges in England – Edubase, which contains information on sponsored and converter academies alongside their date of opening. Using data on school links, also available through Edubase, academy schools are linked to their predecessor schools. In a similar fashion, schools that merge can be linked together, and those who change their school identifier be tracked. When defining school fixed effects, groups of schools that merge in the sample period are aggregated as the same school; likewise, academies and their predecessor schools are treated as the same school.

In addition to teacher characteristics and time-invariant school effects, time-varying school characteristics can influence head teacher salaries. Data on time-varying school

[131] We further check the quality of the match by comparing the DTR and the SWC data in overlapping 2010/11 academic year, since the data for this year is provided in both datasets. We see that, again, the match is very good, although there is a small drop in the number of schools reporting in the first year of the SWC data.
[132] Yearly variation on observations comes from secondary schools closing in the timeframe that we study, mergers of already existing schools, and missing observations. Appendix Table A5.4 shows that results are unchanged if we focus on a balanced sample of schools.

characteristics was obtained via a number of data sources: the Local Education Authority School Information Service (LEASIS) for school level data on pupil composition (% eligible for FSM and the number of full time equivalent pupils) and staff characteristics (number of full time equivalent teachers) for school years prior to 2009/10; for subsequent years, the same information was obtained from the school level component of the National Pupil Database (NPD) and the publicly available, school level, version of the SWC; attainment variables, such as the percentage of pupils achieving 5 A*-C in school leaving exams, were obtained from publicly available school performance tables.[133]

5.5 Empirical Results

5.5.1 Academisation and Head Teacher Salaries

Table 5.2 shows the first set of empirical findings from the basic least squares regressions that estimate conditional academy/state school salary premia for head teachers. The initial specification, shown in column (1), pools all academies together and conditions only on year effects and head teacher characteristics.[134] Columns (2) and (3) break the academy variable down, first to reveal differences between pre and post-2010 academies, and then to the full three way breakdown of pre-2010 sponsored, post-2010 sponsored, and converter academies. This more general breakdown is maintained in columns (4)-(7) which additionally allow for various combinations of time invariant school and individual unobservables.

In line with the descriptive evidence discussed earlier, column (1) shows that head teachers at academies are paid just over 9% more than heads at other state-maintained schools.[135] But this average differential is not spread evenly across types of academy school.

[133] A school leaving exam grade in England is considered a 'pass' if it falls within the grade range A*-G, while a 'good' pass is in the grade range A*-C. The proportion of pupils getting five or more of 'good' grades is used by the government as the headline measure of school performance and published annually in school performance tables.
[134] Head teacher characteristics are age, tenure, gender, and the squares of age and tenure.
[135] (9.2% = [exp(0.088) – 1] X 100)

As column (2) shows, head teachers at pre-2010 academies enjoy a premium about four times as high as those at schools that became sponsored academies after 2009/10. The spread across academy school type is larger when additionally splitting out the post-2010 group into sponsored and converter academies, as shown in column (3). The range of salary premia for the three run from 22% for head teachers in pre-2010 sponsored academies, to 7% for post-2010 sponsored academies down to 6% in converters. Thus, there is a salary differential hierarchy that accords to the super head hypothesis raised earlier.

Column (4) and (5) add time-varying school characteristics and prior attainment. This mostly attenuates the estimates a little: the differentials move down to a very similar ranking hierarchy of 20%, 9% and 3% across the academy types. Prior differences in test score performance across schools do not appear to be driving the overall results about academies (see Besley, Machin and Telhaj 2018, for evidence on increased salary-performance sensitivities over time using the same data as in this paper). Figure 5.5 shows event study estimates with and without lagged controls for pupil performance to echo this finding.

One important additional takeaway from Figure 5.5 is that, at least for post-2010 sponsored and converter academies, there is no pre-trend in head teacher salary gaps before conversion. For the schools where the premium is the largest – the pre-2010 sponsored academies – there is a very modest pre-trend, showing a small increase prior to conversion.

The selection question is addressed in the fuller models reported in the final two columns of the Table. Column (6) shows results from including school fixed effects, whilst column (7) additionally includes head teacher fixed effects to identify salary premia from switchers. Once time-invariant school and individual effects are accounted for, the academy salary premium falls across all three types of academy. It maintains the same hierarchy, but drops to 8%, 3% and 1% respectively in pre-2010 sponsored academies, post-2010 sponsored and converter academies. Thus, the estimated differentials are still present, but become quite a lot smaller in magnitude.

It is worth noting that the column (7) results that include both school and individual fixed effects do not rely solely on individuals who are in the same school before and after it gains academy status. It is possible to allow for individual and school level unobserved heterogeneity and to identify the effect of switching into an academy school. This is because the specifications reported in Table 5.2 pool together distinct labour market transitions, all of which may bring different salary gains or losses.

As noted above, there is a slight upward trend in salaries for the pre-2010 academies. This is investigated further through implementation of a different, more stringent research design for these academies. The main results of Table 5.2 are replicated for the pre-2010 conversions now using schools that obtain sponsored academy status in 2009/10 and 2010/11 as a control group (this is the research design studying pupil performance effects from sponsored academy conversion in Eyles and Machin, 2019). There are no pre-trends in this case. As the schools themselves become treated post-2010, the sample is restricted to the years 1995/96-2008/09. Very reassuringly, the results – shown in Table 5.3 – prove very similar to those columns related to sponsored academies in Table 5.2. Importantly, this use of future sponsored academy converters as a control group fully alleviates the minor concerns about the issue of pre-trends in earnings. [136]

Table 5.4 shows results from more flexible salary equations that allow for separate effects for individuals joining an academy in year *t* and those remaining in one across consecutive years. The transition rates are given in the final column of the Table and clearly differ across academy type. The reported estimates of salary differentials first come from

[136] To give more detail, whereas a joint test of significance for the pre coefficients in Figure 5.5 gives p-values of 0.04 and 0.07 (depending upon whether attainment is controlled for), the same test when only future conversions are used as control schools gives p-values of 0.97 and 0.89 respectively.

models that do not initially contain school and individual fixed effects (column (1)), before they are added in columns (2) and (3). The picture that emerges from the Table is consistent with the findings in Table 5.2 where pre-2010 sponsored academies tend to pay more than academies opening after 2009/10. Across the board, the salary premiums are similar irrespective of whether the head teacher has just joined the academy or was present in the previous year. A minor exception can be seen in Column (3) where post-2010 sponsored academies pay a slightly higher premium for new head teachers than those continuing from a previous year.

The near uniformity of salary premiums in the academy sector between switchers and stayers suggests that premiums in academy schools are not accounted for by greater turnover within the academy school sector. The most consistent finding that emerges from Tables 5.2 and 5.3 is that academies do have salary premiums relative to state schools, but these are - at least for post-2010 sponsored and converter academies - very modest once one allows for unobserved school and individual heterogeneity. Similarly, raw salary differences overstate the premium attached to leading a pre-2010 sponsored academy, but in this case only the premium remains reasonably sizeable at about 8% once the full set of controls are included.

Interestingly, sponsored academies – specifically those that opened prior to May 2010 - have been found to benefit from gaining academy status. Eyles and Machin (2019) provide evidence that these schools significantly improved pupil performance at Key Stage 4.[137] They also show that one of the main changes that these schools made upon gaining greater operational autonomy was to change their leadership. Taken alongside the fact that these schools were disadvantaged prior to conversion, the estimated salary premiums found for these academies adds to a narrative stating that super heads were parachuted in to improve performance in these schools.

[137] Key Stage 4 consists of secondary school students in grades 10-11.

5.5.2 Academisation and Head Teacher Salary Inequality

Moving to the inequality analysis, Table 5.5 reports the variance decomposition results. As well at looking at how the salary variance changes between 2001/02 and 2013/14, three other distributional statistics (the 90-10, 90-50 and 50-10 salary differentials) are considered. Academies are separated into the three types to consider their effect on the salary distribution separately.

The first exercise reported in Table 5.5 is purely descriptive. To give a rough idea of the contribution of each academy type to rising dispersion they are simply dropped from the sample in 2013/14 and the distributional statistics recomputed to see how the 2001/02 - 2013/14 change compares with the change calculated using the full sample. Column (1) shows that each measure of head teacher salary dispersion has increased over time, the variance rising by 0.022 from an initial level of 0.020, the 90-10 going up by 0.098 from an initial level of 0.34, and this reflecting an increase in both the upper and lower parts of the distribution (the 90-50 going up by 0.040, and the 50-10 by 0.058)

Columns (2), (3) and (4) respectively look at the effects of the three academy types, and column (5) at all academies jointly. It should be noted that these 'decompositions' are not sequential so that when, for instance, post-2010 academies are dropped and the statistics recomputed, pre-2010 and converter academies remain in the sample. The entries in the columns represent the difference between the actual change in the statistics and the same change without the relevant academy observations in the 2013/14 sample. Overall, there is not much evidence that academies have caused rising salary inequalities for head teachers. The column (5) specification shows around a 18% contribution, a small effect suggesting that most of the inequality rise would have happened anyway. Academies appear to have opposite sized effects on the 50-10 and 90-50 salary differentials but taken as a whole they make at best a modest contribution to the overall increase in salary inequality.

Columns (2)-(5) simply drop academy observations from the sample when recalculating statistics. The problem with this is that, in so far as academy status and other relevant attributes are correlated, dropping these observations also changes the joint distribution of relevant attributes in 2013/14. To hold these attributes fixed at their 2013/14 level, DFL weights were derived and applied. Columns (6)-(9) shows the results with better counterfactuals based upon academies remaining at their 2001/02 level, but all other attributes remaining at their 2013/14 level. Under the, admittedly strong, assumption that the salary schedule and distribution of attributes would have remained unchanged had the observed rise of academies not occurred, these estimates can be interpreted as causal.

Once again, academies taken have little influence over the increase in salary variance over time. The largest individual effect, observed in Column (9), is that of academies on the 50-10 differential where we see academies contributing to an increase in this statistic over time. While this effect is large, most of the rise in dispersion occurs towards the bottom of the salary distribution where academies, irrespective of type, have a lesser influence.

Taken together, the reweighted estimates suggest that around 5% of the increased variance observed between the years can be attributed to the entrance of academy schools into the English secondary sector. While this effect is non-negligible, it is hard to argue for anything other than at best a modest, second order role played by academy schools in driving increased head teacher salary inequalities over time. The findings are qualitatively similar to those in Appendix Table A5.4 where we perform a two-way Oaxaca-Blinder decomposition using recentred influence function regressions. Using this method, we find that academies account for around 8% of the total rise in the variance of salaries between 2001/02 and 2013/14. Again, this is modest at best.

The final question that arises is to how to square this with earlier evidence showing that a subset of academies – pre-2010 sponsored – offer significant salary premiums. But this subset is very small as a share of the academies operating in the English secondary school sector. The simple explanation is that most academies – those opening after the implementation of the Academies Act 2010 – are precisely the ones with the least impact on head teacher salaries.

5.5.3 Extensions

Several robustness exercises were undertaken to determine whether the estimates reported to date can plausibly be interpreted as causal. Columns (2) - (3) of Table 5.6 reports on two robustness exercises, presenting them alongside the most conservative baseline estimate from before in column (1) (the model that includes school and head teacher fixed effects in column (7) of Table (2)).[138] Column (2) shows results from a 'fake policy' experiment. To do this, the sample is ended in 2009/10, and the opening date of all the academies shifted back five years to define treatment based upon a 'fake', five year lagged, conversion year, and remove all individuals from the sample who are treated. If salary differences between head teachers are due, solely, to them being at academy schools, the 'fake' treatment estimates should be statistically indistinguishable from zero. As can be seen in Table 5.6, this is the case for post-2010 sponsored academies, but not for converters or pre-2010 sponsored academies. Despite this, the significant pre-conversion coefficient for these academies is either economically insignificant – around 1% for converter academies – or, in the case of pre-2010 sponsored academies, very small at 1.5% when benchmarked against the large post conversion coefficient.

As there appears to be a slight salary premium for head teachers in some academies even before conversion, linear school specific trends were also incorporated. The addition of trends reduces the estimates across the board, but the qualitative conclusions are largely unchanged: converter academies pay a negligible premium for head teachers while sponsored academies do pay more after conversion. The principal difference is that the pre-2010 premium

[138] To make column (1) comparable to the fake policy experiment the first 5 years of data were excluded.

is now attenuated towards the post-2010 estimate and lies around 4% rather than the 7% estimated previously. Thus, the salary premium hierarchy remains, but its spread is reduced and the overall conclusion that academy head salary differentials are modest remains.

5.5.4 Sources of Pay Discrepancy.

This section considers how the estimates relate to the institutional salary setting changes introduced earlier, where schools are increasingly setting salaries outside of regulatory salary ranges. For example, one source of the academy salary premium could be that academy heads find themselves in different pay groupings to heads of non-academies.

Table 5.6 presents results from a more flexible version which allows the Table 5.2 estimates to differ according to the pay group, *G*, of the school as follows:

$$w_{ist} = \sum_{k=1,2,3} \sum_{j} \beta_4^{jk} G_{it}^{jk} + \delta_4 X_{it} + \gamma_4 S_{st} + a_t + a_s + a_i + e_{4ist} \tag{9}$$

In (9), types are indexed by *j* so that G_{it}^{jk} takes value 1 if individual *i* is in an academy of type *j*, in pay grouping *k*, at time *t*. Since some pay groupings contain few schools, the eight available groupings are aggregated into three subsets, indexed by *k*.[139]

Table 5.7 shows that salary premiums still arise within these detailed groups, although not all are statistically significant. Across the board, the premiums are highest at pre-2010 academies, with post-2010 sponsored next, and then converters offering minimal salary premiums. Pay premiums do not appear to be systemically larger within any one grouping; for instance, pre-2010 and post-2010 academy premiums are highest within the lowest and middle groupings – covering relatively smaller schools – while pay premiums for converter academies only arise in the final grouping – covering the largest secondary schools. The absence of large

[139] Because there are different minimum and maximum salary values within these groups according to school location, we additionally control for school location in the regressions.

pay premiums occurring within any one grouping, and the fact that pre-2010 sponsored academies still offer the highest premiums, suggests that academy pay differentials are not driven by academies systematically being in different groupings than non-academies due to location or the age distribution and size of their intake. Rather, this is suggestive of academies using their greater freedom to deviate from the STPCD document that determines head teacher pay for state schools.

Although state schools are increasingly setting salaries outside of the pay scales set out in the STPCD, the descriptive evidence presented earlier suggests that academies are more likely to do so. Table 5.8 examines this more formally reporting results that come from probit models with a binary dependent variable for whether the head teacher's salary exceeds the maximum for the school group. The estimates show that academy salaries are much more likely to exceed the maximum threshold value than it is in state schools. In line with the previous results, effects are more pronounced for sponsored academies – specifically pre-2010 sponsored. These schools are over 40% more likely to pay outside of pay spines than comparable state schools. While post-2010 sponsored academies show a similar pattern, albeit with around half the effect size, there is a large difference between sponsored and converter academies as converter academies are only 4% more likely to exceed the maximum pay for their group.

Section 5.2.2 of this chapter discussed pay setting in English schools and argued that there has been an increased emphasis over the previous decades on the importance of effective head teachers, or super heads, in education production. Descriptive evidence has already shown pre-2010 sponsored and converter academies to be more likely to hire super heads than state schools. Table 9 shows estimates of the probability of hiring a super head, which also control for school characteristics while doing so. As defined in section 5.2.2, a super head is someone who, in any previous academic year, was at or above the 90th percentile of the residual earnings distribution. Residual salary is the salary once age, experience, and gender effects are netted out. Only the salary distribution for non-academies is considered as academy pay mixes the premium to being at an academy with the premium to being a super head.

Unlike the results of Table 5.8, where the addition of school level controls does little to change the results, school controls have a large effect upon the descriptive results for super heads. Column (1) of Table 5.9 shows probit estimates that are in line with the descriptive evidence given in Figure 5.4. However, when school level demographic and attainment controls are added, the results reverse. In the latter two cases – shown in columns (2)-(3) of Table 5.9 - post-2010 sponsored academies and converters are more likely to hire super heads than similar state schools, but pre-2010 sponsored academies are no more likely to do so.[140]

Can this be reconciled with the earlier findings? First of all, it is consistent with the small salary premiums found for post-2010 academies once individual fixed effects are included. These academies do not offer a large premium per se, but they are able to attract head teachers who were already well compensated in the state sector. Secondly, and converse to this, pre-2010 sponsored academies do seem to offer a genuine 'academy premium' – they are no more likely to attract super heads than schools like themselves, but they do offer a significant salary boost for those head teachers that they attract.

This is not to say that pre-2010 academies are less likely to attract super heads than other schools. Looking across the three columns of Table 5.9, one can see that the estimated likelihood of these schools attracting high paying head teachers only falls once school characteristics are conditioned upon. In sum, while pre-2010 sponsored academies are more likely to attract super heads relative to the average state school, they are no more likely to

[140] The results in Table 5.9 are robust to alternative definitions of 'super head'. For instance, if we restrict super heads to be those who, in the previous five years, were in the top 10% of the residual income distribution, the estimates in column (3) become 0.006 (0.022), 0.060 (0.027), and 0.035 (0.011). Restricting to the previous 5 years alone removes the mechanical under assignment of super head status to teachers joining sponsored academies that opened early in the sample.

attract a highly paid head teacher than schools similar to themselves – namely those that are disadvantaged and, typically, characterised by poor performance.[141]

5.6 Conclusion

This paper studies the introduction of academy schools in England to assess the contribution of school autonomy – in particular, discretion in setting remuneration levels – on leadership salary differentials and inequality. It uncovers evidence of head teacher salary differentials between academy schools which operate with more autonomy and traditional state schools, and of 'super heads' receiving higher salaries to run academies. However, a large part of these salary differentials can be accounted for by observable teacher and school characteristics.

Moreover, the majority of the observed increase in head teacher salary inequality that occurred in English secondary schools remains unexplained by the growth of autonomous schools. The emergence of "super heads", whose high levels of remuneration have contributed to the rise in head teacher salary inequality, is therefore not a direct consequence of the academisation of English schools and the extra freedoms they have. Instead, it reflects a more general shift towards the market determination of salaries in the English secondary school sector for school leadership personnel. The exception to this is the rise in remuneration for head teachers at pre-2010 sponsored academies. As highlighted elsewhere, these schools are both highly disadvantaged and, upon attaining academy status, significantly raised pupil performance. Although a direct causal link between increased pay and performance is not the focus of this paper, the evidence is suggestive that the ability to increase the pay of school leaders is a mechanism through which pre-2010 sponsored academies were able to raise

[141] This is not at all surprising. As shown in Appendix Figure A5.3, if disadvantage is measured by the proportion of FSM pupils in a school, super heads have become more concentrated in disadvantaged schools over time.

attainment levels of their disadvantaged student body. Even though post-2010 sponsored academies are like their pre-2010 counterparts, they opened during a period of declining school funding, and this may have inhibited their ability to use their greater autonomy to pay school leaders more. The results highlight that the ability to 'scale up' effective schooling interventions depends crucially upon the broader educational landscape.

Finally, the study context is England. However, the findings are of relevance to a range of other settings where autonomous schools have arrived on the education scene (for example, charter schools in the US and free schools in Sweden as discussed in Eyles, Hupkau and Machin, 2016a) and/or where collective bargaining for teachers has been dismantled or salary determination has become more flexible (for example, see the US reform discussed in Biasi, 2021; and Biasi and Sarsons, 2021). It will be an important aim of further research to gain a better understanding of whether different forms of salary determination can deliver better running education systems. Our analysis asks whether academy schools lead to greater pay inequality and rising wage levels, but we do not consider the overall effect that spending reforms have on the efficacy of the education system. Understanding the overall effect of pay reforms in a general equilibrium context can help clarify whether pay decentralisation merely changes the allocation of teachers and school leaders between schools or whether it increases the supply of effective school leaders.

It will also be of interest to look at these kinds of questions for public sector settings more generally, other than schools where reforms have been introduced.

References

Abowd, John, Francis Kramarz, and David Margolis. 1999. "High Wage Worker and High Wage Firms", *Econometrica*, 67: 251-337

Academies Commission. 2013. "Unleashing Greatness: Getting the Best from an Academised System." http://www.educationengland.org.uk/documents/pdfs/2013-academies-commission.pdf

Acemoglu, Daron and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings", in *Handbook of Labor Economics* Vol.4b (Elsevier Science North Holland), ed. Orley Ashenfelter and David Card.

Atkinson, Anthony. 2015. "*Inequality: What Can be Done?*", Harvard University Press.

Besley, Timothy, Stephen. Machin and Shqiponja Telhaj. 2018. "Are Public Sector CEOs Different? Pay for Performance and Salary Inequality in Schools", London School of Economics mimeo.

Biasi, Barbara. 2021. "The Labour Market for Teachers Under Different Pay Schemes", *American Economic Journal: Economic Policy*, 13: 63-102.

Biasi, Barbara. and Heather Sarsons. 2021. "Flexible Wages, Bargaining and the Gender Gap", *Quarterly Journal of Economics*, forthcoming.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen. 2015. "Does Management Matter in Schools?", *Economic Journal*, 125: 647-74

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. "Revisiting Event Study Designs: Robust and Efficient Estimation", working paper, University College London.

Burgess, Simon, Ellen Greaves, and Richard. Murphy. 2019. "Deregulating Teacher Labor Markets*", IZA Discussion Paper* No. 12592

Callaway, Brantly, and Pedro Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods", forthcoming, *Journal of Econometrics*

Card, David, Anna Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and Labor Market Inequality: Evidence and Some Theory", *Journal of Labor Economics*, 36, S13-S70.

Cirin, Rob (2014) "Do Academies Use Their Autonomy?" Department for Education Research Report. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/401455/RR36 6_-_research_report_academy_autonomy.pdf

de Chaisemartin, Clement, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects", *American Economic Review* 110: 2964–2996

Department for Education and Skills (2002) "Statistics of Education: Schools in England", http://webarchive.nationalarchives.gov.uk/20120506195920/http://www.education.gov.uk/rsgateway/DB/VOL/v000359/index.shtml

DiNardo, John, and Thomas Lemieux. 1997. "Diverging Male Wage Inequality in the United States and Canada, 1981-1988: Do Institutions Explain the Difference?", *Industrial and Labor Relations Review*, 50: 629-651.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach" *Econometrica*, 64: 1001-1044.

Dolton, Peter. 2013. "The Status of Teachers", CentrePiece, 18, Autumn: 13-15.

Earley, Peter, Sara Bubb, David Eddy-Spicer, Megan Crawford, and Chris James. 2016. "Governing Bodies, Headteacher Performance and Pay: the Role of External Advisers" *Educational Review,* 68: 257-273

Eyles, Andrew, Claudia Hupkau, and Stephen Machin. 2016a. "Academies, Charters and Free Schools: Do New School Types Deliver Better Outcomes?", *Economic Policy,* 31: 453-501.

Eyles, Andrew, Claudia Hupkau, and Stephen Machin. 2016b. "School Reforms and Pupil Performance", *Labour Economics*, 41: 9-19.

Eyles, Andrew and Stephen Machin. 2019. "The Introduction of Academy Schools to England's Education", *Journal of the European Economic Association*, 17: 1107-1146.

Eyles, Andrew, Stephen Machin, and Olmo Silva. 2018. "Academies 2 – The New Batch: The Changing Nature of Academy Schools in England", *Fiscal Studies,* 39: 121-158.

Firpo, Sergio, Thomas Lemieux, and Nicole M. Fortin. 2018. "Decomposing Wage Distribution Using Recentered Influence Functions", *Econometrics,* 6: 1-40.

Fortin, Nicole M, Thomas Lemieux, and Sergio Firpo. 2011. "Decomposition Methods in Economics", in *Handbook of Labor Economics* Vol.4a (Elsevier Science North Holland), ed. Orley Ashenfelter and David Card.

Gibbons, Robert and Kevin Murphy. 1992. "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence", *Journal of Political Economy*, 100(3): 468-505.

Hanushek, Eric A, Susanne Link, and Ludger Woessmann. 2013. "Does school autonomy make sense everywhere? Panel estimates from PISA", *Journal of Development Economics*, 104: 212–232.

Hutchings, Merryn and Becky Francis. 2017. "Chain Effects 2017: The Impact of Academy Chains on Low-Income Students" The Sutton Trust Research Report, https://www.suttontrust.com/wp-content/uploads/2017/06/Chain-Effects-2017.pdf

Jackson, Kirabo. 2012. "School Competition and Teacher Quality: Evidence from Charter School Entry in North Carolina", *Journal of Public Economics*, 96: 431-438.

Jakubson, George. 1991. "Estimation and Testing of the Union Wage Effect Using Panel Data", *Review of Economic Studies*, 58(5): 971-991.

Jensen, Michael and Kevin Murphy. 1990. "Performance Pay and Top-Management Incentives", *Journal of Political Economy*, 98(2): 225-264.

Murphy, Kevin. 1999). "Executive Compensation", in *Handbook of Labor Economics* Vol.3b (Elsevier Science North Holland), ed. O. Ashenfelter and D. Card.

National Foundation for Educational Research (2017) "Evaluation of Teachers' Pay Reform". https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/652763/Evaluation_of_Teachers__Pay_Reform-Final_Report.pdf

Piketty, Thomas. 2014. "*Capital in the Twenty-First Century*", Cambridge: Belknap Press.

School Teachers' Review Body. 2018. School Teachers' Review Body: Twenty-Eighth Report-2018, (July),https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/728381/CCS207_CCS0518679568-1_STRB_Book_Web_accessible.pdf

Sun, Liyang and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects", forthcoming, *Journal of Econometrics*.

The New Schools Network. 2015. "Comparison of Different Types of School: A Guide to Schools in England" http://www.newschoolsnetwork.org/sites/default/files/files/pdf/Differences%20across%20school%20types.pdf

Wooldridge, Jeffrey. 2021. "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators", Working paper, Michigan State University

Figure 5.1: Head Teacher Salary Growth Over Time

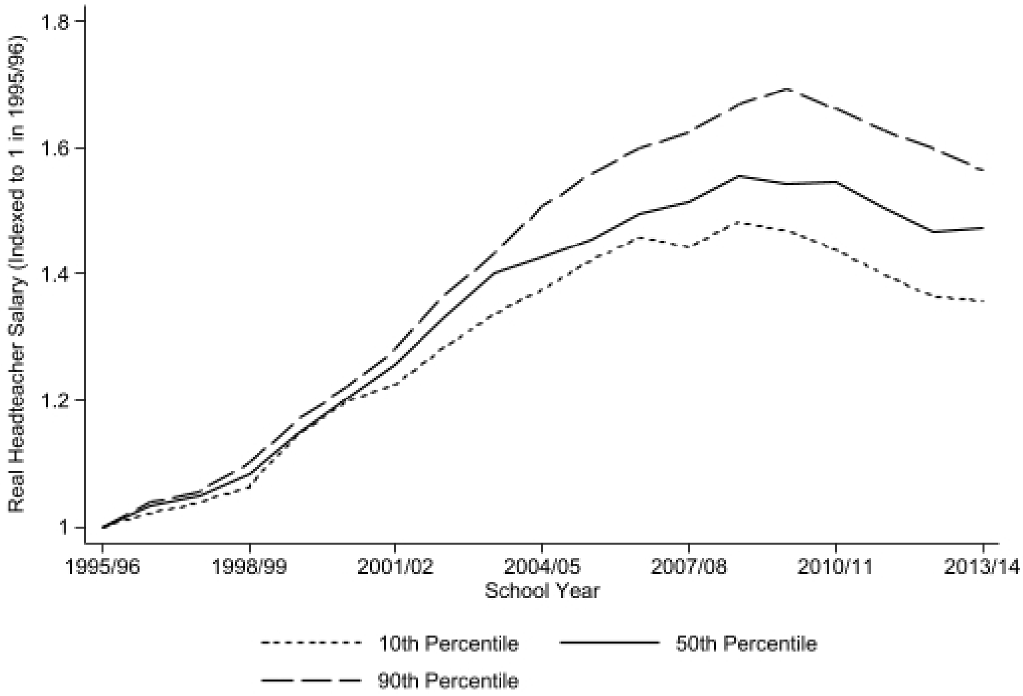


Figure 5.2: Academy Head Teacher Salaries Over Time

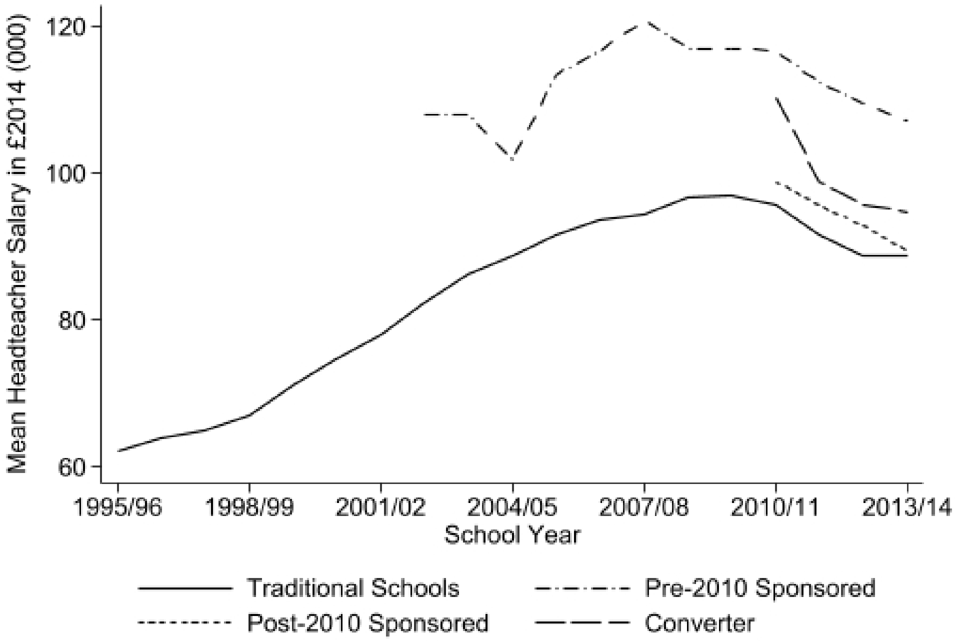
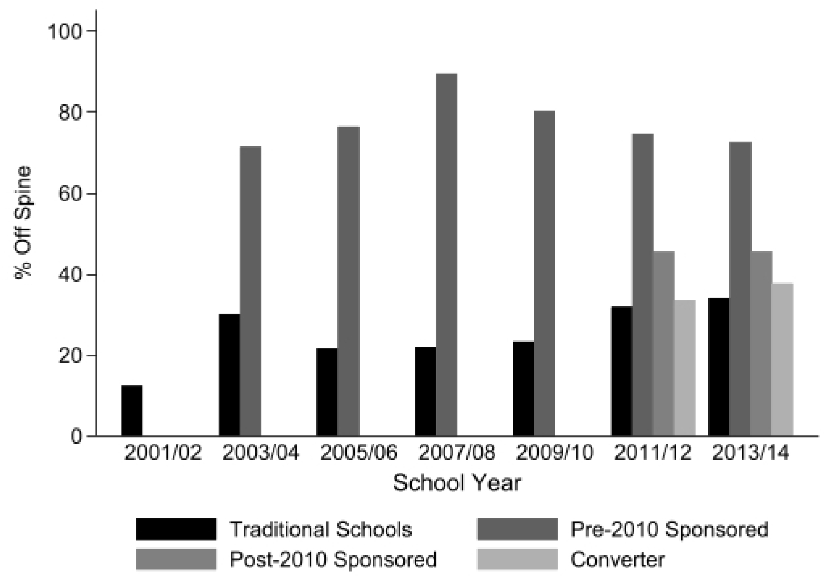
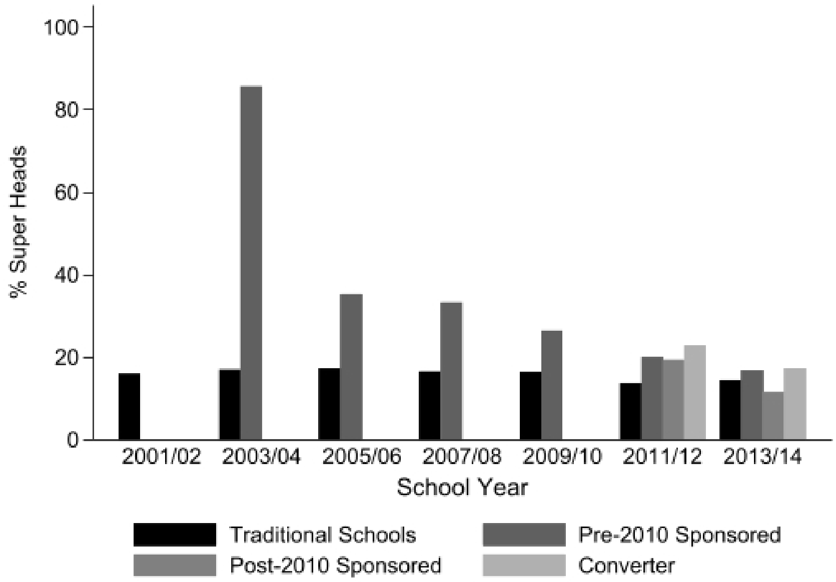


Figure 5.3: % Paying Off Spine



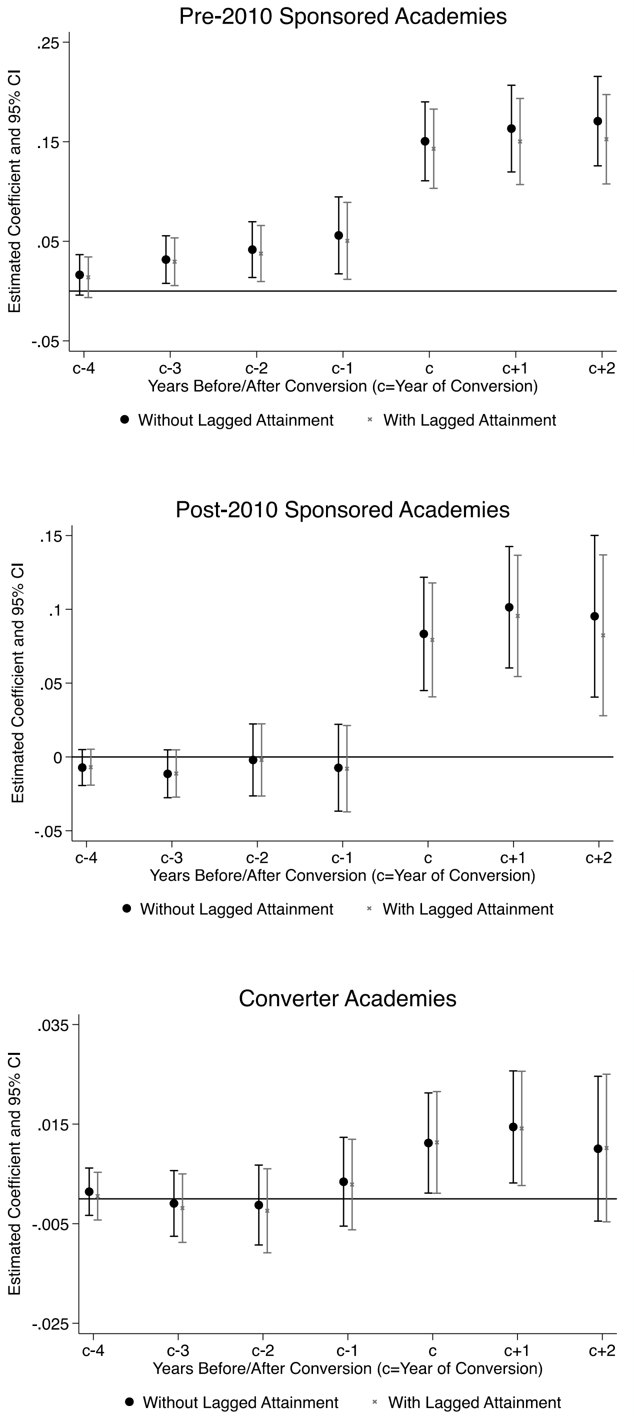
Notes: Off spine means being paid above the maximum salary, as set out in the STPCD, for a school of school of a given size, in a given locality, and with a given intake of SEN pupils. A detailed description of how we construct the maximum salary is given in the data appendix.

Figure 5.4: % Super Heads



Notes: We define a super head as someone who, in any previous academic year, was at or above the 90th percentile of the residual earnings distribution for traditional schools (as academy pay mixes the premium to being at an academy with the premium to being a super head). As some heads are observed only at academy schools, 6% of head teachers are lost when doing this calculation. A more in-depth discussion of the definition of super heads can be found in section 5.2 and in the Appendix.

Figure 5.5: Event Study Estimates



Notes The coefficients (and associated confidence intervals) are from a regression of log salary on head teacher characteristics, school characteristics, year effects, school effects, and dummies for years before/after academy. We omit years prior to 5 years before the data of conversion and year 3 or more years after. Standard errors are clustered at the school level. Estimates with attainment include interactions between observation year and the attainment measure.

Table 5.1: Head Teacher and School Characteristics By Type of School

	Type of School				Gap Compared to Traditional School		
	Traditional	Pre-2010	Post-2010	Converter	Difference (2)-(1)	Difference (3)-(1)	Difference (4)-(1)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Head Teacher Characteristics							
Log (Salary)	11.02	11.20	11.02	11.07	0.17 (0.01)	-0.01 (0.01)	0.04 (0.00)
Age	50.45	49.29	48.15	51.04	-1.17 (0.25)	-2.30 (0.37)	0.59 (0.12)
Male	0.78	0.75	0.75	0.79	-0.03 (0.02)	-0.03 (0.02)	0.01 (0.01)
Tenure	3.04	2.71	1.00	4.55	-0.69 (0.17)	-2.40 (0.22)	1.15 (0.10)
School Characteristics							
5 or more A*-C Key Stage 4	0.42	0.4	0.36	0.45	-0.02 (0.01)	-0.06 (0.01)	0.04 (0.00)
Free school meals	0.18	0.33	0.28	0.13	0.14 (0.00)	0.10 (0.01)	-0.05(0.00)
Log (Total teachers)	3.95	3.93	3.7	4.06	-0.02 (0.02)	-0.25 (0.02)	0.11 (0.01)
Log (Total pupils)	6.76	6.69	6.51	6.88	-0.07 (0.01)	-0.25 (0.02)	0.12 (0.01)

Notes: Robust standard errors in parentheses. Tenure refers to overall tenure as a head teacher, while 5 A*-C refers to the variable observed in the previous year. To avoid conflating genuine differences across schools with general time trends in variables, residuals, net of year effects, are used in Table 5.1. Table 5.1 therefore presents averages across the entire sample period, for each school type, once year effects are accounted for.

Table 5.2: Academies and Head Teacher Salaries

	Log(Head Teacher Salary), 1995/96 to 2013/14						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Academy Pooled	0.088 (0.007)						
Academy*Pre-2010 Opening		0.201 (0.017)					
Academy*Post-2010 Opening		0.051 (0.006)					
Academy*Pre-2010 Sponsored			0.201 (0.017)	0.184 (0.017)	0.172 (0.017)	0.132 (0.013)	0.073 (0.014)
Academy*Post-2010 Sponsored			0.064 (0.015)	0.088 (0.016)	0.084 (0.016)	0.075 (0.014)	0.028 (0.010)
Academy*Converter			0.049 (0.006)	0.030 (0.005)	0.031 (0.005)	0.020 (0.004)	0.013 (0.004)
Sample Size	54,248	54,248	54,248	54,248	54,248	54,248	54,248
Number of Schools	2917	2917	2917	2917	2917	2917	2917
Number of Head Teachers	7904	7904	7904	7904	7904	7904	7904
Head Teacher Characteristics	Y	Y	Y	Y	Y	Y	Y
Year Effects	Y	Y	Y	Y	Y	Y	Y
School Characteristics	N	N	N	Y	Y	Y	Y
School Attainment (t-1)	N	N	N	N	Y	Y	Y
School Fixed Effects	N	N	N	N	N	Y	Y
Head Teacher Fixed Effects	N	N	N	N	N	N	Y

Notes: Robust standard errors (in parentheses) are clustered at the individual level. Head teacher characteristics are age, tenure, gender, and the squares of age and tenure. School level characteristics include the proportion of free school meal eligible pupils, the log of full time equivalent pupils, and the log of full time equivalent teachers. School attainment includes the proportion of pupils achieving 5 or more A*-C at Key Stage 4 in the previous year.

Table 5.3: Pre-2010 Sponsored Academy Estimates, 1995/96 to 2008/09 (2009/10 and 2010/11 Conversions as Control Group)

	Log(Head Teacher Salary), 1995/96 to 2008/09				
	(1)	(2)	(3)	(4)	(5)
Academy*Pre-2010 Opening	0.198	0.182	0.156	0.126	0.062
	(0.037)	(0.038)	(0.038)	(0.026)	(0.016)
Sample Size	2,209	2,209	2,209	2,209	2,209
Number of Schools	158	158	158	158	158
Number of Head Teachers	489	489	489	489	489
Head Teacher Characteristics	Y	Y	Y	Y	Y
Year Effects	Y	Y	Y	Y	Y
School Characteristics	N	Y	Y	Y	Y
School Attainment (t-1)	N	N	N	Y	Y
School Fixed Effects	N	N	N	Y	Y
Head Teacher Fixed Effects	N	N	N	N	Y

Notes: As for Table 5.2. Estimates use 2009/10 and 2010/11 conversions (sponsored) as a control group.

Table 5.4: Head Teacher Transitions and Salaries

	Log(Head Teacher Salary), 1995/96 to 2013/14			% Transitions
	(1)	(2)	(3)	(4)
Switch (Traditional School)	0.020	0.022	0.001	6.6
	(0.002)	(0.002)	(0.001)	
Same (Pre-2010 Sponsored)	0.171	0.131	0.073	1.2
	(0.018)	(0.013)	(0.015)	
Switch (Pre-2010 Sponsored)	0.207	0.159	0.072	0.2
	(0.024)	(0.022)	(0.018)	
Same (Post 2010 Sponsored)	0.066	0.059	0.030	0.7
	(0.015)	(0.014)	(0.009)	
Switch (Post 2010 Sponsored)	0.176	0.166	0.048	0.1
	(0.025)	(0.023)	(0.014)	
Same (Converter Academy)	0.032	0.021	0.012	5.2
	(0.005)	(0.004)	(0.004)	
Switch (Converter Academy)	0.020	0.018	0.014	0.2
	(0.012)	(0.013)	(0.006)	
Sample Size	54,248	54,248	54,248	
Number of Schools	2917	2917	2917	
Number of Teachers	7904	7904	7904	
Head Teacher Characteristics	Y	Y	Y	
Year Effects	Y	Y	Y	
School Characteristics	Y	Y	Y	
School Attainment (t-1)	Y	Y	Y	
School Fixed Effects	N	Y	Y	
Head Teacher Fixed Effects	N	N	Y	

Notes: As for Table 5.2 Same refers to those remaining in the same school across years (even if the school becomes an academy between t and t-1) while switch refers to those taking up a new post in the current year. The omitted category is those that remain in the same traditional school between t and t-1. Column (4) refers to the % based on the overall sample.

Table 5.5: Salary Inequality and Variance Decompositions

	Unweighted					Weighted			
					Type of Academy				
	Raw	Pre-2010	Post-2010	Converter	All	Pre-2010	Post-2010	Converter	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Variance	0.022	0.001	0	-0.001	0.004	0.001	0	-0.002	0.001
		(5%)	(0%)	(-5%)	(18%)	(-5%)	(0%)	(-9%)	(5%)
90-10	0.098	0.011	-0.005	-0.014	0.015	0.008	-0.026	-0.006	0.014
		(11%)	(-5%)	(-14%)	(15%)	(8%)	(-27%)	(-6%)	(14%)
90-50	0.040	0.002	-0.005	-0.023	0.005	-0.002	-0.01	-0.015	-0.013
		(5%)	(-13%)	(-58%)	(13%)	(-5%)	(-25%)	(-38%)	(-33%)
50-10	0.058	0.009	0	0.009	0.009	0.009	-0.015	0.009	0.026
		(16%)	(0%)	(16%)	(16%)	(16%)	(-26%)	(16%)	(45%)

Notes: Bracketed terms refer to the % of the explained change. Columns (5) and (9) display how academy schools have contributed to changes in the statistics of interest. In columns (6)-(9), we reweight observations to account for the fact that when dropping academies, we alter the distribution of school characteristics and head teacher characteristics in the sample (where the characteristics are as set out in Table 5.2. We reweight by the inverse of the probability of remaining in the sample as a function of school and head teacher characteristics. The probabilities used in the reweighting are estimated via probit.

Table 5.6: Robustness Checks

	Log(Head Teacher Salary)		
	2001/02-2013/14	2001/02-2013/14	1995/96-2013/14
	Baseline	Fake Policy	Time Trend
	(1)	(2)	(3)
Academy*Pre 2010 Sponsored	0.068 (0.013)	0.015 (0.007)	0.037 (0.011)
Academy*Post 2010 Sponsored	0.027 (0.010)	-0.000 (0.007)	0.035 (0.008)
Academy*Converter	0.012 (0.004)	0.009 (0.003)	0.005 (0.003)
Sample Size	39,710	40,569	54,248
Number of Schools	2917	2917	2917
Number of Teachers	6633	6511	7904
Head Teacher Characteristics	Y	Y	Y
Year Effects	Y	Y	Y
School Characteristics (t-1)	Y	Y	Y
School Attainment	Y	Y	Y
School Fixed Effects	Y	Y	Y
Head Teacher Fixed Effects	Y	Y	Y

Notes: As for Table 5.2. Column (1) replicates column (7) of Table 5.2 and trims the sample so as to have the same sample structure as the fake policy experiment. This requires starting the sample five years later, in 2000/01, to ensure that each school is observed in the same number of years as it is in column (2). Column (2) replicates Table 5.2, but assumes a fake year of academy conversion that precedes the actual opening date by five years. All post conversions observations are then dropped and treatment variables are defined with reference to the fake opening date.

Table 5.7: Within Group Estimates

	Log(Head Teacher Salary), 2001/02 to 2013/14				
	(1)	(2)	(3)	(4)	(5)
Academy*Pre 2010 Sponsored*First Group	0.275	0.222	0.197	0.127	0.133
	(0.049)	(0.046)	(0.046)	(0.040)	(0.061)
Academy*Pre 2010 Sponsored*Second Group	0.209	0.172	0.163	0.135	0.056
	(0.022)	(0.022)	(0.022)	(0.015)	(0.013)
Academy*Pre 2010 Sponsored*Third Group	0.191	0.163	0.151	0.081	0.069
	(0.021)	(0.022)	(0.022)	(0.019)	(0.018)
Academy*Post 2010 Sponsored*First Group	0.180	0.146	0.145	0.124	0.017
	(0.039)	(0.040)	(0.041)	(0.037)	(0.016)
Academy*Post 2010 Sponsored*Second Group	0.077	0.075	0.070	0.076	0.036
	(0.017)	(0.017)	(0.017)	(0.015)	(0.012)
Academy*Post 2010 Sponsored*Third Group	0.082	0.059	0.058	0.066	0.007
	(0.042)	(0.043)	(0.042)	(0.036)	(0.018)
Academy*Converter*First Group	-0.012	-0.006	-0.012	-0.003	-0.016
	(0.029)	(0.027)	(0.026)	(0.020)	(0.011)
Academy*Converter*Second Group	-0.014	0.022	0.019	0.009	0.003
	(0.007)	(0.006)	(0.006)	(0.005)	(0.005)
Academy*Converter*Third Group	0.036	0.043	0.045	0.025	0.018
	(0.006)	(0.006)	(0.006)	(0.005)	(0.004)
Sample Size	36,793	36,793	36,793	36,793	36,793
Number of Schools	2917	2917	2917	2917	2917
Number of Teachers	6361	6361	6361	6361	6361
Head Teacher Characteristics	Y	Y	Y	Y	Y
Year Effects	Y	Y	Y	Y	Y
School Characteristics	N	Y	Y	Y	Y
School Attainment	N	Y	Y	Y	Y
School Fixed Effects	N	N	N	Y	Y
Head Teacher Fixed Effects	N	N	N	N	Y

Notes: Robust standard errors (in parentheses) are clustered at the individual level. Groupings are calculated based upon the number of pupils at each key stage and the proportion of these with special needs statements. Controls are included in each regression for the location of the school – fringe, inner London, or outer London. As some groups are small, we cluster together the first 4 groups, the next two groups, and the final two groups. Group 1 refers to those 4 groups with the lowest pay spine points as determined by the STPCD. Group 3 refers to the two groups in the highest pay band. Note that unlike our main specification, this Table only uses data from 2001/02 onwards given that we can only assign schools to groups (which is necessary to calculate the dependent variable) from this year due to data limitations.

Table 5.8: The Probability of Paying Outside of Pay Spine

	Pr[Paying Outside Pay Spine], 2001/02 to 2013/14			
	(1)	(2)	(3)	(4)
Academy*Pre 2010 Sponsored (Marginal)	0.441	0.453	0.438	0.430
	(0.029)	(0.029)	(0.029)	(0.029)
Academy*Post 2010 Sponsored (Marginal)	0.177	0.202	0.196	0.193
	(0.027)	(0.028)	(0.027)	(0.027)
Academy*Converter (Marginal)	0.057	0.034	0.033	0.034
	(0.013)	(0.013)	(0.013)	(0.013)
Sample Size	36,695	36,695	36,695	36,695
Number of Schools	2917	2917	2917	2917
Number of Teachers	6352	6352	6352	6352
Year Effects	Y	Y	Y	Y
Head Teacher Characteristics	N	Y	Y	Y
School Characteristics	N	N	Y	Y
School Attainment	N	N	N	Y

Notes: School level and individual controls are the same as those listed in the notes for Table 5.2. Note that unlike our main specification, this table only uses data from 2001/02 onwards given that we can only assign schools to groups (which is necessary to calculate the dependant variable) from this year due to data limitations. Off spine takes a value 1 if the head teacher is paid above the maximum payment spine for a school of their type (where type is defined by enrolment, intake of SEN pupils, and locality). A detailed description of how we compute the maximum payment, according to the STPCD, is given in the data appendix. Estimates are marginal effects from a probit model computed at the means of the covariates.

Table 5.9: The Probability of Employing a 'Superhead'

	Pr[Superhead] , 2001/02 to 2013/14		
	(1)	(2)	(3)
Academy*Pre 2010 Sponsored (Marginal)	0.079	0.041	0.026
	(0.029)	(0.028)	(0.029)
Academy*Post 2010 Sponsored (Marginal)	0.013	0.070	0.067
	(0.033)	(0.033)	(0.033)
Academy*Converter (Marginal)	0.070	0.044	0.045
	(0.014)	(0.012)	(0.012)
Sample Size	36,695	36,695	36,695
Number of Schools	2917	2917	2917
Number of Teachers	6352	6352	6352
Year Effects	Y	Y	Y
School Characteristics	N	Y	Y
School Attainment	N	N	Y

Notes: See the notes for Figure 5.4 for a discussion of how we define super heads. Unlike Table 5.7 that controls for head teacher characteristics, we do not do so in this table as the super head variable is defined based upon salary residuals that already have the characteristics netted out. Estimates are marginal effects from a probit model computed at the means of the covariates.

Appendix

Table A5.1: Secondary School Types in England, Biannually 2001/02 to 2013/14

	2001/02	2003/04	2005/06	2007/08	2009/10	2011/12	2013/14
Traditional School	2,914	2,910	2,900	2,869	2,702	1,754	1,201
Pre-2010 Sponsored Academies	0	7	17	48	117	115	113
Post-2010 Sponsored Academies	0	0	0	0	0	67	191
Converter Academies	0	0	0	0	0	701	1,135
Number of Schools	2,914	2,917	2,917	2,917	2,819	2,637	2,640

Notes: Calculations based upon the sample of schools that are used in the analysis. We plot types from 2001/02 due to the absence of academy schools, of any type, in all years prior to this.

Table A5.2: Average Estimates Using the Callaway and Sant'Anna (2021) Methodology

	Log (Head Teacher Salary)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Academy Pooled	0.051	0.044	0.047	0.046	0.042	0.039	0.045	0.042	0.040
	(0.006)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Head Teacher Characteristics	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
School Characteristics	N	Y	Y	N	Y	Y	N	Y	Y
School Attainment (t-1)	N	N	Y	N	N	Y	N	N	Y
School Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
Head Teacher Fixed Effects	N	N	N	N	N	N	N	N	N
TWFE	Y	Y	Y	N	N	N	N	N	N
Control Group	-	-	-	Never treated	Never treated	Never treated	Not yet treated	Not yet treated	Not yet treated
Number of Schools	2483	2483	2483	2483	2483	2483	2483	2483	2483
Number of Head Teachers	6867	6867	6867	6867	6867	6867	6867	6867	6867
Sample Size	47,177	47,177	47,177	47,177	47,177	47,177	47,177	47,177	47,177

Notes: Columns (4)-(9) use the procedure proposed in Callaway and Sant'Anna (2020) to aggregate group (in this case conversion year) year specific treatment effects into an overall average treatment effect. These estimates average over treatment effects that are specific to each academy conversion cohort and time period and weights these by group size to form an overall average. Columns (4)-(6) use schools that never convert to academy status as a control group while columns (7)-(9) use schools that have not yet converted as controls. Columns (1)-(3) present estimates using standard two-way fixed effects by way of comparison. Head teacher characteristics are age, tenure, gender, and the squares of age and tenure. School level characteristics include the proportion of free school meal eligible pupils, the log of full time equivalent pupils, and the log of full time equivalent teachers. School attainment includes the proportion of pupils achieving 5 or more A*-C at Key Stage 4 in the previous year. Standard errors, which account for clustering at the school level, are computed using the multiplier bootstrap. Estimates in the table derive from a balanced sample of schools.

Table A5.3: Average Estimates Using the Callaway and Sant'Anna (2021) Methodology, Heterogeneity by Academy Type

	Log (Head Teacher Salary)								
	Pre-2010 Sponsored			Post-2010 Sponsored			Converter		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Academy	0.124	0.113	0.116	0.083	0.084	0.085	0.022	0.018	0.012
	(0.021)	(0.022)	(0.022)	(0.020)	(0.021)	(0.021)	(0.005)	(0.005)	(0.005)
Head Teacher Characteristics	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
School Characteristics	N	Y	Y	N	Y	Y	N	Y	Y
School Attainment (t-1)	N	N	Y	N	N	Y	N	N	Y
School Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
Head Teacher Fixed Effects	N	N	N	N	N	N	N	N	N
Number of Schools	1293	1293	1293	1347	1347	1347	2917	2917	2917
Number of Head Teachers	3862	3862	3862	4029	4029	4029	6032	6032	6032
Sample Size	24,567	24,567	24,567	25, 593	25, 593	25, 593	42,066	42,066	47,177

Notes: As for Table A5.2. Table A5.3 allows for separate effects for each academy type by retaining only never treated schools and the schools that convert to the academy type of interest. The sample in columns (1)-(3) consists of never treated schools and those that become pre-2010 sponsored academies. The sample in columns (4)-(5) consists of never treated schools and those that become post-2010 sponsored academies while the final 3 columns consist of never treated schools and schools that later become converter academies.

Table A5.4: The Effect of Academies on Head Teacher Salaries, Balanced Panel

	Log (Head Teacher Salary)				
	(1)	(2)	(3)	(4)	(5)
Academy*Pre 2010 Sponsored	0.209	0.190	0.178	0.138	0.076
	(0.015)	(0.015)	(0.015)	(0.013)	(0.015)
Academy*Post 2010 Sponsored	0.063	0.090	0.085	0.077	0.032
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Academy*Converter	0.049	0.030	0.031	0.019	0.013
	(0.006)	(0.005)	(0.005)	(0.004)	(0.004)
Sample Size	47,177	47,177	47,177	47,177	47,177
Number of Schools	2483	2483	2483	2483	2483
Number of Head Teachers	6867	6867	6867	6867	6867
Head Teacher Characteristics	Y	Y	Y	Y	Y
Year Effects	Y	Y	Y	Y	Y
School Characteristics	N	Y	Y	Y	Y
School Attainment (t-1)	N	N	Y	Y	Y
School Fixed Effects	N	N	N	Y	Y
Head Teacher Fixed Effects	N	N	N	N	Y

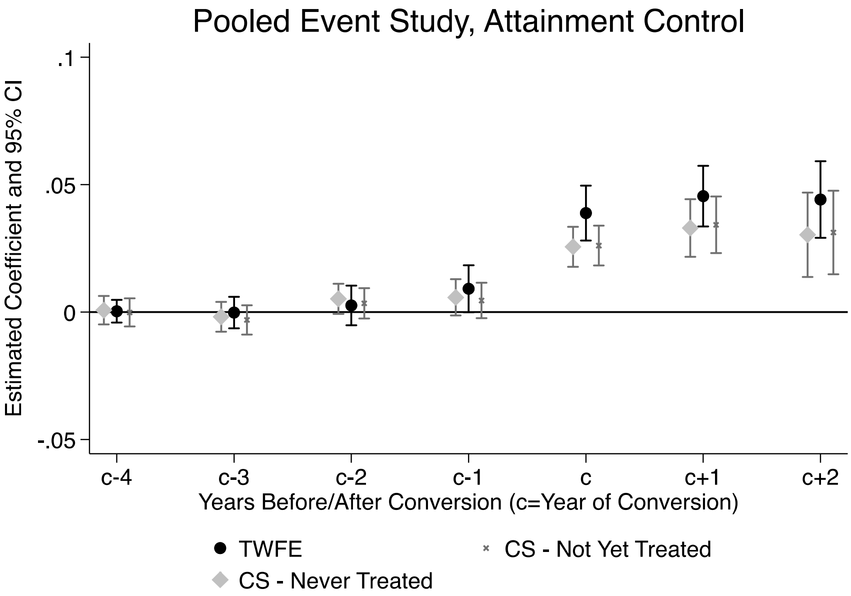
Notes: Robust standard errors (in parentheses) are clustered at the individual level. Head teacher characteristics are age, tenure, gender, and the squares of age and tenure. School level characteristics include the proportion of free school meal eligible pupils, the log of full time equivalent pupils, and the log of full time equivalent teachers. School attainment includes the proportion of pupils achieving 5 or more A*-C at Key Stage 4 in the previous year.

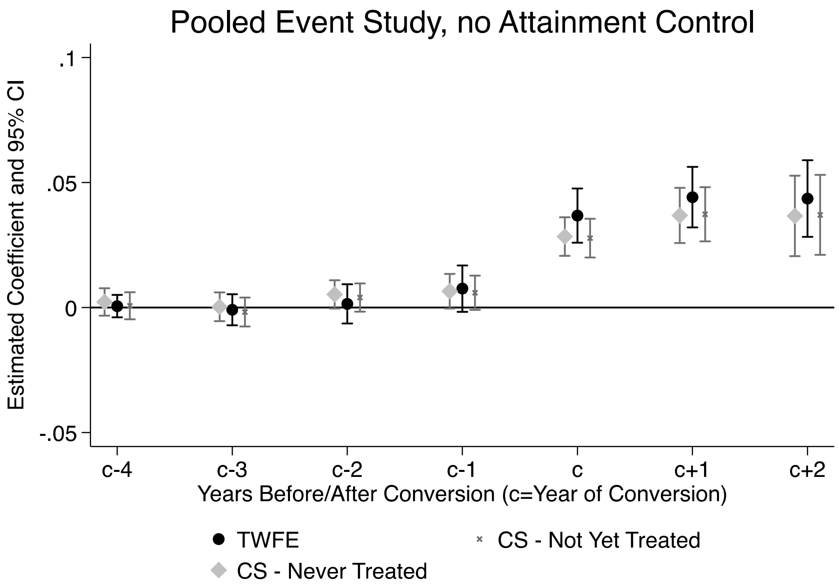
Table A5.5: RIF Oaxaca-Blinder Decomposition

	90-10	90-50	50-10	Variance (x100)
2001/02	0.325	0.163	0.162	2.011
2013/14	0.439	0.212	0.228	4.178
Change	0.114	0.049	0.065	2.168
Composition Effect				
School Characteristics	0.001	0.000	0.000	0.008
Teacher Characteristics	-0.009	-0.005	-0.003	-0.114
Performance	0.015	-0.007	0.021	0.197
Total Composition Effect	0.007	-0.012	0.018	0.091
Salary Structure				
School Characteristics	-0.097	-0.047	-0.050	0.359
Teacher Characteristics	-0.113	-0.111	-0.002	-0.592
Performance	-0.074	-0.011	-0.062	0.100
Pre-2010 Sponsored	0.008	0.002	0.005	0.119
Post-2010 Sponsored	-0.001	-0.004	0.003	0.054
Converter	0.007	-0.010	0.017	0.091
Constant	0.377	0.241	0.136	1.946
Total Salary Effect	0.108	0.060	0.047	2.077
Sample Size	5,554	5,554	5,554	5,554

Notes: The results here are from a two-way unweighted Oaxaca Blinder decomposition that uses RIF regressions to uncover the effect of marginal changes in covariates on functionals of the unconditional salary distribution (see Firpo, Fortin, and Lemieux 2018 for a detailed description of this method). Here we use 2001/02 as the baseline year. Because there are no academies at baseline we fix the 'price' of academies to 0 in the baseline year. In doing this, we can only measure the total effect of academies (the sum of the salary structure and composition term) on the statistics of interest. Head teacher characteristics are age, tenure, and gender. School level characteristics include the proportion of free school meal eligible pupils and the number of full time equivalent pupils. School performance includes the proportion of pupils achieving 5 or more A*-C at Key Stage 4 in the previous year.

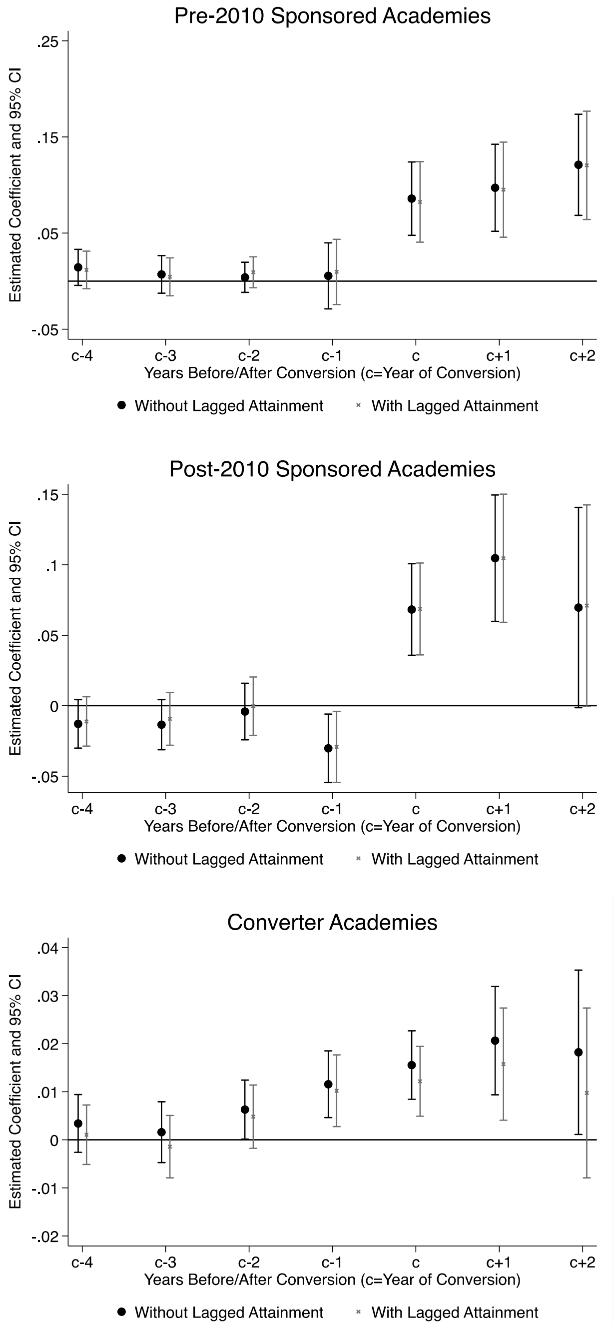
Figure A5.1: Pooled Event Study using the Callaway and Sant'Anna (2021) Methodology





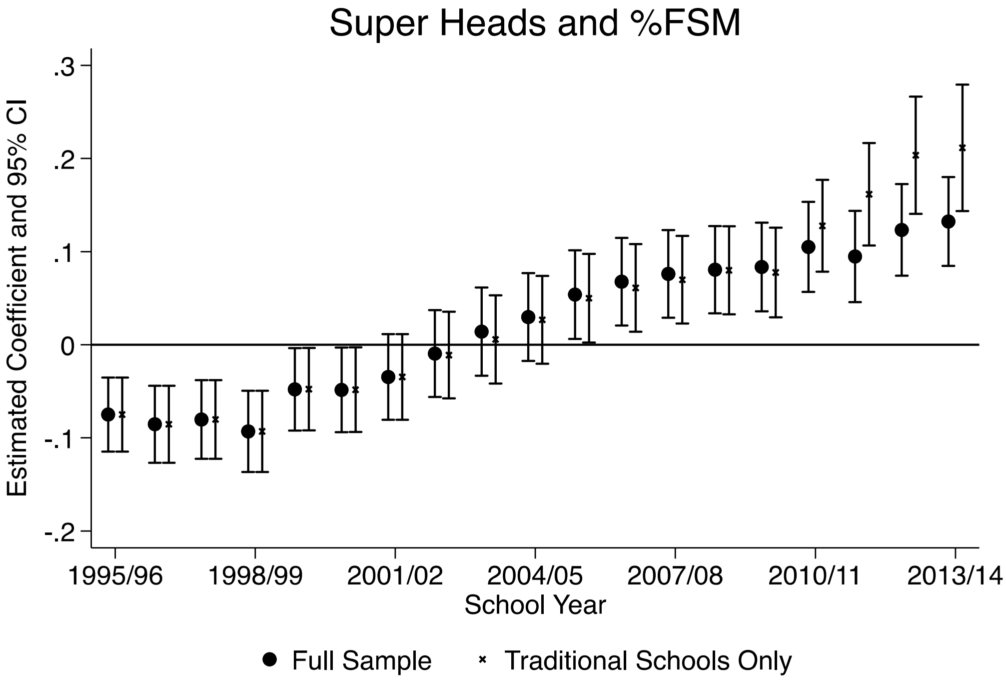
Notes: Figures use the methodology proposed in Callaway and Sant'Anna (2021), to correct for bias in event study estimates where treatment timing differs, and effects are heterogeneous across conversion cohorts. Event dummies in this case pool across the three academy types and we partially balance our sample as we do in our main event study estimates (we omit years prior to 5 years before the data of conversion and year 3 or more years after). In line with Figure 5.5, we also estimate our event study both with and without lagged attainment measures. As Callaway and Sant'Anna (2021) propose using either the never treated (those that do not convert to academies) or not yet treated (those who convert in the sample period but have not converted in the year of interest) to estimate group/time specific effects, we present estimates using both potential control groups in the above. As can be seen, either of these appear to be a suitable control and give quantitatively similar results to standard TWFE estimates.

Figure A5.2: Event Study using the Callaway and Sant'Anna (2021) Methodology



Notes: This figure replicates Figure 5.5 using the methodology proposed in Callaway and Sant'Anna (2021). To estimate separate effects by academy type, we retain only observations of schools that never convert and schools that convert to the academy type of interest. We then run separate event studies, using the CS method on these three separate samples. We use a balanced panel and truncate each sample to exclude observations that fall more than 3 years after conversion or 4 years before.

Figure A5.3: The Relationship between Hiring a Super Head and %FSM



Notes: Coefficients in A5.4 are obtained by calculating the percentile of each school in the %FSM distribution each year. We then regress whether the school has a super head, as defined in the main text, on this percentile for each year in the sample before plotting the coefficients. The figure shows estimates of this relationship for all schools as well as estimates from a restricted sample where the percentiles are based upon traditional state schools only and academies are excluded from the sample.

Data Appendix

Sample selection

The focus of the paper is to study whether academy schools have changed the salary structure for head teachers over time. In line with this, we consider a sample of headteachers drawn from state run secondary schools between the years 1995/96 and 2013/14. Our final dataset excludes head teachers at special schools and independent schools, both of which are subject to different rules regarding remuneration. Our final sample imposes a number of additional restrictions. As our data is drawn from the universe of the school workforce, we have to identify head teachers using a service indicator variable. For the limited number of state-run secondary schools for which we cannot do so, we drop these observations. We include only those schools that we can match with control variables using school performance table data. Therefore, we retain records for headteachers employed at schools for whom we have data on the school number of on roll pupils, the proportion of pupils eligible for FSM, and the proportion receiving 5 A*-C at GCSE. Our final restriction is to focus on schools that have a record for 15 consecutive years. Headteachers at schools that are recently opened, with respect to the sample timeframe, or schools that open late in our sample period are removed from the sample.

Academy Schools

While we remove recently opened schools, we do include academy schools that are in operation for less than 15 years. The reason we do so is that most academies in our sample period are conversions of already existing schools. While these schools gain extra freedoms, and become exempt from pay legislation, they operate on the same site and pupils enrolled in academies prior to conversion need not re-apply for admission. The fact that already open schools can become academies also provides us with useful variation in specifications where both headteacher and school fixed effects are included.

We use Edubase data to identify academies that open during our sample period and to link these schools to their predecessor school. In all specifications that include school fixed effects, we treat the academy and its predecessor as the same school[142]. Edubase also provides us with information on the date at which schools become academies and the type of academy that they become – either sponsored or converter. Post 2010, a significant number of academy schools report opening mid-year i.e. after the 1st of September of the current school year. We classify schools as opening in a given academic year if they open in that calendar year. Therefore, if a school opens between 1st January and 31st December 2010 they are deemed to open in the 2010/11 academic year. We use this same cut-off when assigning sponsored academies to either post 2010 or pre-2010.[143]

Pay Spine estimates

A key part of our analysis is understanding whether the liberalisation of salary setting in academies led to increases in both the level and dispersion of head teacher pay. One indicator of whether this was so is whether academies were more likely to pay outside of pay spines set out in annual releases of the School Teachers' Pay and Conditions Document (STPCD).

[142] There are a few instances in which multiple schools consolidate into a single academy. We retain records only for schools where a single schools converts to academy status.
[143] All converter academies open pursuant to the Academies Act 2010.

In the period for which we use our ‘off spine’ measure, the document changes very little. In each year, schools are assigned to one of 8 groupings based upon the number of pupils in each key stage and the number of pupils on roll with special educational needs (with a statement). Once assigned to a group, a payment structure is agreed whereby head teachers progress along a number of pay spine points for a school within that group. As there is a maximum pay spine for each school group, the STPCD sets out an upper bound on what each school can pay head teachers in any year. While academies are free to pursue their own pay policies, state-maintained schools may also pay above the maximum of their head teacher group range when the school is a school causing concern, paying below the maximum would lead to difficulty in recruiting a head teacher for a vacant post, or where paying below the maximum would lead to difficulty in retaining an existing headteacher.

To determine the maximum of the group range for a given school in a given year, we use the formula set out in the STPCD. This formula gives a point score based upon the number of pupils in each key stage with children in Key Stage 2 (grades 3-6) and below given 7 points, those in Key Stage 3 (grades 7-9) given 9 points, those in Key Stage 4 (grades 10-11) given 11 points, and those in Key Stage 5 (grades 12-13) given 13 points. Each pupil with a special needs statement is given 3 times the points for their key stage. The sum of points over a school gives a total unit score that then assigns the school to a group as follows:

Total Unit Score	Group
up to 1000	1
1001 to 2200	2
2201 to 3500	3
3501 to 5000	4
5001 to 7500	5
7501 to 11,000	6
11,001 to 17,000	7
17,001 and above	8

For each group, the document sets out a maximum salary that varies slightly depending upon the region with head teachers in inner London, outer London, and the London fringe being paid slightly more. We use data from the school census that allows us to calculate points scores for each school, based on numbers of pupils in each key stage, and data from Edubase, which allows us to classify schools by region, coupled with the annual STPCD documents to derive a maximum salary for each year/school pair in our sample. Our ‘off spine’ measure then takes a value of one if the headteacher is paid above this maximum. As we do not have comprehensive data on enrolment by key stage, and fraction of SEN pupils, prior to 2002, our analysis using this variable begins in the 2001/02 school year.

Had we been able to extend our analysis before 2001/02, we would have encountered a slightly different pay regime for the earliest years in our sample - 1995/96, 1996/97, and 1997/98. As can be seen by the table comparing the groupings and pay ranges in the crossover years, the changes in 1999 increased pay at every level of school size and led to a vast increase in potential remuneration for those in the largest (when weighted by age) schools.

STPCD Documentation in 1998/99

Key Stage	Units	Total Unit Score	Group	Spine Points	Pay Range
For each pupil under 14 years of age	2 units	up to 300	1	3-15	27,204-32,382
For each pupil aged 14 and under 15	4 units	301-700	2	8-22	29,355-35,691
For each pupil aged 15 and under 16	5 units	701-1,300	3	15-29	32,382-39,726
For each pupil aged 16 and under 17	7 units	1,301-2,400	4	23-37	36,270-45,483
For each pupil aged 17 and over	9 units	2,401-4,600	5	31-44	41,163-52,533
		4,601 and above	6	38-51	46,488-59,580

STPCD Documentation in 1999/00

Key Stage	Units	Total Unit Score	Group	Spine Points	Pay Range
For each pupil at key stage one or two	7 units	up to 1,000	1	1-9	31,155-37,947
For each pupil at key stage three	9 units	1,001 to 2,200	2	3-12	32,733-40,848
For each pupil at key stage four	11 units	2,201 to 3,500	3	6-15	35,244-43,971
For each pupil at key stage five	13 units	3,501 to 5,000	4	9-18	37,947-47,322
		5,001 to 7,500	5	13-22	41,868-52,194
		7,501 to 11,000	6	16-26	45,060-57,570
		11,001 to 17,000	7	19-30	48,498-63,498
		17,001 and above	8	23-34	53,490-70,002

Notes: Pay ranges and groupings are set out in the annual STPCD documents. The pay ranges above apply to head teachers at state-maintained schools outside of inner London, outer London, and the London fringe. Schools in these regions are subject to the same formula, but have slightly higher upper and lower pay bounds. Our analysis takes into account these differences when classifying head teachers as being above the maximum pay range for the school group

Estimating Transitions and Classification of 'Super Heads'

Table 3 estimates salary returns to being a head teacher at an academy and allows them to differ depending upon where the teacher was in the previous year. Table 8 classifies heads as being super heads according to their position in the salary distribution in previous years.

In each of these cases, we use the full sample of head teachers before we impose restrictions to derive measures of both pay percentile and previous school type. We do this to avoid dropping heads who transition between schools in our sample and those not in our sample. In practice, this means those who transition from a school that is in operation less than 15 years, over the course of our sample period, to a school that we include in our analysis.

The same reasoning holds when classifying head teachers as ‘super heads’. We define a super head as someone who, in any previous academic year, was at or above the 90^{th} percentile of the residual earnings distribution. We only look at the earnings distribution for traditional schools as academy pay mixes the premium to being at an academy with the premium to being a super head. When looking at salary percentiles, we include head teachers who are at schools that are not in our final sample.

General Conclusions

The five chapters of this thesis explore social mobility and educational reforms in the United Kingdom.

Past research has highlighted that income mobility in the UK is low relative to international standards (Corak, 2013) and has declined over time (Krutikova et al 2023).

Chapter 1 of this thesis explores whether the same holds true for wealth mobility. Wealth mobility is notoriously hard to measure due to the difficulty of collecting both accurate wealth data and linking this data for a given cohort and their offspring at an appropriate point in the life cycle. This chapter bypasses this difficulty by using detailed data linking one's wealth to both housing tenure and the value of one's main property. This is then combined with intergenerational home ownership correlations to back out estimates of intergenerational wealth transmission. Evidence from cohort studies shows that the relationship between housing tenure and that of one's parents has strengthened significantly over time. When combined with evidence of the connection between home ownership and wealth, these results are suggestive of a fall in intergenerational wealth mobility across birth cohorts.

While Chapter 1 focuses on the measurement of mobility, Chapter 2 instead asks whether the determinants of mobility have changed across birth cohorts. I use from the machine learning explainability literature, combined with a latent factor framework, to highlight how patterns of upward mobility have changed between those born in 1958 and 1970. This methodological approach brings new insight into the literature that tries to explain the determinants of income mobility (Blanden, Gregg, and Macmillan, 2007; Bolt et al, 2021). While a host of factors influence the likelihood of upward mobility – household composition, socioemotional skills, school quality, and parental time investment – I highlight that these factors exert little independent influence on mobility prospects once years of schooling and cognitive ability in late childhood are accounted for. This is true even when allowing for flexible patterns of interactions between, say, socioemotional skills and cognition.

After establishing the primacy of educational attainment and cognition in driving mobility outcomes, the final three chapters focus on one of the most encompassing educational reforms to take place in England in the last two decades – the academies programme. Academies are state funded schools that are allowed to run in an autonomous manner outside of local education authorities. Chapter 3 focuses on the initial batch of academies that were conversions of low-performing state schools. These schools were in deprived urban areas and catered to disadvantaged student populations. It is shown that, relative to comparable schools, these schools generated strong test score gains for pupils; however, soon after conversions we also find that these schools begin to attract a more advantaged school population.

Although the initial programme was focused on poorly performing secondary schools, the later expansion of academies covered both primary schools and secondary schools. Chapter 4 highlights how the initial success of academies failed to replicate when scaled up to improve primary schools. We highlight that test results in primary converters remained unchanged upon conversion. We link this null result to the fact that primary converters were strong performing schools to begin with. Looking at financial data, we also find that extra money that accrued to these schools upon conversion tended to be spent on operational expenses that are unrelated to academic performance.

Chapter 3 offers some insights as to why early sponsored academies managed to increase test score performance; namely, greater operational autonomy coupled with changes in school leadership. The final chapter of the thesis explores how these schools – and the academies that opened after 2010 – shaped head teacher pay in England. Although head teacher pay has increased (and become more unevenly distributed) in the last two decades, we argue that this is largely unrelated to academy schools. The emergence of "super heads", whose high levels of remuneration have driven the inequality rise, is not a direct consequence of the academisation of English schools but reflects a more general shift towards the market determination of salaries in the English secondary school sector.

Taken together, the thesis highlights how intergenerational mobility – both in terms of wealth and income - is changing in Britain and the primacy of education and cognition in shaping overall mobility outcomes. This suggests that educational policies are a key lever by which the decline, and subsequent stasis, in mobility prospects can be reversed. However, educational reform is hard. Reforms such as the early 2000s academies program, which did raise attainment for disadvantaged children, do not always scale up. Along with this, secular trends in education, such as the growing disparity in headteacher incomes have the potential to increase educational inequality and exacerbate the forces that keep upward mobility in the UK at internationally low levels.

References

Blanden J, Gregg P, and Macmillan L (2007) Accounting for Intergenerational Income Persistence: Noncognitive Skills, Ability and Education, *The Economic Journal*, 117, 43-60.

Bolt U, French E, Hentall-MacCuish J, and O'Dea C (2021) Intergenerational altruism and transfers of time and money: a life cycle perspective, Institute for Fiscal Studies.

Corak, M (2013) Income Inequality, Equality of Opportunity, and Intergenerational Mobility, *Journal of Economic Perspectives*, 27 (3), 79-102**.**

Krutikova, S, L Macmillan, L van der Erve, and D Sturrock (2023), Intergenerational mobility in the UK, Institute for Fiscal Studies.