The London School of Economics and Political Science

Does geography matter? An empirical investigation into neighbourhood, peer effects and electricity consumption

Felix Julian Weinhardt

A thesis submitted to the Department of Geography and Environment of the London School of Economics for the degree of Doctor of Philosophy, London, February 2012.

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

London, 07/02/2012
Felix Julian Weinhardt

Statement of conjoint work

Two out of the four chapters that form part of this thesis involve conjoint work, as specified below:

Chapter III "Do your neighbours matter? Evidence from students' outcomes in England" is conjoint work with Steve Gibbons and Olmo Silva. This is work undertaken in the preparation for a joint paper titled: "Everybody needs good neighbours? Evidence from students' outcomes in England.", which is currently published as IZA working paper and submitted to a peer-reviewed journal. Together with the co-authors I have contributed to the paper in all of the following aspects:

Development of research question and methodological approach

Data construction

Statistical analysis using STATA

Conceptual framework and robustness checks

Presentation of the work at seminars and conferences, including at the SOLE 2011 conference in Vancouver and various LSE seminars.

Full first write-up of the paper.

Overall, my contribution amounts to two thirds of the total paper.

Chapter IV "The good, the bad and the average: evidence on ability peer effects in schools" is conjoint work with Olmo Silva and Victor Lavy. A later version of this project is accepted for publication by the Journal of Labor Economics.Together with the co-authors I have contributed to the paper in all of the following aspects:

Data construction

Statistical analysis using STATA

Conceptual framework and robustness checks

Presentation of the work at seminars and conferences

Drafting of the paper.

Overall, my contribution amounts to one third of the total paper.

Abstract

This thesis consists of four distinct projects which sit at the crossroad between Labour, Education and Environmental Economics. The underlying and unifying theme is the examination of social and geographical inequalities using applied econometrics.

In the first project, I estimate the effect of moving into a deprived high-density social housing neighbourhood on the educational attainments of teenagers in England. I exploit the timing of moving, which can be taken as exogenous because of long waiting lists for social housing in high-demand areas, to avoid the usual sorting problems. Using this strategy, I find no evidence for negative effects.

The second project investigates the effect of neighbours' characteristics and prior achievements on teenagers' educational outcomes. The study relies on mover-induced variation in neighbourhood quality, whilst controlling for general gentrification trends and other unobservables. The results provide little evidence for significant effects on pupil test score progression.

The third project looks at the size, significance and heterogeneity of ability peer effects in secondary schools in England. The methodological innovation is to identify ability peer effects using within-pupil-across-subject variation in students' test scores and peer prior achievements. The chapter shows that it is the low- and high-achievers, who account for most or all of the effect of average peer quality on the educational outcomes of other pupils and that this effect varies across genders.

The final project presents -to the best of my knowledge- the first nationwide empirical assessment of residential electricity use in response to the timing of daylight for the US. Employing Geographical Information Systems (GIS), I calculate the solar times of sunrise and sunset for all locations in mainland US and show that two distinct sources of geographical variation can be used to estimate county-level responses in residential electricity consumption. Using both approaches I find that early sunrise is associated with lower residential electricity use in the North, but higher consumption in the South. This is a novel finding with potentially significant policy implications and I offer some suggestions about how future research should examine the behavioural channels that could cause these results.

Acknowledgements

First and foremost I would like to thank my supervisor Olmo Silva, who has always supported me and my work with his insights and suggestions. Indeed, from the beginning of my PhD I have felt lucky to have Olmo as a supervisor, and later also as a co-author and a friend.

Further, I thank my review supervisor Henry Overman. Henry has also been my tutor during my undergraduate and master degrees and supported me throughout. His support has indeed been critical for my academic development in a number of occasions, including but not limited to critical discussions of my work, funding applications and setting up my visit to John Quigley at UC Berkeley in Spring 2011.

Over the past years I have also greatly benefited from working with my co-authors Steve Gibbons and Victor Lavy, from whom I have learnt a lot both academically and personally.

ESRC PhD funding is gratefully acknowledged (Ref: ES/F022166/1).

I also gratefully acknowledge countless discussions with fellow students, in particular with Ben Faber, Richard Murphy, and with Rosa Sanchis-Guarner Herrero.

I thank Christian Hilber, who encouraged me to pursue a PhD in my early days as undergraduate student. I thank Sandra McNally for her support with data applications.

The work presented in the various chapters also benefited from discussions and feedback from a number of individuals.

For Chapter II, I thank Christian Hilber, Victor Lavy, Tim Leunig, Richard Murphy, Henry Overman, Frédéric Robert-Nicoud, Olmo Silva and Yves Zenou, and participants of the EALE/SOLE conference in London, the NARSC Urban Economics Association Conference in San Francisco, the LSE-SERC Urban Economics Seminar, the LSE-CEP Labour Market Workshop, the 2010 IZA Summer School, the 2010 ZEW Workshop on "Evaluation of Policies Fighting Social Exclusion"; and the CEP-CEE Education Workshop for helpful comments and discussions.

For Chapter III, I (and my co-authors) would like to thank Peter Fredriksson, Hilary Hoynes, Rucker Johnson, Jens Ludwig, Richard Murphy, Henry Overman, Marianne Page, Stephen Ross, Jesse Rothstein, and seminar participants at the SOLE Annual Meeting 2011, UC Berkeley, Bocconi University, the CEP Labour Workshop, the CEE Informal Meetings, CEIS-Tor Vergata, HECER Finland, the CEP Annual Conference 2010, IFAU Uppsala, the SERC Annual Conference 2010, University of Bologna, University of Milan 'Bicocca', University of Stockholm and the Workshop on "The Economics of the Family and Child Development" June 2010 at the University of Stavanger for helpful comments and suggestions.

Regarding Chapter IV, I (and my co-authors) thank the Centre for Economic Performance (CEP) for seed money for this project and acknowledge comments and discussions from/with Rebecca Allen, Josh Angrist, Kenneth Chay, Steve Gibbons, William R. Johnson, Francis Kramarz, Steve Machin, Henry Overman, Michele Pellizzari, Steve Pischke, Steve Rivkin, Yona Rubinstein, Hongliang (Henry) Zhang and seminar participants at Bocconi University, Brown University, CEP-LSE, CMPO-Bristol University, CREST in Paris, EIEF in Rome, ESSLE 2009, Hebrew University of Jerusalem, IFS and IoE in London, Royal Holloway University, and Tel Aviv University.

Regarding Chapter V, I thank Ben Faber, David Grover, Charles Kolstad, Henry Overman, Steve Pischke, John Quigley and Olmo Silva, as well as LSE-SERC seminar and LSE Labour market workshop participants for helpful comments and suggestions.

Last but not least, I have to thank my family for everything they have given me. I have always felt the invaluable support and love of my mother Beate and sister Clara. I am saddened that my father has not been able to witness most of my academic development. His love and inquisitive mind have greatly influenced my personal development and helped me to build confidence regarding my own judgements, which has certainly influenced my decision to pursue a PhD.

Most especially, I would like to dedicate this work to Barbara, my wife. It is for her that I will always look back at these years as a time full of academic challenges and personal happiness. I also thank my daughter Helena for the additional joy she brought into my life and for keeping up without seeing much of me especially during the final phase of the write-up.

Contents

Index of tables and figures

Chapter II

Chapter III

Chapter IV

Chapter V

CHAPTER I

INTRODUCTION

1 Does geography matter?

Until the early twentieth century Geography was a discipline based around local, detailed knowledge of particular places. The underlying assumption was that we shoud study each individual location[1] in order to gain an understanding of the world. The importance of Geography was thus implied by the approach of the discipline in itself.

With the 'quantitative revolution' in the middle of last century, some geographers started to study variations in phenomena over space rather than just focussing on a particular location. Geography has always been a largely quantitative science, but today increased attention is paid to theory. An example is the sub-field of *Economic Geography*, in which the Marxist geographer David Harvey is the most influential figure. David Harvey theorises how space, in particular the housing market and the built environment, is instrumental to the survival of capitalism, most prominently in Harvey (1973). This perspective is then used to study socioeconomic phenomena. Other examples taken from the natural sciences are gravity theory or entropy-maximising models. From Regional Science came 'Central Place Theory', or Weber's industrial location theory (Johnston and Watts, 2000).

This introduction of theory had important consequences for the way geographers study location. With the move of the discipline towards theory, the importance of geography has become testable. Rather than being presumed according to the discipline, theoretical models can be used on the data, and we can test the importance of geography in a positivist sense.

Similarly, Paul Krugman started to re-introduce space into mainstream theoretical economics in the early 1990s (Krugman, 1992). Today many economists study the importance of location for individual or firm outcomes, and the new sub-discipline of *Spatial Economics* has emerged (Duranton and Rodríguez-Pose, 2005; Overman, 2004).

To the present day, for many questions it remains highly contested whether geography, or space, simply offers a useful 'container' or perspective to study social phenomena, or whether geography has an independent explanatory power.

The approach taken by this thesis is to assess the importance of location-mediated phenomena in a number of distinct settings using applied econometrics. Thus this thesis presents an indirect test of geography. I briefly summarise the four main chapters

[1] In this thesis I will use the words geography, environment, space and location interchangeably.

and how each of them tries to assess causal importance of geography in section 3 below. First, the next section outlines the key methodological challenges to be overcome when trying to assess causality in a geographical setting.

2 The quantitative empirical approach

2.1 Potential outcomes and selection

Parallel to the increase in awareness of location patterns in mainstream economics, there has a development in applied research that Pischke and Angrist term the 'credibility revolution'(Angrist and Pischke, 2008). Led by the sub-discipline of *Labour Economics*, applied researchers have increasingly turned their attention to the question of causality. In particular, we have a much better understanding today of how research design can support the causal interpretation of statistical associations. The applied quantitative researcher is equipped with a multitude of statistical tools to support her studies and claims of causality. In this section I outline the key challenges that need to be addressed in order to claim causality in geographical settings.

All quantitative empirical work that tries to assess the impact of one particular variable on another is faced with a fundamental problem: the counterfactual is never observed. To understand the importance of the environment for individual outcomes, for example, one would need to observe the same person at the same time in different environments in order to support causal argumentation. This is not possible.

As first shown by Rubin (1974) we can derive the consequences this has for empirical work using the potential outcomes notation. For simplicity, let us call some characteristic of the environment that we are interested in 'treatment', denoted by D, which takes the values 1 and 0. Following Angrist and Pischke (2008), when Y denotes the outcome of interest for the same individual i, there are two possible states:

$D = 1$, the person was treated. The expected outcome is denoted by $E(Y|D = 1)$.[2]

$D = 0$, the person was not treated. The expected outcome is denoted by $E(Y|D = 0)$

The causal effect of the treatment γ can then be represented by the difference in

[2]Subscipt i omitted for simplicity. We also assume homogeneous treatment effects.

expected outcomes:

$$\gamma = E(Y|D=1) - E(Y|D=0) \tag{1}$$

The fundamental problem is that each individual can only be observed in one or the other outcome. In other words, we can only observe the effect of the treatment for people who were actually treated. Denoting the outcome of an individual who was treated with Y_t and an outcome for an individual who was not treated with Y_c, what we can actually observe is:

$$\theta = E(Y_t|D=1) - E(Y_c|D=0) + E(Y_t|D=0) - E(Y_t|D=0) \tag{2}$$

rearranging yields:

$$\theta = E(Y_t|D=1) - E(Y_t|D=0) + E(Y_t|D=0) - E(Y_c|D=0) \tag{3}$$

The first two terms now represent the 'treatment on the treated', that is, the effect of the treatment on people who were actually treated. The latter two terms represent the difference in expected outcomes between individuals in the treatment group and those in the control group in a world where neither were actually treated. This quantity is referred to as 'selection bias'. In programme evaluations selection bias becomes a problem if people self-select into treatment and control groups or if treatment status is not assigned randomly in an experimental setup, for example. However, selection bias can equally occur in other situations. Any pre-treatment outcome-relevant unobserved differences between those treated and those who were not would cause omitted variable bias. As a result of this, we need to carefully assess whether there might be any differences between individuals -correlated with the outcome under investigation- that are potentially not independent of treatment assignment in each situation.

The most straightforward way to achieve this would be to randomise treatment status, and as a result randomised experiments are regarded as the 'gold standard' in applied empirical work. Unfortunately randomisation is rarely possible especially once geographical treatments are considered. As a result, this thesis pays particular attention to research design. Each section will outline the strategy followed to identify causal relationships and the required assumptions will be discussed.

2.2 The selection problem in the geographical context

The selection problem outlined in section 2.1 is particularly severe when studying geographical patterns. This is because it is very hard to find distinct locations that only differ along a clearly specified and observable dimension. Further, when asking about the importance or influence of geography or some location-mediated phenomena on some particular outcomes, we often do not have clear mechanisms in mind.

Many of the methodological challenges for quantitative empirical research which tries to assess in a causal way the importance of location-mediated phenomena - or what happens around us - have been formalised in the neighbourhood effects literature. Indeed, all parts of this thesis have to deal with methodological challenges that can be understood in reference to the neighbourhood effects research setup. Therefore, I now present a synopsis of 'identification'-problems in neighbourhood effects research.

In a seminal paper Manski (1993) spelled out these identification problems. He considers the following empirical model:

$$Y_{in} = \alpha_0 + \alpha_1 X_i + \alpha_2 X_{n,-i} + \alpha_3 Y_{n,-i} + \delta_n + \epsilon_{in} \tag{4}$$

In this specification, the term Y_{in} represents an individual i's outcome living in neighbourhood n. The term X_i represents individual-i-specific characteristics that have an influence on the outcome under consideration. The two terms that can be understood as capturing location-mediated phenomena are $X_{n,-i}$ and $Y_{n,-i}$. The former presents a measure of characteristics of people who live in neighbourhood n. The subscript $-i$ denotes that this measure is calculated net of the characteristics of individual i. The latter term presents a measure of outcomes or actions that are undertaken by the individuals who live in neighbourhood n, again measured net of the outcome of individual i.

The coefficients of interest are α_2 and α_3. Since Manski (1993), α_2 is called the 'exogenous' or 'contextual' effect, which is the effect that comes from the characteristics X of the neighbours. When considering effects on school outcomes for teenagers, this could, for example, capture the socioeconomic background of the neighbours. The coefficient α_3, on the other hand, would capture the effect of what neighbours are doing, that is an effect of neighbours' outcomes, which are denoted by $Y_{n,-i}$. In the education-related example, this could represent average school results of neighbours. Finally, the term δ_n captures neighbourhood-specific, potentially unobserved characteristics that influence the outcome. Manski (1993) calls this the 'correlated effect'. In this setting a number of identification problems are present but the selection mechanism is particularly problematic.

To see this, note that δ_n potentially captures any common shock or neighbourhood characteristic that has an effect on the outcome. In the study of neighbourhood effects on educational outcomes, a common candidate would be local school quality, for example. This is problematic because the shocks or neighbourhood level variables that are captured by δ_n are potentially correlated with both neighbours' characteristics ($X_{n,-i}$) and their actions ($Y_{n,-i}$). The reason for this is that people 'sort' themselves into their neighbourhoods based on preferences and income through the housing market. Therefore, it is likely to find a correlation between potentially unobserved neighbourhood quality variables and the characteristics of the people living in these neighbourhoods. However, if unobserved δ_n correlates with the neighbourhood measures, this results in omitted variable bias similar to the selection bias discussed in the previous section.

To clarify this further, consider the following example: assume that we are interested in measuring the effect of the presence of very well-performing teenagers at school on the educational outcomes of other teenagers. In reference to equation (4), we want an estimate for α_3. To do this, our approach is to compare the outcomes of individuals living in two different neighbourhoods, A and B, and for simplicity let's assume that only neighbourhood A has such high-performing individuals.

In this setting, following the discussion in section 2.1, the problem is that we cannot observe the same individual living in both neighbourhoods. Instead, we can only compare the outcomes of different individuals across neighbourhoods. Thus we can only hope to measure the effect of the 'treatment on the treated'. As we have seen, the critical condition that needs to hold is that the expected outcomes of the individuals in the two neighbourhoods would have been equal in the hypothetical absence of the high-performers in neighbourhood A. Using the potential outcomes notation, this selection term can be written as $E(Y_A|D = 0) - E(Y_B|D = 0)$, where Y_A represents teenagers' test scores in neighbourhood A, and D denotes treatment status - the presence of high-performers.

In this neighbourhood setting it is very unlikely that this term equals zero. This is because the two neighbourhoods are likely to differ along other dimensions that correlate with the presence of high performers. In particular, consider the possibility that neighbourhood A has better schools, partially unobserved by the researcher. Since school quality is likely to correlate with individual test outcomes, this would clearly induce bias. Worse, the selection problem in the geographical setting has a further dimension: We know that neighbourhood-level amenities such as school quality capitalise into house prices, as for example shown by Black (1999). As a result of the higher house prices, we would further expect neighbourhoods to differ in terms of the socioeconomic status of their residents, and potentially many other local amenities that could influence school results. Because of this kind of sorting of people with higher socioeconomic status through the housing markets all these potentially unobserved 'correlated' effects could be wrongly attributed to the effect of the presence of high performers on individual outcomes.

Throughout this thesis, I will pay particular attention to this sorting problem, which is the geographical counterpart of the more general selection problem in applied quantitative work.[3]

The next section gives a brief summary of the four research projects presented here.

3 Overview of thesis

This thesis consists of four main research projects, presented in Chapters II to V.

3.1 Chapter II: Evidence on the importance of deprived neighbourhoods

The starting point for the question addressed by the second chapter is that poverty and levels of educational attainment are not distributed evenly across space. Conversely, we observe that poverty is concentrated in a small number of deprived neighbour-

[3] A separate -and often confused- identification problem in social interaction research is that α_3 cannot be identified due to reverse causality issues. Recall that α_3 is meant to capture the effects of neighbours' outcomes on individual outcomes. If these effects are non-zero, then individual i's outcomes will naturally also affect the neighbours' outcomes $Y_{n,-i}$. Chapters II, III and IV solve this problem using predetermined measures of outcomes for the neighbours, which will be explained in detail. The reflection problem does not cause any issues for the analysis presented in Chapter V.

hoods, often in the largest cities. Here, we also observe students with the some of the poorest outcomes at school. This gives rise to the possibility that the concentration of poverty itself causes school disadvantage, which would have important consequences for social mobility. Indeed, the generation and perpetuation of social inequalities and the concept of 'equality of chances' lies at the heart of our understanding of modern democratic society, i.e. (Sen, 1980).

Researchers in both Economics and Sociology have identified a number of mechanisms to explain why living in deprived areas might exert negative influences in its own right: peer group and role model effects could explain why our behaviour depends on others around us (Akerlof, 1997; Glaeser and Scheinkman, 2001). Alternative candidates are social networks (Granovetter, 1995; Calvo-Armengol and M.O., 2004; Bayer et al., 2008; Zenou, 2008; Small, 2009; Gibbons et al., 2010) or conformism (Bernheim, 1994; Fehr and Falk, 2002). Even if the people in your neighbourhood do not exhibit any influences, differences in the quality of local resources might generate the observed concentration of poverty (Durlauf, 1996).

The work presented in Chapter II argues that if living in a bad neighbourhood does indeed have negative effects on outcomes such as school results, in England these effects will be most extreme in high-density social housing neighbourhoods. In this study, I estimate the effect of moving into a very deprived neighbourhood, as identified by a high density of social housing, on the educational attainment of fourteen years old (9th grade) students in England. Neighbourhoods with markedly high concentrations of social housing have very high unemployment and extremely low qualification rates, as well as high building density, over-crowding and low house prices. In order to identify the causal impact of moving into permanent social housing in a highly deprived neighbourhood, this study exploits the timing of moving into these neighbourhoods. I argue that the timing of a move can be taken as exogenous because of long waiting lists for social housing in high-demand areas. I will argue that by focussing on the timing of the move, I can single out variation in neighbourhood quality that is not confounded by the selection problem. Thus this new strategy bypasses the usual sorting and endogeneity problems.

Using this approach, there is no evidence for otherwise negative effects, which has potentially wide-ranging implications for housing policy.

3.2 Chapter III: Evidence on the importance of peers in the neighbourhood

As we know from the sorting discussion in section 2.2, it is notoriously difficult to cleanly identify neighbourhood effects. Moreover, there is some confusion in the literature about what constitutes a neighbourhood effect. Using Manski's taxonomy, Chapter II of this thesis estimates the reduced form effect from both the 'exogenous' and the 'endogenous' effects. I do not distinguish between effects that might be caused by particular actions of neighbours, for example peer or gang pressure, and effects coming from the characteristics of those neighbours. Furthermore, it is left open what kind of neighbours might matter. I argue, however, that I carefully control for potential correlated effects, and include school fixed effects or neighbourhood fixed effects in some of the specifications.

The work presented in the third chapter further disentangles potential channels that could give rise to neighbourhood effects. In particular, we[4] address the question whether teenagers in similar age-groups affect other teenagers' outcomes. This is possible using very detailed pupil-level and school-level administrative data matched to detailed geographical information on pupils' residence and geographical mobility, as well as other administrative data sources. This combined set of information is used to analyse to effect of quality of peers in the neighbourhood - and more generally quality of the neighbourhood - on individual progress through secondary education.

To measure the quality of one student's neighbourhood, we construct several aggregate indicators based on geographical areas which group a handful of postcodes surrounding a pupil's place of residence. These include aggregate information about the educational quality of the peers in the neighbourhood - such as average attainments - as well as proxies for the overall quality of the neighbourhood - such as the incidence of unemployment. The fact that we can use various levels of geographical aggregation helps us to address the problem that there exists no clear-cut delimitation of 'the neighbourhood'.

Regarding the geographical selection (sorting) problem, different strategies are used to account for the fact that place of residence is determined by unobservable family characteristics and preferences regarding school quality and neighbourhood amenities. In the main specifications we focus on pupils who do not change their residential

[4] Since this chapter presents co-authored work I am using plural rather than singular pronouns when referring to the author(s).

locations themselves, and use variation in the neighbourhood composition that is generated by the mobility of other pupils around them. Given the detail of our data, we are able to carefully control for parental sorting and other neighbourhood-level correlated effects by individual, secondary-school-by-cohort-by-year, primary school fixed effects and neighbourhood trends. We carefully discuss this strategy regarding external and internal validity and provide a series of robustness checks to support our specification.

All in all, our results show little evidence of sizeable and significant neighbourhood effects on young people's educational attainments.

3.3 Chapter IV: Evidence on the importance of peers at school

The previous two chapters argued that neighbourhood-level variables have little influence on teenagers' school outcomes in England, controlling for school quality. However, local school quality can also be regarded as a neighbourhood amenity. Rather than focussing on other characteristics of neighbourhoods and social interactions in the neighbourhood, differences in peer composition at school could also cause differences in outcomes in later life. Chapter IV looks at schools and studies the scale and nature of ability peer effects in secondary schools in England.

The identification problems inherent to peer effects studies at schools are largely similar to neighbourhood effects research. In the context of peer effects the main difficulty lies in convincingly controlling for the sorting of students into schools and local differences in school quality, which might result in spurious correlations between peers' characteristics or ability and individual outcomes. From the methodological perspective these issues are closely linked to geographical sorting into neighbourhoods and unobserved neighbourhood infrastructure characteristics in neighbourhood effects research.

In order to shed light on the nature of peer effects in secondary schools in England, this chapter investigates which segments of the peer ability distribution drive the impact of peer quality on students' achievements. Additionally, we[5] study which quantiles of the pupil ability distribution are affected by different measures of peer quality.

[5] Since this chapter presents co-authored work I am using plural rather than singular pronouns when referring to the author(s).

To do so, we use census data for four cohorts of pupils taking their age-14 national tests in 2003/2004-2006/2007, and measure students' ability by their prior achievements at age-11. We base our identification strategy on within-pupil regressions that exploit variation in achievements across the three compulsory subjects (English, Mathematics and Science) tested both at age-14 and age-11. It is demonstrated that this pupil fixed effect approach controls for geographical sorting.

In terms of findings, this chapter documents significant and sizeable negative peer effects arising from students at the very bottom of the ability distribution. In contrast, there is little evidence that average peer quality and the highest achieving peers significantly affect pupils' academic achievements. However, these results mask some significant heterogeneity along the gender dimension, with girls benefiting significantly from the presence of very academically bright peers, and boys significantly losing out. We further provide evidence that the effect of the very best peers substantially varies by the ability of other pupils. On the other hand, the effect of the very worst peers is similarly negative and significant for boys and girls of all abilities.

Given the heterogeneity of the effect it is difficult to draw general policy conclusions but the paper offers some discussion of potential policy interventions aimed at increasing overall attainment.

3.4 Chapter V: Evidence on the importance of time zone assignment

The final research project that forms part of this thesis offers a slightly different angle on the importance of geography. The three previous chapters present evidence on neighbourhood effects and social interactions in schools and neighbourhoods making use of rich individual-level panel data in order to examine how students are affected by their environment. The last chapter of this thesis poses the question how people are affected by their environment in a rather different setting: I examine effects of the timing of daylight on electricity consumption. Electricity consumption presents an interesting case because households spent 125 billion US-$ for electricity in the US in 2005 alone (USdOT, 2010). I argue that a better understanding of the effect of timing of daylight on electricity consumption could potentially result in significant cost savings and welfare improvements.

We generally know very little about the economic effects of local time and daylight on human activity. While there is an established medical literature on the positive effects of daylight, for example van den Berg (2005), there is very little empirical evidence in other areas. For a given location local times of sunrise and sunset only vary in a very smooth pattern over the year, which makes credible empirical estimation difficult. In terms of variations in local time, the exceptions are changes due to daylight-savings time (DST)[6], and this variation has indeed been used to estimate effects on residential energy consumption (USdOT, 1975; Rock, 1997; CEC, 2001; Kandel, 2007; Kotchen and Grant, 2012; Kellogg and Wolff, 2008), coordination costs (Hamermesh et al., 2008), effects on trade and FDI (Marjit, 2007), on financial markets (Kamstra et al., 2000), and car accidents (Sood and Ghosh, 2007). Overall the results of the literature on electricity consumption are inconclusive.

In Chapter V, I argue that rather than focussing on local changes in DST regimes, nationwide geo-temporal variation in local times of sunrise can be used for estimation. To do this, I employ Geographical Information Systems (GIS) to calculate the solar times of sunrise and sunset for all geographical locations in the mainland US. Combing this information with institutional factors of time zone assignment and the daylight savings regime, I can uncover the non-standard variation in sunrise times in standard local time over space. I show that the local time of sunrise depends on the time-zone, daylight-saving time, and geographical position within zones, and identify two distinct geographical sources of variation in the timing of daylight. This variation is subsequently used to uncover county-level responses in residential electricity consumption to changes in the timing of daylight.

In terms of findings, there is no robust overall effect of sunrise times. However, using both sources of variation I find that early sunrise is associated with lower residential electricity use in the North, but higher consumption in the South. These results would suggest that additionally splitting the US into time zones horizontally could potentially generate welfare effects through substantial cost savings. I also present a first rationalisation of these new stylised facts, but acknowledge that further research is needed to examine behavioural channels in more detail.

The rest of the thesis is structured as follows: the next chapter presents evidence on social housing neighbourhoods and school performance. Chapter III looks at peer

[6] Daylight savings time is also often referred to a British Summer Time.

effects at the neighbourhood level, and Chapter IV examines the scale and nature of peer effects in secondary schools in England. Finally, Chapter V assesses how the timing of daylight, which is a function of (exogenous) geography and (endogenous) institutions, affects residential electricity consumption.

References

Akerlof, G. (1997). Social distance and social decisions. *Econometrica*, 65(5):1005–1027.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, First edition.

Bayer, P., Ross, S. L., and Topa, G. (2008). Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy*, 116(6):1150–1196.

Bernheim, B. D. (1994). A theory of conformity. *Journal of Political Economy*, 102(5):841–77.

Calvo-Armengol, A. and M.O., J. (2004). The effect of social networks on employment and inequality. *American Economic Review*, 94(3):426–454.

CEC (2001). Effects of daylight saving time on california electricity use. california energy commission (cec). Technical report.

Duranton, G. and Rodríguez-Pose, A. (2005). When economists and geographers collide, or the tale of the lions and the butterflies. *Environment and Planning A*, 37(10):1695–1705.

Durlauf, S. N. (1996). A theory of persistent income inequality. *Journal of Economic Growth*, 1(1):75–93.

Fehr, E. and Falk, A. (2002). Psychological foundations of incentives. *European Economic Review*, 46(4-5):687–724.

Gibbons, S., Silva, O., and Weinhardt, F. (2010). Do neighbours affect teenage outcomes? evidence from neighbourhood changes in england. SERC Discussion Papers 0063, Spatial Economics Research Centre, LSE.

Glaeser, E. and Scheinkman, J. (2001). Measuring social interactions. In Durlauf, S. and P., Y., editors, *Social Dynamics*. Boston, MA: MIT Press.

Granovetter, M. (1995). *Getting a Job*. Chicago, IL: University of Chicago Press.

Hamermesh, D. S., Myers, C. K., and Pocock, M. L. (2008). Cues for timing and coordination: Latitude, letterman, and longitude. *Journal of Labor Economics*, 26(2):223.

Harvey, D. (1973). *Social justice and the city*. Johns Hopkins University Press.

Johnston, R.J.; Gregory, D. T. G. and Watts, M. (2000). *Human Geagraphy*. Blackwell Publishing, (eds.) 4th edition.

Kamstra, M. J., Kramer, L. A., and Levi, M. D. (2000). Losing sleep at the market: The daylight saving anomaly. *The American Economic Review*, 90(4):1005–1011.

Kandel, A. (2007). The effect of early daylight saving time on california electricity consumption: A statistical analysis. Technical report.

Kellogg, R. and Wolff, H. (2008). Daylight time and energy: Evidence from an australian experiment. *Journal of Environmental Economics and Management*, 56(3):207–220.

Kotchen, M. J. and Grant, L. E. (2012). Does daylight saving time save energy? evidence from a natural experiment in indiana. *The Review of Economics and Statistics*.

Krugman, P. (1992). *Geography and Trade*, volume 1. The MIT Press.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3):531–42.

Marjit, S. (2007). Trade theory and the role of time zones. *International Review of Economics & Finance*, 16(2):153–160.

Overman, H. G. (2004). Can we learn anything from economic geography proper? *Journal of Economic Geography*, 4(5):501–516.

Rock, B. (1997). Impact of daylight saving time on residential energy consumption and cost. *Energy and Buildings*, 25(1):63 – 68.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:688–701.

Sen, A. (1980). Equality of what? In McMurrin, S., editor, *Tanner Lectures on Human Values*. Cambridge: Cambridge University Press.

Small, M. (2009). *Unanticipated Gains: Origins of Network Inequality in Everyday Life*. Oxford Univerity Press.

Sood, N. and Ghosh, A. (2007). The short and long run effects of daylight saving time on fatal automobile crashes. *The BE Journal of Economic Analysis & Policy*, 7(1):11.

USdOT (1975). The daylight saving time study: A report to congress from the secretary of transportation. u.s. department of transport (usdot). Technical report.

USdOT (2010). *Annual Energy Review 2009, U.S. Department of Energy (USdOT)*. Energy Information Administration.

van den Berg, A. (2005). *Health impacts of healing environments: a review of evidence for benefits of nature, daylight, fresh air, and quiet in healthcare settings*. Foundation 200 years University Hospital Groningen.

Zenou, Y. (2008). Social interactions and labour market outcomes in cities. IZA Discussion Paper 3283, IZA.

CHAPTER II

EVIDENCE ON THE IMPORTANCE OF DEPRIVED NEIGHBOURHOODS

1 Introduction[1,2]

Over 170,000 families used an emergency shelter or a transitional housing programme in the US during the 12-month period preceding September 30, 2009, 7% up from the previous year. Indeed, demand for affordable housing (or 'social' or 'public' housing) exceeds supply in many countries. In England, 4.5 million people are on social housing waiting lists. This is despite neighbourhoods with markedly high concentrations of social housing having very high unemployment and extremely low qualification rates, as well as high building density, rooms over-crowding and low house prices.

Researchers in both economics and sociology have identified a number of reasons why living in deprived areas might exhibit negative influences on its own right: peer group and role model effects could explain why our behaviour depends on others around us (Akerlof 1997; Glaeser and Scheinkman 2001). Others have pointed to the importance of social networks (Granovetter 1995; Calvó-Armengol and Jackson 2004; Bayer et al. 2008; Zenou 2008; Small 2009; Gibbons et al. 2010) or conformism (Bernheim 1994; Fehr and Falk 2002). Finally, local resources such as school quality or other environmental amenities could also induce neighbourhood effects (Durlauf 1996).

If living in a bad neighbourhood has negative effects on outcomes such as school results, in England these effects will be most extreme in high-density social housing neighbourhoods. This can have wide-ranging implications, as it could in extreme cases constitute a locking-in of the disadvantaged into a spatial poverty trap: 'once you get into a bad neighbourhood, you and your children won't get out'. As a consequence of these concerns, the relationship between place and poverty has become a key issue for

---

[11] The work presented in this section builds and improves on my MSc dissertation submitted to the Department of Geography and Environment at the London School of Economics in 2008. The title of my MSc dissertation was 'Detecting Neighbourhood Effects: An empirical investigation of Social Housing and School Performance in England'. Compared to my MSc thesis, the analysis presented here uses data spanning twice as many students, as well as additional data on house prices, and a different and improved specification. The analysis presented here, as well as all Stata-code and the full write-up, was created during the PhD-research stage.

[2] A previous working-paper version of this chapter is currently published as SERC discussion paper Weinhardt (2010).

policy makers worldwide and poses interesting questions regarding where to construct new affordable (or 'permanent supportive') housing, which is one of the key policies of the US Department for Housing and Urban Development. In England, the debate has focused on constructing new social housing in 'mixed communities' (e.g. Cheshire et al. 2008) to avoid potential negative effects.

However, the existing evidence on negative neighbourhood effects is inconclusive. The majority of empirical studies do not find evidence for negative neighbourhood effects on school outcomes using a variety of research strategies to overcome selection, reflection and other problems with unobserved correlated effects. Most American studies that do not find evidence for neighbourhood effects on school outcomes define neighbourhoods using very large US Census blocks. Arguing that a much finer level of spatial aggregation is needed to uncover highly localised effects, Goux and Maurin (2007) document large negative effects of close neighbours on school outcomes in France.

The aim of this paper is to establish whether moving into localised high-density social housing neighbourhoods causes a deterioration in the school attainment of 14-year-old pupils. The English setting offers a unique opportunity to answer this research question for a number of reasons. Firstly, the data makes it possible to identify highly localised neighbourhoods using the Census 2001 'Output Areas' (OAs), which are small-scale and contain only 125 households on average. This avoids potential bias from using aggregated data, although the sign of the bias from over-aggregation cannot be theoretically determined. Second, in order to identify the causal impact of neighbourhood deprivation on pupil attainment, this study exploits the timing of moving into these neighbourhoods. In England, the timing of a move can be taken as exogenous because of long waiting lists for social housing in high-demand areas. Naturally, a pupil's result in the age-14 Key Stage 3 (KS3) test, a centralized nationwide assessment of school attainment, can only be influenced by the low quality of her new neighbourhood if she moves into this neighbourhood before taking the test. Finally, nationwide census data makes it possible to track individual residential mobility for two cohorts of school children; the study is therefore not limited to a small number of neighbourhoods or of cities. This setting hence allows us to compare the school results of pupils who move into deprived social housing neighbourhoods before, versus after, taking the national KS3 test. Controls for a potential direct effect of moving, earlier attainments, family background and school quality are also included.

Importantly, this strategy controls for segregation that is purely an outcome of sorting. It is a well-established fact that housing markets lead to spatial income segregation (Kain and Quigley 1972; Black 1999; Gibbons et al. 2009). Parental income also correlates with school attainments (e.g. Taubman 1989); hence we expect to find that the weakest pupils live in the worst neighbourhoods purely based on the sorting mechanism. As a result, individual characteristics may correlate with neighbourhood characteristics, which confounds causal interpretation (Manski 1993; Moffitt 2001). This problem is addressed directly by focusing on the exogenous timing that results from social housing waiting lists. As a result, I can include neighbourhood fixed effects to relax the usual assumption that social housing neighbourhood allocation is quasi-random (e.g. Oreopoulus 2003).

The main finding of this study is that early movers into deprived social housing neighbourhoods experience no negative effects on their school attainment relative to late movers. While it is demonstrated that there are large associations between moving into deprived areas and school outcomes, these correlations cease to exist once controlling for group-specific characteristics in a difference-in-difference framework. The finding of no negative effects of low social housing neighbourhood quality on school attainment raises interesting questions regarding the current housing policy of mixed communities in England, and affordable housing policy worldwide.

The rest of the paper is structured as follows: The next section explains in detail the institutional background and identification strategy used by this study. Section 3 describes the data; section 4 discusses the results; and section 5 presents a battery of robustness checks before I conclude.

2 Empirical strategy

2.1 Institutional background: the English school system

The English school system is organized into four key stages, in which learning progress is assessed at the national level. Of interest for this study is the Key-Stage 2 (KS2) assessment at the end of primary/junior school, and the Key-Stage 3 (KS3) assessment, which assesses pupils' progress in the first three years of compulsory secondary education (figure 1). The KS2 assessment is at the age of 10/11, while the KS3 is carried out at the age of 13/14. I use the average performance across the three core subjects, English, Mathematics and Science, to measure attainment. Since I compute cohort-specific per centiles of the respective KS2 and KS3 scores, individual results between the two tests and cohorts are directly comparable. The KS3 score is of no direct importance to parents or housing organisations and is not a high-stakes test in a sense that anyone would specifically avoid moving before the test or time a move around it. On the other hand, it correlates highly with later school and labour market outcomes and is therefore of general policy interest.

2.2 Institutional background: Social Housing

The quality and social composition of social tenants has changed greatly over the past 60 years. After the Second World War, when Britain, like most other European countries, faced an acute housing shortage, social housing provided above-average quality accommodation. A move into social housing was regarded as moving up from private renting and most houses had gardens and good amenities (Lupton et al. 2009). The social housing sector continued to expand during the 1960s and 1970s and peaked at 31 per cent of the total English housing stock in 1979 (Hills 2007, p. 43). Social housing still provided much diversity in terms of both, quality and social and neighbourhood composition but some of the older stock required refurbishments. As a response to this, housing associations, non-profit entities that provide social housing, started to grow in number and importance (Lupton et al. 2007).

From the 1980s until today the social sector shrank both in absolute size and importance relative to other types of tenure. Construction activity in the social sector declined sharply from almost 150,000 dwellings to 50,000 dwellings/year in the early 1980s and stagnate on the historically lowest level since the Second World War at around 20,000/year since the late 1990s (Hills 2007). In 2004, councils and housing associations provided about 4 million social dwellings (18.5 per cent of stock), down from almost 6 million dwellings in 1979. This decline of social housing resulted from a combination of the 'right-to-buy' scheme introduced by Margaret Thatcher in 1980 and public spending cuts on new construction (Hills 2007, p. 125). The 'right-to-buy' scheme also altered the socioeconomic composition of social tenancy as it allowed those who could afford it move into owner-occupation (Hills 2007; Lupton et al. 2009). Admission criteria also changed during this period when the Homeless Persons Act of 1977 forced councils to provide accommodation to certain groups in extreme need (Holmans 2005). These trends continued through the 1980s and 1990s, and since 1991 growing demand has confronted a negative net supply of absolute numbers of social rented dwellings (Hills 2007). As a result of these changes and the increasingly needs-based allocation, in 2004 70 per cent of social tenants belonged to the poorest two-fifths of the income distribution and hardly anyone to the richest fifth. This is in contrast to 1979 when 20 per cent of the richest decile lived in social housing (Hills 2007, pp. 45, 86).

Today, demand for social housing grossly exceeds supply. Currently, nine million social renters live in four million social dwellings (Turley 2009). With negative net changes in social housing supply, spaces can only free up if existing tenants die or move out. Yet movement within or out of the sector is very low and 80 per cent of social tenants in 2007 were already there in 1998, if born (Hills 2007, p. 54). As a result, there are currently 4.5 million people (or about 1.8m households) on waiting lists for social housing. Taking these numbers at face value, if nothing were to change and no one were born into social housing, this would mean that about 800,000 dwellings (20 per cent of 4m) could free up every 10 years. Even assuming zero new demand over the coming years, it would take over 22 years to provide housing to all of those who are currently on a waiting list in this scenario.

The social housing allocation system as it exists today continues to operate on a needs-based system where the Homelessness Act 2002 defines beneficiaries. Importantly, families with children are treated as a priority. In the current situation of excess demand it is in fact very difficult to get into social housing without belonging to one of the needy groups. While the needy groups are defined nationally, provision is decentralised and administered through councils or housing associations. Local authorities operate different systems, some using a banding system and others a points-based system to ensure that those with the highest need and waiting time get a permanent place in social housing next (Hills 2007).

About a third of local authorities complement their waiting list system with a choice-based element, where new social housing places are announced publicly and prospective tenants are asked to show their interest in each specific place (Hills 2007, p. 163). The prospective tenant with the highest score as determined through the waiting list mechanisms then gets the offer. However, most places are still directly allocated through the council or housing association. Regan (et al. 2001) writes that one of their interviewees in Reading who rents from a social landlord complained: "Most of the people I know who have been offered flats or houses or anything have no choice… it is that or nothing" (2001, p.22). As I will argue later, it is not central to our identification that people cannot exert influence on the neighbourhood or place where they are offered social housing.

As already mentioned, mobility within the social housing sector is extremely low. Reagan (ibid., executive summary, no page numbers) concludes in a qualitative study on housing choice and affordability in Reading and Darlington that "Moving within social housing was curtailed by allocation procedures and a lack of opportunity to move or swap properties". Quantitative evidence confirms that mobility within the social rented sector is extremely low, in spite of the mobility schemes that the government started to implement in the recent years (Hills 2007, p. 109). It is still the exception to move within the social housing sector once one gets in.

Finally, there is a widespread perception that immigrants receive priority in social housing allocation. If this were true, changes in migration flows could confound my analysis. However, this is not the case because immigrants are generally ineligible for social housing, as pointed out by Rutter and Latorre (2009).

2.3 Institutional background: Housing Benefits

In England, parents on low incomes or who are unemployed can claim housing benefit, which essentially covers part or up to 100 per cent of their payable rent. The current eligibility rules were set in 1988, which is prior to the period of this study (Hills 2007, p. 115). Importantly, housing benefit is awarded independently of tenure status and is equally available to parents living in the private rented or social housing sector or even in temporary private accommodation. The exact amount of housing benefit paid depends on a number of factors including the number of children, income and savings but also on the 'local reference rent', which is determined by local housing officials and effectively sets a maximum for what constitutes a 'reasonable' rent in the private sector. Depending on these circumstances housing benefits can cover the full rent. For central London, for example, the corresponding rent for a 2-room flat (i.e. one bedroom, one living room) was 290 pounds per week in December 2005 (Hills 2007, p.116). Importantly, housing benefits are extremely responsive to residential changes or changes in rent; hence families who get into social housing where rents are 50-60 per cent lower than in the private rented sector will face an immediate reduction in housing benefits. This institutional setting gives rise to a unique situation where we do not expect any direct income effects from moving into social housing. I will return to the question of potential income effects from moving into social housing in the robustness section, where I show that there is no significant association between free school meal status, an indicator for low income, and the timing of moving. The social housing sector is not attractive because of any income effects but because it offers stability that cannot be offered by private landlords (Hills 2007, p.18). The fear of 'Rachmanism', which is exploitation of tenants by unscrupulous landlords, is another factor that raises the attractiveness of the social rental sector relative to the private sector (Hills 2007, p.18).

2.4 The general identification strategy

Figure 1 illustrates the identification strategy in the context of the English school system. The time when the Key Stage 3 test is taken is denoted with *t*, the time for the Key Stage 2 with *t-1*. Hence, *t-1* to *t* spans the academic years 7 to 9, between the KS2 and KS3 tests. Conversely, *t* and *t+1* cover the years 10 and 11 after the KS3. I compare test scores of pupils who move into deprived social housing neighbourhoods before taking the KS3 test, in the period from *t-1* to *t*, to pupils who also move into deprived social housing neighbourhoods, but in the period between *t* and *t+1*. Formally, this reads:

$$\begin{aligned} \text{test score}_{i,n,t} = \alpha_i \; & + \gamma_1 \, d(\text{SH-Move})_{i,t,t-1} \\ & + \gamma_2 \, d(\text{SH-Move})_{i,t-1,t+1} \\ & + \gamma_3 \, d(\text{Move})_{i,t,t-1} \\ & + \gamma_4 \, d(\text{Move})_{i,t-1,t+1} \\ & + \mathbf{D}_{n,t+1}{}' \, \boldsymbol{\gamma}_5 \\ & + \gamma_6 \, \text{test score}_{i,t-1} \\ & + \text{further controls}_{i,t} \\ & + \phi \, d(\text{cohort})_i + \varepsilon_{i,n,t} \end{aligned} \tag{1}$$

The dependent variable is the age-14 KS3 test score, which is the national test taken at academic year *t*. The first four dichotomous dummies constitute a difference-in-difference setup where γ_1 is the coefficient of main interest. The first group consists of all pupils who do not move at all during the observed period. This group is only included in the constant a_i and control variables in order to gain precision. The second group consists of pupils who move once, denoted by the fourth dummy $d(Move)_{t-1,t+1}$. The third group, which is a sub-group of the second, consists of pupils who move into social housing neighbourhoods $d(SH\text{-}Move)_{t-1,t+1}$. Some pupils who move once, also move before the KS3 test at time *t*, and they are captured by $d(Move)_{t-1,t}$. Finally, of those pupils who move once and before the test, some move into social housing neighbourhoods, identified by $d(SH\text{-}Move)_{t-1,t}$. Hence an estimate of γ_1 gives the association between moving once and into a social housing neighbourhood before the test and the Key Stage 3 test result, controlling for moving once γ_4, moving before the test γ_3, and other effects that potentially correlated with moving into social housing at some point γ_2.

As further variables, the KS2 test score is included to proxy pre-treatment ability, denoted by $test\ score_{i\ t-1}$. $\boldsymbol{D}_{n\ t+1}$ is a matrix of neighbourhood-dummies that captures all unobservable constant neighbourhood characteristics and *further controls*$_{i,t}$ include information on parental income, proxied by free school meal eligibility, ethnicity, gender and a variable on the pupil teacher ratio at school. Finally, *d(cohort)*$_i$ allows for different intercept values for the two cohorts of students that I can follow over this period.

As a results of focusing on pupils who move into social housing neighbourhoods at different times, the hope is to single out variation in neighbourhood quality that is exogenous, i.e. independent of prior characteristics. This strategy exploits the fact that people who apply for social housing in England are not directly allocated a place but have to remain on waiting lists for quite a while. Since identification relies on long waiting lists, as additional safety net I only include in the analysis local authorities in which at least 5 per cent of the population have been on a waiting list in the year 2007.[3] Crucially, this should ensure that families who get into social housing at different points in time are on average very similar in their characteristics. That is, the timing of the move, but neither the decision to move itself nor the wish to get into social housing, should be exogenous in high demand areas. In these areas, pupils with parents who apply to social housing at different times should share similar observable and unobservable characteristics but have different 'exposure' times to a social housing neighbourhood, as generated through the precise timing of when they are offered a place. As I show later, our data makes it possible to support this identifying assumption directly regarding observable characteristics. Technically, if this assumption is met, this ensures that γ_2 in specification (1) captures constant 'correlated effects' that could otherwise confound causal interpretation of the estimate for γ_1.

In the following I discuss the assumptions that need to hold for identification in more detail. Importantly, I argue that identification is achieved even if there is discrimination or an institutional preference for certain types of families. I can still obtain causal estimates if families or children with certain unobserved characteristics are favoured in the social housing allocation process. This is because in expected outcomes the setting can be represented as follows:

[3] There is very little variation over time in this indicator.

$$E\{\text{test scores}_i \mid \text{SH-Move}_{i,t-1,t}=1\} - E\{\text{test scores}_i \mid \text{SH-Move}_{i,t-1,t}=0\} \quad (2)$$
$$= E\{\text{test scores}_{1i} - \text{test scores}_{0i} \mid \text{SH-Move}_{i,t-1,t}=1\}$$
$$+ [E\{\text{test scores}_{0i} \mid \text{SH-Move}_{i,t-1,t}=1\} - E\{\text{test scores}_{0i} \mid \text{SH-Move}_{i,t-1,t}=0\}]$$

The first line defines the difference in the expectation of test score results for the same individual *i* who either moves into social housing between *t-1* (KS2) and *t* (KS2), denoted by $SH\text{-}Move_{i,t-1,t}=1$, or does not, i.e. conditional on $SH\text{-}Move_{i,t-1,t}=0$. Of course, both of these outcomes can never be observed simultaneously for the same individual. But as we know from the introduction, these terms can be rearranged into an effect of Treatment on the Treated (ToT) and Selection Bias. Here, the term in the second row represents the ToT, and the term in the square brackets sorting into treatment, which is the selection bias. One concern is that this expression $[E\{test\ scores_{0i} \mid SH\text{-}Move_{i,t-1,t}=1\} - E\{test\ scores_{0i} \mid SH\text{-}Move_{i,t-1,t}=0\}]$ does not equal zero. It represents the difference between test scores of pupils who do not move into social housing, compared to the counterfactual of what pupils who move into a social housing neighbourhood would have obtained, had they not moved.

Indeed, there are many reasons to believe that this term does not equal zero. This is because pupils who move into social housing have parents who are eligible for social housing, and who thus have lower incomes and further characteristics that are likely to correlate to a pupil's test score at school. As a result, we cannot interpret the observed differences in test scores between pupils who move into social housing neighbourhoods, compared to pupils who did not move, as a causal effect of the neighbourhood. This is an example of sorting, the geographical counterpart to the more general selection problem.

The identification assumption of this study is that the timing of a move is independent of individual characteristics conditional on moving into a social housing neighbourhood in a high demand area at some point. Formally, where $test\ scores_{1i}$ and $test\ scores_{0i}$ denote the two potential outcomes for individual *i*, this assumption can be written out formally thus:

$$\{\text{test scores}_{1i}, \text{test scores}_{0i}\} \quad \text{SH-Move}_{i,t-1,t} \mid \text{SH-Move}_{i,t-1,t+1}=1 \qquad (3)$$

where the LHS denotes test scores under the two potential outcomes for individual *i.* We are concerned that the occurrence of the potential outcomes is not independent of moving into social housing. The identification assumption taken here is that moving (or not moving) into social housing before the KS3 test $SH\text{-}Move_{i,t-1,t}$ is not related to the potential outcomes conditional on moving into social housing at some point either before or after the KS3 test, denoted by $SH\text{-}Move_{i,t-1,t+1}=1$. This means that conditional on moving into social housing at some point, the timing of the move is not related to observable and unobservable characteristics and this variation can be used for estimation. If this holds, we get:

$$\begin{aligned} &E\{\text{test scores}_{i} \mid \text{SH-Move}_{i,t-1,t}=1\} - E\{\text{test scores}_{i} \mid \text{SH-Move}_{i,t-1,t}=0\} \qquad (4) \\ &\quad = E\{\text{test scores}_{1i} - \text{test scores}_{0i} \mid \text{SH-Move}_{i,t-1,t+1}=1\} \end{aligned}$$

The selection term disappears since it equals zero conditional on $SH\text{-}Move_{i,t-1,t+1}=1$. This is because there are no differences between early and late movers conditional on moving at some point. Therefore the observed difference in test scores between early and late movers into social housing can be interpreted as the treatment–on-the treated-effect, as long as the independence assumption in equation (3) holds.

Importantly, this setting does not rule out the existence of any institutional factor, discrimination against certain types of applicants or selection that is constant over time. By the same logic, if a social planner always offers places in nicer neighbourhoods to families with certain characteristics, this is going to happen equally before and after the KS3 test. As long as these factors remain unchanged over the study period, they are not correlated with the timing, conditional on moving at some point, just as spelled out by equation (3). Therefore, γ_1 in specification (1) would still uncover the effect of the treatment on the treated, conditional on moving into a social housing neighbourhood at some point. Note that this is a relaxation of the assumption that discrimination or institutional preferences for certain types of families do not exist at all. Here it is only required that these factors do not change over the time of the study period.

A second reason why constant unobservable factors do not cause any bias is because I can include neighbourhood fixed effects in the specification. This means that I will effectively compare pupils who move into the same neighbourhood at different points in time. Any constant unobservable characteristic that is then related to neighbourhood quality will be captured by the fixed effect.

To summarise, the identifying assumption is that the average characteristics of pupils whose parents move into highly concentrated social housing neighbourhoods do not change over the study period. If this assumption is met, identification is not obscured by individual or any constant unobservable factors such as sorting preferences or institutional discrimination that influence both neighbourhood quality and school results.

A remaining threat to this identification strategy is the possibility that a negative shock which made a family eligible for social housing, and which occurred several years before the observed move, may affect the test scores of early and later movers differently later in life. We will return to this issue in section 5, where I present a quasi-test of the identification assumption of the independence of the timing of moving. As I will show, I fail to detect any differences in student characteristics or detailed data on prior test scores between pupils who move into social housing neighbourhoods before or after the KS3 test, which supports my approach.

3 Data

3.1 The Pupil Level Annual School Census

The Department for Children, Schools and Families (DCSF) has collected pupil-level census information from all state schools in England since 1996. From the 2001/02 cohort onwards, detailed pupil-level information such as ethnic background, free school meals eligibility (FSME), and pupils' postcode of residence is collected in the pupil-level annual school census (PLASC). People eligible for FSME are likely to receive Income Benefits, Job-seekers' Allowance and to be single parents with a dependent child (Hobbs *et al.* 2007). This variable serves as proxy for the lowest income groups. Overall, given the extent of the census data, I can construct a pupil-level panel of two cohorts for five consecutive years and track individual pupils from their first (year 7) to fifth (year 11) year in secondary education. For the first cohort this corresponds to the period from 2001/02 to 2005/06, and for the second from 2002/03 to 2006/07.

The PLASC is collected in the middle of each January, close to when the Key Stage 3 tests are taken in May. I ignore this time mismatch of four months here, but address it directly in one of the robustness checks. I can use the residential information on the Census 2001 Output Area (OA) level to identify all pupils who have moved during the academic years 8, 9, 10 or 11. OAs were originally constructed to include a comparable number of households: each contains about 4 to 5 postcodes and on average 125 households. I use the OA to define what I understand as a 'neighbourhood'.

Unfortunately, the PLASC does not contain any information on housing tenure. Hence the next and crucial step is to identify who lives in a social housing neighbourhood and who does not. I do this using neighbourhood information from the 2001 Census of Population. The 2001 Census of Population is the most recent available decennial survey of all people and households living in England and Wales. A wide range of socio-economic variables was collected and made available at various levels of spatial aggregation. This census was collected one year before my analysis starts and I extract pre-treatment neighbourhood-level information on the total number of households that rent from the council (local authority) or a registered social landlord or housing association, the male unemployment rate, the level of education, the level of car ownership, building density, overcrowding, average number of rooms per household and the percentage of lone parents with dependent children. The first two variables are used to calculate the percentage of households living in social housing for each OA. There has been very little change in the stock of social housing since 2001, and mobility is limited, as discussed in section 2.2. As a result, it is unlikely that these neighbourhoods have changed dramatically since the 2001 Census (Hills 2007, pp.

169ff). Therefore I can use that census to identify high-density social housing neighbourhoods for the entire study period.[4]

Following our identification strategy, the timing of movers into 100-per cent social housing neighbourhoods must be exogenous, whereas movers into zero-per cent social housing neighbourhoods, at the other extreme, are never constrained by social housing waiting lists. However, only a very few OAs are completely social housing. This is why I am forced to use a lower threshold of 80 per cent. If 80 per cent of all households in a particular OA live in social housing, then it is still very likely that a pupil who lives in that OA also lives in social housing. Therefore, everyone living in an OA with 80 per cent or more households being in social housing is treated as living in a social housing neighbourhood, and all others are not. Using this threshold, by tracking OA changes over the years it is now possible to identify those who move out of an area with less than 80 per cent of social tenants into an area with 80 per cent or more. As I already know, mobility within the social housing sector is close to zero. Hence to identify pupils who move into social housing I focus the analysis on those who move into an OA with more than 80 per cent of households in social housing and stay there. From now on this will be referred to as 'moving into a social housing neighbourhood'.

Finally, the analysis is restricted to comprehensive, grammar, secondary modern and technical schools that span the whole period between KS2 and two years after the KS3. Other less common school types such as middle schools are not organised around the Key Stages the same way and often require school changes after year 9, which could confound any analysis that focuses on moves between years 7 and 11. The schools included cover 90 per cent of pupils in English state education.[5]

In my final dataset, 2,094 pupils move into such social housing neighbourhoods between their 7th and 11th academic year. 703 pupils move into social housing from year 7 to 8, 516 from year 8 to 9, 433 from year 9 to 10 and 442 between the academic years 10 and 11. Numbers are slightly higher for the earlier years, but this merely

[4] Even if annual information was available I would prefer to use the pre-dated 2001 Census information because later changes in neighbourhood quality could be endogenous to variation that I am using for estimation.

[5] Note also that there is a small fraction of pupils who move more than once during the study period. These students are not representative of 'stayers' and are not included in the main analysis.

reflects the general decline in mobility and is not social housing neighbourhood specific.

3.2 Descriptive statistics

Table 1 contains summary statistics for the main dataset. The first two column pairs give information for pupils who either live in a social housing neighbourhood throughout their academic years 7 to 11 (columns 1), and for pupils who move into social housing neighbourhoods during this period (columns 2). Column pairs (3) and (4) are for pupils who stay in a non-social housing neighbourhood, and who move between non-social housing neighbourhoods respectively. The table is further split into three panels, where panel A shows descriptive statistics on pupil characteristics, panel B on neighbourhood characteristics, and panel C on school characteristics.

Column (1) shows descriptive statistics for the about 10,000 pupils who live in a high-density social housing neighbourhood during the whole period. We can see from panel A that these pupils have Key Stage test scores much below the national average, which is about 50. Their KS2 scores average at only 38.64 points and the respective KS3 scores are even lower at 35.63 percentile points. These pupils are the weakest when starting secondary school, but results deteriorate even further up to KS3. Moreover, about half of them are eligible for free school meals (FSME), which is a proxy for a low-income background.

Still focussing on column pair (1), it further becomes evident from panel B that the neighbourhoods where these pupils live are characterised by a very high average unemployment rate of almost 12 per cent, low qualification levels, room overcrowding, high building densities and low property prices. Only half of the households have access to a car or van, about one fifth of the household heads are lone parents with at least one dependent child, and 43 per cent have at least one household member with a limiting long-term illness.

To summarise, pupils who live in social housing neighbourhoods throughout the entire study period underperform at school, and their neighbourhoods are characterised by indicators of high deprivation.

Next, column (2) shows statistics for the 2,097 pupils who move into a social housing neighbourhood during the study period. Panel A shows that they have individual characteristics very similar to pupils who live in a social housing neighbourhood throughout. Their KS2 and KS3 test scores also average far below the national mean: at 37.258 and 33.332 respectively, they are even slightly lower compared to the 'stayers'. The only remarkable difference is in the share who change secondary school: about ten per cent of the 'movers' change school, compared to only 4.3 per cent of the 'stayers'. This suggests that controlling for school level characteristics will be important. As discussed, one general problem in neighbourhood research is that neighbourhoods do not change much over time. As a result I have to rely on movers to identify the effect. It is hence comforting to see that 'movers' are generally similar to 'stayers' with respect to their observable characteristics. This is important for the external validity of this study.

Panel B shows the respective neighbourhood characteristics for these pupils. Note that these are the characteristics of the neighbourhoods those pupils move out of, since I summarise area characteristics before the relocation. Pupils who move into social housing hence move out of the neighbourhoods described in column (2), and into social housing neighbourhoods described in column (1). We can see that the non-social-housing neighbourhoods are significantly better than those of the social-housing-neighbourhood stayers, something that we will examine in detail in table 2 below.

Columns (3) and (4) give summary statistics for pupils who lived in non-social housing neighbourhoods throughout, or move between non-social housing neighbourhoods respectively. Panel A shows that individual Key Stage scores are much higher compared to the social housing groups. Note, however, that movers (columns 4) have slightly lower scores than 'stayers' in non-social housing neighbourhoods (47.064 compared to 51.317), but still much higher than the social housing groups (around 35). Also, only about 14 to 20 per cent (compared to almost 50 per cent) of pupils in these non-social housing neighbourhood groups are eligible for free school meals.

Secondly, panel B shows that non-social housing neighbourhoods are much 'nicer' places to live, with unemployment rates around 5 per cent, high qualification levels, lower shares of lone parents with dependent children, about 10 percentage points lower shares of residents with limiting long term illnesses, lower levels of overcrowding, larger homes, lower populations densities and higher house prices.

Finally, comparing panel C across columns and tables, it turns out that teacher-to-pupil ratios do not differ much for the various groups of pupils.

To summarise, there are small differences between the 'mover' and 'stayer' groups, but it is evident that pupils who live in or move into social housing neighbourhoods underperform in their KS2 and KS3 national tests. Furthermore, these areas present some of the most deprived neighbourhoods in the UK.

As noted above, pupils who move into social housing neighbourhoods experience deterioration in their neighbourhood quality. Table 2 looks explicitly at the neighbourhood-level changes that the 2,094 pupils who move into social housing neighbourhoods experienced. The neighbourhoods they move into are described in column (1) and column (2) gives the percentage change in neighbourhood quality for each indicator compared to the neighbourhood these pupils move out of. The first row of table 2 shows that unemployment rates are 50% higher in the new social housing neighbourhoods. In fact, we can see that neighbourhood quality deteriorates in all characteristics for pupils who move into a social housing neighbourhood. Pupils who move into a social housing neighbourhood move into a neighbourhood with a 54 per cent higher unemployment level, 14 per cent lower qualification levels, 23 per cent lower access to a car or van, and 56% more lone parents with dependent children. Furthermore, their new neighbourhoods have 14 per cent more inhabitants with limiting long-term illnesses, a 28 per cent higher overcrowding index, 10 per cent fewer rooms in the average household, 29 per cent higher population density, and 23 per cent lower house prices. The third column of table 2 expresses these changes in terms of standard deviations. Overall, the changes experienced by social housing movers are substantial; they vary between a 0.3 to over 1.1 standard deviations of the underlying variables. Note that what this study identifies is the aggregate effect on school results that arises from this general deterioration in neighbourhood quality. To summarise, table 2 shows that pupils who move into social housing neighbourhoods experience significant deteriorations in their overall neighbourhood quality. The next section presents the main results.

4 Results

4.1 'Traditional' approach

Before I turn to the main results, it is useful to inform the discussion with some benchmark regressions. These regressions are for comparative purpose only and do not focus on identification: they simply correlate KS3 results with the areas where the pupils live or move to. Table 3 shows the results from these regressions and is organized into three panels with three regressions each, where additional controls and school fixed effects are added subsequently in column (1) to (3), (4) to (6) and (7) to (9). Panel A shows estimates for the effect on KS3 scores of living in a social housing neighbourhood at the start of secondary education (year 7). In panel B the effect is estimated for pupils who move into social housing neighbourhoods before the test in year 8 and 9, and panel C shows estimates for pupils who move into social housing neighbourhood before or after the test.

For example, the regression estimated in the last column, panel B, is the following:

$$\begin{aligned} \text{Test Score}_{i,t} \quad = \quad & + \theta_1 \, d(\text{SH-Move})_{i,t-1,t} \\ & + \theta_2 \, d(\text{Move})_{i,t-1,t} \\ & + \mathbf{S}(\text{school})_{i,t-1}\,'\boldsymbol{\theta}_3 \\ & + \theta_4 \, \text{Test Score}_{i,t-1} \\ & + \text{ further controls}_{i,t} \\ & + \phi \, d(\text{cohort})_i + \phi_{i,t} \end{aligned} \qquad (5)$$

The dependent variable is the Key Stage 3 result for individual *i*. The first dummy $d(SH\text{-}Move)_{i,t-1,t}$ equals one for all pupils who move into a social housing neighbourhood before the KS3 test, that is between *t-1* and *t* which is indicated by the subscripts. The second dummy variable $d(Move)_{i,t-1,t}$ controls for the direct effect of moving and equals one for everyone who moved before the KS3, independent of whether the move was into social housing neighbourhoods or between non-social housing neighbourhoods. The third term *S(school)* is a matrix of dummies for each individual school in year 7 and θ_4 is an estimate for the effect of previous attainment (KS2 score). Variables for school changes, FSME eligibility, ethnicity and gender are included are included in *further controls$_{i,t}$*. Finally, *d(cohort)* allows for a different intercept for the two cohorts that I can follow. The other specifications in table 3 only differ in terms of the first two variables that are included here, as specified by the panel headings.

The key difference of this cross-sectional setup compared to the difference-in-difference specification (1) discussed in section 2 is that, essentially, KS3 results of pupils who move into social housing neighbourhoods (or live in these neighbourhoods as for panel A) are compared to other pupils, conditional on a bunch of 'standard' controls. The regressions in table 3 do control for a potential direct effect from moving in itself with the *d(Move)*-variables, but this is not social-housing specific. Conversely, in the difference-in-difference framework we additionally control for 'moving into social housing' and use the timing of the move to estimate our main coefficient of interest. This is not the case for the cross-sectional results presented in table 3.

Turning to the estimates, panel A column (1) in table 3 shows the associations between living in a social housing neighbourhood at the beginning of secondary education and KS3 scores. Without further controls, the estimate in the first row shows that pupils who lived in social housing neighbourhoods in year 7 score 14.84 percentile points lower than their peers. This is an extremely strong association; it is hence not surprising that educational underperformance has been linked to neighbourhood quality in the past. However, this association between place and test score reduces to about 2.9 percentile points once a rich set of controls including prior KS2 results are added (column 2). With school fixed effects, this association reduces further to 1.54 points, while remaining significant at the one-percentage level (column 3). Note that variables such as the number of years of free school meal eligibility – an income proxy - are more important in determining school improvements.

The results are similar in size and significance to panel B, which shows estimates for specification (5) discussed above. Here, the effects are estimated for pupils who move into a social housing neighbourhood between the tests, hence for 'SH-movers' rather than for 'SH-stayers'. The unconditional association is now -13.251 percentile points (column 4) and it again reduces substantially, to 2.772 percentile points, once additional controls (column 5) and to 1.454 once school fixed effects (column 6) are added. These estimates are quite similar to panel A. If anything, the associations between moving into a social housing neighbourhood and the test results are somewhat weaker compared to those who lived in social housing in year 7.

Summarising the results from panels A and B: we see large and negative associations between neighbourhood quality and school results. These associations reduce to about one and a half percentile points once controls for a rich set of background characteristics including previous test scores and school fixed effects are included. However, these neighbourhood effect estimates are purely cross-sectional comparisons. As discussed earlier, unobserved correlated effects potentially bias these results. Therefore these results cannot be interpreted as causal effects.

Panel C of table 3 presents a first quasi-test for the importance of sorting. Here I estimate the neighbourhood effect for pupils who move into social housing before or after the KS3 test. If the previous negative associations were causal estimates for the true effect of the social housing neighbourhoods, then the estimates in panel C should be much smaller than the previous ones. This is because a substantial share of pupils who move into a social housing neighbourhood before or after the test did of course only move after the KS3 test was taken, and hence only received a 'placebo' treatment. This is because for those pupils, the new neighbourhood cannot exhibit any negative influences on educational attainment by definition. The fact that the estimates in panel C are very similar to the previous ones might hence suggest that it is not the neighbourhood, but unobserved correlated effects that cause the negative findings. The next section looks at this specifically using the difference-in-difference framework.

4.2 Main results: early and later movers into social housing neighbourhoods

Table 4 is divided into two panels horizontally and shows the main results. The upper part shows descriptive statistics (means) for groups moving before KS3/after KS3 and into social housing/non-social housing neighbourhoods. Pupils who move into a social housing neighbourhood before the test have average KS3 scores of 33.598, pupils who moved during the two years after the test score on average 32.962 (column 1). The corresponding figures for non-social-housing neighbourhood movers are 46.849 and 45.847, as shown in column (2). In column (3) the first differences are shown for pupils either moving before or after the KS3 test. Pupils who move into social housing before the KS3 score 13.251 points worse than pupils who move between non-social-housing neighbourhoods. Note that this simple difference in means is equivalent to the unconditional OLS-estimate presented in table 3 column (4). In the last column in panel A of table 4 I difference the first differences again, which results in the difference-in-differences of -0.364 KS3 points for pupils moving into social housing before versus after the test. This is equivalent to the unconditional difference-in-difference OLS estimate shown in the first column of panel B in table 4.

Turning to panel B, column (1) shows this unconditional estimate only controlling for a potential direct effect of moving, column (2) additionally includes previous test scores, ethnicity, school characteristics and gender, and in column (3) school fixed effects are added to the specification. Finally, in column (4) school fixed effects are replaced with neighbourhood fixed effects. This is the specification (1) discussed in section 2.4.

The first row shows estimates for moving into a social housing neighbourhood before the test γ_1, which are now non-significant in all specifications, and even positive in most cases. The simple mean-difference-in-difference of -0.364 in columns (1) is not significantly different from zero. Adding controls, this causal estimate of moving into social housing before the KS3 test even turns positive in columns (2) to (4), and is estimated at 0.426, 0.539 and 0.267 respectively. However, none of these estimates is significantly different from zero at conventional levels. This result is in contrast to the cross-sectional estimates presented table 3. Importantly, it is not driven by increases in the standard errors but by actual changes in the absolute sizes of the estimates.[6] This means that although pupils who move into a social housing neighbourhood before the

[6] I cluster standard errors at the neighbourhood level. Using robust standard errors instead does not alter any of the conclusions.

KS3 test underachieved, they did not underachieve to any different degree compared to their peers who move into a similar neighbourhood after the KS3 test.

This becomes directly evident when we compare the 'traditional' estimates from table 3 with table 4. For example, column (4) from table 3 gives a negative association of 13.251 percentile points for early SH-movers. In table 4, this association is now fully captured by the dummy variable that controls for moving into social housing before or after the test, which is estimated at -12.886 in the second row, panel B, column (1). This strongly suggests that the previous negative associations between moving into social housing neighbourhoods are driven by unobservable characteristics common among all pupils who move into social housing neighbourhoods at some point, and not at all by exposure to social housing neighbourhoods.

These conclusions are further substantiated in column (4), which includes neighbourhood destination fixed effects. Here, the estimate in the first row shows the difference in KS3 results for pupils who moved into the same social housing neighbourhood before or after the test. Again, there is no evidence for detrimental effects on test scores. This is an important finding because the neighbourhood fixed effect absorbs any constant selection of groups or individuals into specific social housing neighbourhoods, as well as for potential institutional discrimination. Note that the coefficient in row 2, the pure association of test scores with moving into social housing neighbourhoods at some point, is now also insignificant, which illustrates that the KS3 performance of 'SH-movers' does not generally differ from 'SH-stayers'.

It is worth noting that the main findings hold for all specifications and are not at all sensitive to the inclusion of control variables such as previous test scores or fixed effects. This is a direct result of the strong balancing of individuals who move into social housing neighbourhoods at different times. I will return to the issue of balancing in the next section. In fact, if the timing of moving is exogenous, the inclusion of control variables should not make any difference. In small samples, however, there is a trade-off between precision and finite sample bias.

To summarise the results, the traditional approach results in large and significant negative associations between living in or moving into social housing neighbourhoods, and schooling. These associations persist despite the inclusion of a rich set of control variables including a test score measure of prior ability and school fixed effects. However, the difference-in-difference results show that the negative associations between moving into deprived social housing neighbourhoods and test scores are driven by characteristics common to pupils who move into these neighbourhoods at some point, and not by neighbourhood exposure before taking the test. Using the timing of a move as source of exogenous variation, there is no evidence for detrimental short-term effects from moving into a deprived social housing neighbourhood.

4.3 Heterogeneity

The previous section showed results for effects of general deteriorations in neighbourhood quality. As already discussed in the lower half of section 3.2 (and table 2), pupils who move into social housing move into a neighbourhood with higher unemployment levels, lower qualification rats, lower access to transport vehicles, a higher share of lone parents and people with a limiting long term illness, more overcrowding, fewer rooms per household, higher density and lower house prices. My results so far suggest that there is no overall effect on KS3 test scores of these 'treatments' combined. However, this finding does not preclude the possibility of heterogeneous effects. In this section, I present tests for potential heterogeneity in four different dimensions.[7]

Before discussing the findings of this exercise, I note that when allowing for heterogeneous effects in my difference-in-difference framework, interactions need to be included for all relevant group variables. Therefore all regressions presented in table 5 include main interaction effects and interactions with the general moving dummies as well. This means that for each specification, five additional terms are added: one main effect, two in interaction with the general moving dummies, and two further interactions that are social-housing-move specific. In table 5, I only report the coefficients for the interactions with the social housing move, which are of main interest. Notice that I use the unconditional specification from table 4, column (1) as

[7] I tried further interactions but never found significant effects, which is why the discussion here is limited to four potential dimensions.

reference point in this exercise, although using additional control variables does not alter any of my conclusions below.

Column (1) of table 5 presents results for a regression that allows for a different treatment effect for pupils who move into a social housing neighbourhood and also change secondary school. It is possible that lower neighbourhood quality only matters if the school environment changes as well. If this is the case, then there should be significant differences between those two groups. Indeed, the estimate for the interactions between changing school and moving into social housing before the KS3 test is negative at 1.770 percentile points (first row). However, the standard error is very large and this estimate is not significant.[8]

Next, I split the treatment by gender to allow for the possibility that boys and girls experience different effects. This is motivated by some of the recent literature finding gender differences in neighbourhood effects. Kling (*et al.* 2005), for example, find different neighbourhood effects for female and male youth on criminal activity. In column (2) I find a negative effect interaction effect for boys of -2.453. To the contrary, the effect of moving before the KS3 test for females, now captured by the dummy indicating a pre-KS3 social housing move shown in the third row, is positive at 0.795. Taken together, girls and boys could be affected differentially by up to three percentile points. However, again neither of the coefficients, nor the difference in these estimates, are significant in a statistical sense.

Summarising the findings from the first two columns of table 5: it is possible that effects are heterogeneous, especially along the dimensions, but it is difficult to draw final conclusions because of the imprecision of the estimates. This is because the effects are estimated of interactions with further dummy variables, limiting the amount of variation available for estimation.

Next, I consider interactions with continuous variables, namely the change in the neighbourhood level unemployment rate (column 3) and the change in the percentage of lone parents with dependent children (column 4). Overall, pupils moving into social housing experiences large deteriorations in these indicators (see table 2), so hopefully

[8] Including school fixed effects moves this estimate closer to zero in magnitude (-0.66), remaining insignificant at any conventional level.

these examinations are more informative. Indeed, the standard errors in columns (3) and (4) of the main estimate of interest are much lower, allowing us to draw economically relevant conclusions. However, the estimates themselves are very close to zero: the interaction of a one percentage-point increase in the neighbourhood unemployment-level change is estimated at 0.010 for pre-KS3 social housing movers. The corresponding coefficient for the lone parent indicator is estimated at -0.016, very close to zero and not significant despite the greater precision in the estimates.

To conclude the discussion on heterogeneous effects: interacting the difference-in-difference framework with dichotomous indicators like school-changes or gender results in imprecise estimates that make it difficult to draw final conclusions. However, looking in detail for potential heterogeneity in continuous neighbourhood level indicators, I fail to detect any significant results. In particular, there is no evidence of heterogeneity in the effect for changes in neighbourhood level unemployment and the share of lone parents with dependent children. Overall, these results confirm the previous conclusions that there is no evidence for negative neighbourhood effects for teenagers moving into social housing.

5 Assessing the identification strategy

5.1 Balancing of individual and neighbourhood characteristics: graphic analysis

Recall the identifying assumption of this study that early and late movers into social housing neighbourhood are statistically identical. If early and late movers had different characteristics, this could potentially confound the analysis that links differences in exposure-times to social housing neighbourhoods to school performance.

The data allows me to directly address this concern. Figure 2 shows averages of individual characteristics and neighbourhood change for pupils moving into social housing neighbourhoods, by year. Panel A shows the percentage of pupils who were eligible for free school meals in year 7, their gender and KS2 result. Notably, all these characteristics are determined before anyone moves and can hence not be endogenous to the quality of the new neighbourhoods. The figure clearly shows that pupils who move into social housing neighbourhoods are very similar across the years. Regardless of the year, about 50 per cent are eligible for free school meals, slightly less than half are male and KS2 results average around 34 percentile points.

As discussed in the last paragraph of section 2.4, one potential threat to the identification might be that the timing of negative shocks that make parents eligible for social housing confounds the comparison of later test scores between early and later movers. If pre-move shocks had differential impacts on test scores, this should equally show up in the KS2. The fact that the KS2 results of early and late movers look extremely balanced is therefore particularly comforting.

In our setting, I can also check whether changes in neighbourhood quality differ depending on the year of the move. This is another way to indirectly test for identification. I would expect the change in neighbourhood quality (the underlying treatment) to be balanced with respect to the year of moving into a social housing neighbourhood. Panel B of figure 2 shows the negative changes in neighbourhood quality that pupils experience by year of move. What we can see is that the shocks are similar over the years. Regardless of the year of relocation, pupils move into neighbourhoods with larger percentages of lone parents, more overcrowding, higher unemployment rates, lower qualification levels, lower access to cars and lower house prices. This further supports the causal interpretation of the social housing neighbourhood effects in our setup.

5.2 Balancing of individual and neighbourhood characteristics: regression analysis

While the figures discussed in section 5.1 are reassuring, we can also test whether early movers differ from post-KS3 test movers into social housing neighbourhoods formally using a probit regression. Here, *Pr=1* denotes the probability of moving into social housing in the years before the KS3, Φ the cumulative distribution function of the standard normal distribution (probit function), X the matrix of regressors and β the coefficients that are estimated by Maximum Likelihood.

$$\Pr(d(\text{SH-Move})t\text{-}1,t{=}1 \mid X) = \Phi\,(X'\beta) \tag{6}$$

Table 6 presents estimates of marginal effects for specification (6). The coefficients reported in column (1) are estimated using the 2,094 pupils who move into social housing at some point and the dependent variable equals one if the pupil moves before the KS3 test. If the identification assumption is violated, the KS2 score which correlates highly with the KS3 should be particularly prone to pick up differences between early and late movers. But as we can see from the marginal effects estimates in the first row of table 6, early and late movers are literally identical in respect to previous attainment. This difference is estimated at -0.0097 and not statistically significant. Notice that similar conclusions hold for the other pre-determined variables like free school meal eligibility in year 7, gender or ethnicity, as shown by the remain estimates in column (1).

The second column presents estimates for 2,977 pupils who moved out of social housing during the study period. I have so far not explicitly focussed on these pupils in the analysis because there are fewer reasons to believe that the timing of moving out of social housing could be exogenous. Essentially, this is because there are no waiting lists for moving out of social housing. However, even for these pupils, I cannot predict the year of move using a rich set of background variables including prior KS2 test scores.

Finally, the third column of table 6 shows that even non-social-housing neighbourhood movers are quite balanced with regard to the timing of the move. For this group, there is a highly significant relation between KS2 test scores and the timing of the move, but the coefficient is small and estimated at 0.0173. This means that each additional point in the KS2 test makes moving early 1.73 per cent more likely. This regression is estimated using over 106,427 pupils who move once and between non-social-housing neighbourhoods during the study period, of which about 56 per cent actually move before the KS3. In other words, early non-social-housing neighbourhood movers do better in terms of pre-determined KS2 test scores, than late movers. Notice that this will bias me towards finding negative neighbourhood effects for the social housing movers in the difference-in-difference framework, which is not what I will find.

Another important assumption for the validity of the difference-in-difference approach is that there are no direct income effects resulting from moving into social housing. If parents who move into social housing before the KS3 had a higher disposable income, this could counteract potential negative neighbourhood influences. However, I have argued before in section 2.3 that housing benefits are administered in such a way as to net out income effects from moving into social housing, and I therefore do not think that the income channel is of particular importance for my setup. To test for this directly, table 6 also includes indicators for the free school meal status in the academic years 7 and 8 as regressors (second and third rows). These estimates are not significantly different to zero, this means that even free school meal eligibility in years 8 and 9, which are not a pre-determined measures for the early movers, fail to predict the timing of the move for social housing neighbourhood movers (column 1). In other words, the time-sensitive free school meal indicator gives does not show any reaction to moving into social housing, which is comforting but also in line with expectations.

To conclude the discussion of table 6, in the last row I test the hypothesis that all coefficients jointly equal zero. It turns out that in column (1) I fail to reject the null for the social housing neighbourhood movers. However, for non-social-housing neighbourhood movers I can reject the null of joint insignificance, although the estimated coefficients are not very dissimilar in terms of magnitude (column (3)). Given these results, I therefore cannot completely rule out the possibility that that social housing neighbourhood movers *look* balanced partly due to large standard errors. Notice, however, that the balancing test presented in table 6 is unconditional on school and neighbourhood fixed effects.

To investigate this possibility further I run additional balancing regressions where I can also include school fixed effects. Since pupils can in fact choose secondary schools relatively independently of exact residential location, sorting into schools does not need to be exactly correlated to the timing of the move. We know that there is a strong sorting mechanism in England; it would hence be comforting to look at the balancing conditional on school fixed effects. This can be done by running balancing regressions where individual characteristics (in particular the KS2 test scores) are used as a dependent variable and predicted by the timing of the move. This setup then allows us to keep the whole sample, including pupils who do not move, which in turn makes it possible to correctly estimate school fixed effects.

Table 7 reports estimates for such balancing regressions that use the KS2 test score as dependent variable. Column (1) and (3) report estimates for social-housing-neighbourhood movers, while columns (2) and (4) focus on non-social-housing neighbourhood moves, and columns (3) and (4) include school fixed effects. The reported results in table 7 show that using the timing of the move as independent variable, OLS regressions on KS2 scores are not significant for SH-movers but again significant for non-SH movers. As before, however, the signs are reversed, clearly indicating that SH-movers are different to other movers at least with regard to the timing of moving. Once school fixed effects are included (columns (3) and (4)), the coefficient for moving into social housing neighbourhoods before the KS3 stays insignificant and gets smaller and very close to zero, whereas the coefficient for non-SH-before-KS3-moves stays significant and becomes larger in size. Again, I read these results as supporting the identification assumption that the timing of SH-moves is quasi-exogenous for social housing neighbourhood movers, and in fact different to the timing behaviour of non-social-housing neighbourhood pupils.

To summarise this section, the observable characteristics of pupils who move into social housing are balanced against the timing of their move. This does not rule out that those moving later are different on unobservable characteristics but makes it unlikely if one assumes that unobservable characteristics track observable characteristics (as in Altonji et al. 2005). Therefore, I conclude that the balancing regressions provide indirect evidence of the validity of the identification assumption.

5.3 Identifying social housing neighbourhood movers

A data limitation of this study is that I am unable to exactly identify pupils who move into social housing neighbourhoods. Instead, I need to rely on Output Area information from the UK 2001 Census of Population to determine if a neighbourhood is social housing or not, as explained in section 3.1. Since only a handful of neighbourhoods have 100 per cent social tenants, all OAs with at least 80 per cent social tenants were classified as social housing neighbourhoods. Note that neighbourhood quality is negatively correlated with the threshold level. Neighbourhoods with at least 20 per cent social tenants are worse than neighbourhoods with at least 10 per cent social tenants, but better than those with at least 30 per cent regarding the various neighbourhood characteristics. I impose this somewhat arbitrary threshold to focus on pupils who move into neighbourhoods with at least 80 per cent of social renters. This means that someone who moves from a neighbourhood with 79 per cent social renters to one with 81 per cent is now coded as 'moving into social housing'. Taking the regression from table 4 (column 4) as a benchmark, the first row of table 8 addresses this concern directly and only counts a move as into social housing if it was out of a neighbourhood with a maximum of 20 per cent and into a neighbourhood with at least 80 per cent social tenants. The results are insensitive to this modification, and we will return to table 8 later.

Another way of testing if the choice of the threshold level influences the findings is to run separate regressions for different cut-off points. The sensitivity of the main result to the definition of this threshold is shown in figure 3. Panel A shows results for the traditional-approach regression as in table 3 (column 5), and panel B for the main results from table 4 (column 2), both for the specifications including school fixed effects. The dashed black line plots the estimates for the 'traditional' control strategy and the solid line for the difference-in-difference estimates. First, we can clearly see that the estimated negative neighbourhood effect becomes larger as we increase the threshold in the 'traditional' approach. The estimated effect of moving from a neighbourhood with less than 10 per cent social tenants to a neighbourhood with at least 10 per cent is zero (panel A) or close to zero (panel B) but increases quickly in size and significance, shifting the threshold level up. The difference-in-difference estimate, on the other hand, remains constant around zero, suggesting that there is no neighbourhood effect regardless of the definition of the threshold. This suggests that the increasing negative effects in the 'traditional' estimates reflect unobserved characteristics that correlate negatively with KS3 results and neighbourhood quality. This is in line with the main finding that the negative association between neighbourhood quality and school results disappears once controlling for moving into the social housing neighbourhood at some point.

5.4 Sample selection, imprecise measure of timing

I further checked the sensitivity of the main finding against specific sample selection issues. These results are reported in table 8, where I only show the estimates of the main coefficient of interest, the effect of moving into social housing before the KS3.

As already discussed in the previous section, in the first row of table 8 I estimate specification (1) but now coding pupils as moving into social housing only, if they moved out of a neighbourhood with a maximum of 20 per cent social tenants and into a neighbourhood with at least 80 per cent. Again, there is no significant result, and early moves are even associated with a 0.400-point increase in KS3 test scores, which is in line with my main findings.

Another concern is that the KS3 test is not taken on the exact date that residential information is collected. In particular, the residential information is collected mid of each January, while the KS3 is taken over the spring. This means that up to a third of pupils coded as moving in year 9 to 10 might in fact have moved just before the KS3 tests were taken, although residential mobility is usually lower during the winter period. In the second row of table 8, I therefore exclude from the analysis all pupils for whom I cannot be fully confident that they moved after the test was taken. This means that I compare KS3 test results of pupils who move into social housing neighbourhoods in the academic years 7-8 or 8-9 to pupils who move into social housing neighbourhoods in the years 10-11 only. The estimates for this sample, negative 0.532 and insignificant) remains in line with our main results.

The third row estimates the specification using the first cohort only. All specifications include a cohort effect but this cohort effect is not interacted with all the other variables. If our results were cohort-specific this would cast serious doubts on the external validity of the findings. However, the effect of moving before the KS3 test is non-significant for both cohorts. As it turns out, for the first cohort, the estimate is negative and non-significant and for the second (not shown here) it is positive and insignificant. This strengthens the interpretation that there is no significant effect from moving into high-density social housing neighbourhoods.

Finally, the last row excludes 'stayers' from the regression. 'Stayers' were included to gain precision but their inclusion does not drive any the results, and the causal effect of moving before the KS3 is estimated at -0.013 points only.

Summarising the findings form table 8: the main results do not seem to be driven by the specific way in which I identify movers into social housing neighbourhoods.

5.5 Lengthening the exposure time to social housing neighbourhoods

Another potential concern is that it takes longer for neighbourhood effects to operate. To at least partially address this concern, table 9 includes pupils who move between the academic years 6, which corresponds to the end of primary school, and year 7, the first year of secondary education. Hence, here we compare pupils who move into high-density social housing neighbourhoods during the three (not two) years prior to taking the KS3 test to pupils who move during the first two years after the test. The cost of this setup is that I can only use one cohort of pupils, which approximately halves the sample size.

Table 9 reports the results of this exercise and is organised in a similar way as lower panel of the main results table 4. Two estimates are reported in the first row, where the first estimate is the effect of moving into social housing before the test, here between the academic years 6 and 9. The second row shows the coefficient for the dummy that indicates if a pupil moves into a social housing neighbourhood at some point over the study period, here the extended period from academic year 6 to year 11. Moving from columns (1) to (4), individual controls including KS2 test scores, school fixed effects and finally neighbourhood fixed effects are included.

Just as in the main results table 4, the estimate for the effect of moving into social housing before the KS3 test is never sizeable nor significant in any of the specifications. The unconditional estimate equals -0.990, but turns positive to 0.883 once control variables are included, remains positive (0.864) in column (3) and becomes close to zero (0.083) once neighbourhood fixed effects are included. Again, none of these estimates is significantly different from zero, and I conclude that moving into a social housing neighbourhood during the three years prior to the KS3 test again does not correlate with the results.

6 Discussion and Conclusions

This study estimates the effect of moving into a very deprived neighbourhood, as identified by a high density of social housing, on the educational attainment of 14-year-old (9th grade) students in England. Neighbourhoods with markedly high concentrations of social housing have very high unemployment and extremely low qualification rates, as well as high building density, over-crowding and low house prices. In order to identify the causal impact of neighbourhood deprivation on pupil attainments, I exploit the timing of moving into these neighbourhoods. The timing of a move can be taken as exogenous because of long waiting lists for social housing in high-demand areas. Contrary to previous studies in the social housing context, this strategy does not rely on exogenous allocation of people to neighbourhoods. Here, it is only required that the timing of such moves is unrelated to personal characteristics. This is a new strategy that bypasses the usual sorting and endogeneity problems.

Using this approach, there is no evidence for otherwise negative short-term effects. This suggests that the underachievement of pupils who move into social housing neighbourhoods cannot be causally linked to place characteristics during the formative teenage years.

This paper makes a number of important contributions. First, I highlight the importance of control strategies in neighbourhood research. I demonstrate that 'traditional' control strategies that simply include more variables on observable characteristics fail to identify the effect. I think that the focus on temporal differentiation must not remain limited to the social housing context but is applicable to all situations where supply restrictions in a specific neighbourhood introduce randomness into the timing of residential moves. Secondly, this study has a strong external validity. I do not need to rely on a small set of neighbourhoods or cities but estimate the effects using census data for the entire English secondary school population over two cohorts and neighbourhoods in all areas of England. Thirdly, the problem of choosing the right level of spatial aggregation is solved by using very disaggregated data. Finally, the focus is on pupils moving into social housing neighbourhoods, which is of the greatest policy relevance in times of excess demand. With the most recent budget cuts it can be expected that the number of people on waiting lists will rise above the current level of 4.5 million. The finding of no negative effect of low social housing neighbourhood quality on school attainments during the formative teenage years raises interesting questions regarding how to meet this demand, and the current social housing policy of mixed communities.

7 Bibliography

Aaronson, D. (1998). Using sibling data to estimate the impact of neighbourhoods on children's educational outcomes, *Journal of Human Resources*, pp. 915-46.

Angrist, J. and Pischke, J.-S. (2009). *Mostly harmless econometrics: an empiricist's companion*, Princeton University Press, London.

Akerlof, G. (1997). Social Distance and Social Decisions, *Econometrica*, 65(5), pp. 1005-1027.

Altonji, G., T. Elder and C. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools, *Journal of Political Economy*, 113(1): pp. 151-184.

Bayer, P., Ross, L.R. & Topa, G. (2008). Place of Work and Place of Residence: Informal Hiring Networks and Labour Market Outcomes, *Journal of Political Economy*, 116(6): pp. 1150-1196.

Bernheim, B.D. (1994). A Theory of Conformity, *The Journal of Political Economy*, 102(5), pp. 841-877.

Briggs, X. de Souza (1998). Brown kids in white suburbs: Housing mobility and the many faces of social capital, *Housing Policy Debate*, 9(1), pp. 177-221.

Black (1999). Do Better Schools Matter? Parental Valuation of Elementary Education*, *The Quarterly Journal of Economics*, 114(2), pp. 577-99.

Calvó-Armengol, A. & Jackson, M.O. (2004). The Effects of Social Networks on Employment and Inequality, *American Economic Review*, 94(3), pp. 426-454.

Cheshire, P., Gibbons, S. & Gordon, I. (2008). Policies for 'Mixed Communities': A Critical Evaluation, *SERC Policy Paper*, 2.

Cutler, D.M. & Glaeser, E.L. (1997). Are Ghettos Good Or Bad?*, *The Quarterly Journal of Economics* 112(3), pp. 827-72.

Durlauf, S. (1996). A Theory of Persistent Income Inequality, *Journal of Economic Growth*, I, pp. 75-93.

Fehr, E. & Falk, A. (2002). Psychological foundations of incentives, *European Economic Review*, 46(4-5), pp. 687-724.

Gibbons, S. (2002). Neighbourhood effects on educational achievement: Evidence from the Census and National Child Development Study, *Centre for the Economics of Education Discussion Paper*, 16.

Gibbons, S., Machin, S. & Silva, O. (2008). Valuing School Quality Using Boundary Discontinuity Regressions, *SERC Discussion Paper*, 16.

Gibbons, C., Silva, O. & Weinhardt, F. (2010). Do Neighbours Affect Teenage Outcomes? Evidence from neighbourhood changes in England, *LSE-CEE Discussion Paper*, 112.

Glaeser, E. & Scheinkman, J. (2001). Measuring Social Interactions, in *Social Dynamics*, Durlauf, S. & Young, P. (eds.), Boston,MA: MITPress.

Gould, E.D., Lavy, V. & Pasterman M.D. (2004). Immigrating to Opportunity: Estimating the Effect of School Quality using a Natural Experiment on Ethiopians in Israel, *The Quarterly Journal of Economics* 119(2): 489-526.

Goux, D. & Maurin, E. (2007). Close neighbours matter: neighbourhood effects on early performance at school, *Economic Journal*, 117(523), pp. 1193-215.

Granovetter, M., (1995): *Getting a Job*, second edition, Chicago, IL: University of Chicago Press.

Gurmu, S., Ihlanfeldt, K.R. & Smith, W.J. (2008). Does residential location matter to the employment of TANF recipients? Evidence from a dynamic discrete choice model with unobserved effects, *Journal of Urban Economics*, 63, pp. 325-351.

Hills J (2007). Ends and Means: The future roles of social housing in England, Report: London School of Economics and Political Science, CASEreport 34.

Hobbs, G. & Vignoles, A. (2007). Is free school meal status a valid proxy for socio-economic status (in schools research)?, *Centre for Economics of Education Discussion Paper,* 84.

Holmans, A. (2005). Housing and Housing Policy in England 1975-2002, Report: Office of the Deputy Prime Minister, London.

Jacob, B.A. (2004), Public housing, housing vouchers, and student achievement: Evidence from public housing demolitions in Chicago, *The American Economic Review*, 94(1), pp. 233-58.

Kain, J.F. & Quigley, J.M. (1972). Note on owner's estimate of housing value, *Journal of the American Statistical Association*, 67(340), pp. 803-6.

Katz, L.F., Kling, J.R. & Liebman, J.B. (2001). Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment, *The Quarterly Journal of Economics* 116 (May 2001), 607-54.

Kling, J., Ludwig, J. & Katz, L. (2005). Neighbourhood Effects on Crime for Female and Male Youth: Evidence From a randomized Housing Voucher Experiment, *Quarterly Journal of Economics*, 120, pp. 87-130.

Lavy, V., Silva, O. & Weinhardt, F. (forthcoming). The Good, the Bad and the Average: Evidence on the Scale and Nature of Ability Peer Effects in Secondary Schools, *Journal of Labor Economics*.

Lupton, R. (2005). Social justice and school improvement: improving the quality of schooling in the poorest neighbourhoods, *British Educational Research Journal*, 31(5), pp. 589-604.

Lupton, R. (2009). Growing up in social housing in Britain, Report: Joseph Rowntree Foundation, York.

Manski, C.F. (1993). Identification of endogenous social effects: The reflection problem, *The Review of Economic Studies*, 60(3), pp. 531-42.

Moffitt, R.A. (2001). Policy interventions, low-level equilibria, and social interactions, in *Social Dynamics*, Durlauf, S. & Young, P. (eds.), Boston,MA: MITPress.

Oreopolous, P. (2003). The Long-Run Consequences of Living in a Poor Neighbourhood*, *The Quarterly Journal of Economics*, 118(4), pp. 1533-1575

Orford, S., Dorling, D., Mitchell, R., Shaw, M. & Davey Smith, G. (2002). Life and death of the people of London: a historical GIS of Charles Booth's inquiry, *Health and Place*, 8(1), pp. 25–35.

Small, M. (2009). *Unanticipated Gains: Origins of Network Inequality in Everyday Life*, Oxford University Press.

Taubman, P. (1989). Role of Parental Income in Educational Attainment, *The American Economic Review*, 79(2), pp. 57- 61.

Turley (2009). Council Housing -Back to the Future?, Report: New Local Government Network, London.

Regan, S. & Patrick, R. (2001). *Squeezed Out*, Institute for Public Policy Research, London.

Rosenbaum, J.E. (1995). Changing the Geography of Opportunity by Expanding Residential Choice: Lessons from the Gautreaux Program, *Housing Policy Debate*, 6(1), pp. 231-69.

Rutter, J. & Latorre, M. (2009). Social housing allocation and immigrant communities, Report: Institute for Public Policy Research, London.

Sampson, R.J., Morenoff, J.D. & Gannon-Rowley, T. (2002). Assessing "Neighbourhood Effects": Social Processes and New Directions in Research, *Annual Review of Sociology*, 26, pp. 443-478.

Sanbonmatsu, L., Kling, J.R., Duncan, G.J. & Brooks-Gunn, J. (2006). Neighbourhoods and academic achievement: Results from the Moving to Opportunity Experiment, *Journal of Human Resources*, 41(4), p. 649.

Weinhardt, F. (2010). Moving into the Projects: Social Housing Neighbourhoods and School Performance in England, SERC Discussion Papers 0044, Spatial Economics Research Centre, LSE.

Zenou, Y. (2008). Social Interactions and Labour Market Outcomes in Cities, *IZA Discussion Paper* 3283.

Table 1: Descriptive Statistics

	(1)		(2)		(3)		(4)	
	Pupil stayed in SH n'hood during study period		Pupil moved into SH n'hood during study period		Pupil stayed in non-SH n'hood during study period		Pupil moved btw. non-SH n'hoods during study period	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
Panel A: Individual characteristics								
Key Stage 2 Score	38.641	24.229	37.258	24.332	51.317	25.902	47.064	25.685
Key Stage 3 Score	35.629	23.721	33.332	23.710	51.507	26.439	46.409	26.111
Changed school before, yr 7-9	0.043	0.202	0.106	0.308	0.021	0.144	0.095	0.293
FSME eligibility year 7	0.494	0.500	0.498	0.500	0.143	0.350	0.205	0.404
FSME eligibility year 8	0.484	0.500	0.494	0.500	0.139	0.346	0.197	0.398
FSME eligibility year 9	0.467	0.499	0.493	0.500	0.133	0.340	0.187	0.390
Gender (male=1)	0.500	0.500	0.484	0.500	0.508	0.500	0.497	0.500
Ethnicity-White British Is.	0.629	0.483	0.694	0.491	0.830	0.376	0.804	0.397
Ethnicity-Other White	0.036	0.187	0.032	0.176	0.017	0.130	0.020	0.139
Ethnicity-Asian	0.065	0.246	0.053	0.223	0.066	0.248	0.065	0.247
Ethnicity-Black	0.166	0.372	0.138	0.345	0.030	0.169	0.044	0.205
Ethnicity-Chinese	0.008	0.088	0.007	0.082	0.003	0.055	0.003	0.054
Ethnicity-Mixed	0.043	0.203	0.036	0.187	0.021	0.145	0.025	0.156
Ethnicity-Other	0.028	0.164	0.019	0.137	0.006	0.080	0.010	0.098
Panel B: Neighbourhood characteristics, pre move (if any)								
Unemployment rate	0.117	0.048	0.079	0.045	0.045	0.037	0.054	0.042
Level 4+ qualification[1]	0.489	0.114	0.548	0.130	0.618	0.131	0.603	0.133
Access to car or van[2]	0.500	0.128	0.649	0.169	0.830	0.151	0.787	0.167
Lone parent with dep. child	0.199	0.090	0.124	0.070	0.073	0.100	0.087	0.066
Limiting long term illness	0.431	0.100	0.386	0.098	0.344	0.100	0.351	0.103
Overcrowding[3]	0.198	0.131	0.132	0.110	0.066	0.076	0.081	0.086
Number of rooms	4.291	0.537	4.782	0.648	5.439	0.824	5.230	0.797
Population density[4]	133.978	158.608	86.643	91.936	53.187	49.823	61.066	62.365
Average house price[5]	0.617	0.630	0.716	0.496	0.931	0.537	0.840	0.499
Panel C: Secondary school characteristics, year 7								
Pupil to teacher ratio	15.734	1.856	15.877	1.808	15.850	1.555	15.894	1.601

Notes: Neighbourhood classified as Social housing if at least 80% of residents in social rented sector. Key stage scores are percentiles computed on the whole cohort. Only pupils who always lived in Local Authority with more than 5% of population on Social Housing waiting list included. For SH stayers 10k observations, SH movers 2,094 observations. For non-SH stayers 474k observations, non-SH movers 109k observations. All movers only move once. Panel B: Neighbourhood characteristics as in academic year 7 (before the move). 1) First degree, Higher degree, NVQ levels 4 and 5, HNC, HND, Qualified Teacher Status, Qualified Medical Doctor, Qualified Dentist, Qualified Nurse, Midwife or Health Visitor, 2) age households that can access at least on car or van, 3) Index as used in Census 2001, a value of 1 implies there is one room too few, 4) people per hectare, 5) Average house price: All property sales in neighbourhood between 2000 and 2006 divided by monthly national average price.

Table 2: The treatment: neighbourhood changes for pupils moving into social housing

	(a) New SH n'hood	(b) % ch.	(c) S.D. ch.
Unemployment rate	0.122	54.43%	1.089
Level 4+ qualification	0.470	-14.08%	-0.589
Access to car or van	0.497	-23.42%	-0.947
Lone parent with dep. child	0.194	56.45%	1.116
Limiting long term illness	0.441	14.25%	0.542
Overcrowding	0.169	28.03%	0.453
Number of rooms	4.333	-9.39%	-0.540
Population density	112.151	29.44%	0.446
Average house price	0.550	-23.08%	-0.312

Notes: Only pupils who always lived in a Local Authority with more than 5% of population on Social Housing waiting list included. 2,094 obs. Variables defined as in previous Tables.

Table 3: Social housing and school performance, traditional approach

	Panel A			*Panel B*			*Panel C*		
	Lived in SH neighbourhood in year 7			*Moved into SH neighbourhood before KS3 test*			*Moved into SH n'hood before or after the test*		
Dependent variable: KS3 test scores	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimated effect on Key Stage 3 score:	-14.837 (0.260)**	-2.899 (0.161)**	-1.540 (0.140)**	-13.251 (0.703)**	-2.722 (0.413)**	-1.454 (0.373)**	-13.077 (0.534)**	-2.882 (0.316)**	-1.667 (0.299)**
Key Stage 2 score	-	0.849 (0.001)**	0.820 (0.001)**	-	0.850 (0.001)**	0.820 (0.001)**	-	0.850 (0.001)**	0.820 (0.001)**
Changed secondary school before KS3	-	-3.060 (0.107)**	-1.669 (0.115)**	-	-3.252 (0.107)**	-1.854 (0.006)**	-	-3.086 (0.107)**	-1.670 (0.115)**
FSME eligibility year 7	-	-2.935 (0.091)**	-1.920 (0.087)**	-	-3.005 (0.092)**	-1.948 (0.087)**	-	-2.998 (0.092)**	-1.944 (0.087)**
FSME eligibility year 8	-	-1.468 (0.112)**	-0.991 (0.106)**	-	-1.494 (0.112)**	-0.999 (0.106)**	-	-1.494 (0.112)**	-0.999 (0.106)**
FSME eligibility year 9	-	-2.118 (0.097)**	-1.459 (0.092)**	-	-2.162 (0.097)**	-1.469 (0.092)**	-	-2.998 (0.092)**	-1.470 (0.092)**
Gender (male==1)	-	-1.411 (0.035)**	-1.249 (0.036)**	-	-1.412 (0.035)**	-1.251 (0.036)**	-	-1.410 (0.035)**	-1.249 (0.036)**
Pupil to teacher ratio, year 7	-	-0.499 (0.014)**	(absorbed)	-	-0.497 (0.014)**	(absorbed)	-	-0.498 (0.014)**	(absorbed)
Control for moving into social housing	No	No	No	No	No	No	No	No	No
Controls for effects of moving	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity-controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
School fixed effects	No	No	Yes	No	No	Yes	No	No	No

Notes: Neighbourhood classified as Social housing if at least 80% of residents in social rented sector. Over 596k obs., errors clustered at neighbourhood level. Only pupils who always lived in Local Authority with more than 5% of population on Social Housing waiting list. Standard errors in brackets. ** sig. at 1%.

Table 4: Main results: social housing and school performance, the causal effect

Dependent variable: KS3 test scores	(1) Moved into SH n'hood	(2) Moved into non-SH n'hood	(3) First Difference	(4) DiD
Panel A: Unconditional means				
Move before KS3 test	33.598	46.849	-13.251	-0.365
Move after KS3 test	32.962	45.847	-12.886	
Panel B: Regression estimates				
Move into SH neighbourhood before KS3 test	-0.365 (1.056)	0.426 (0.628)	0.539 (0.597)	0.267 (0.651)
Move into SH neighbourhood before or after KS3 test	-12.886 (0.801)**	-3.152 (0.481)**	-2.000 (0.454)**	0.097 (0.515)
Key Stage 2 score	-	0.850 (0.001)**	0.820 (0.001)**	0.830 (0.001)**
Changed secondary school before KS3	-	-3.251 (0.107)**	-1.854 (0.116)**	-2.763 (0.120)**
FSME eligibility year 7	-	-3.001 (0.092)**	-1.946 (0.087)**	-1.439 (0.101)**
FSME eligibility year 8	-	-1.494 (0.112)**	-1.466 (0.092)**	-0.924 (0.123)**
FSME eligibility year 9	-	-2.156 (0.097)**	-1.466 (0.092)**	-1.058 (0.107)**
Gender (male==1)	-	-1.412 (0.035)**	-1.251 (0.036)**	-1.525 (0.040)**
Pupil to teacher ratio, year 7	-	-0.497 (0.014)**	(absorbed)	-0.549 (0.019)**
Control for moving into social housing	Yes	Yes	Yes	Yes
Controls for effects of moving	Yes	Yes	Yes	Yes
Ethnicity-controls	No	Yes	Yes	Yes
School fixed effects	No	No	Yes	No
Output Area fixed effects (after move)	No	No	No	Yes

Notes: Neighbourhoods classified as Social housing if at least 80% of residents in social rented sector. SH movers who move only once. Only pupils who always lived in Local Authority with more than 5% of population on Social Housing waiting list. Over 596k obs., errors clustered at neighbourhood level. Standard errors in brackets. ** sig. at 1%.

Table 5: Testing for heterogeneity: interactions with moving into SH neighbourhoods

	(1)	(2)	(3)	(4)
Dependent variable: KS3 test scores	Changed School before KS3	Gender (male=1)	Change in n'hood % unemp't.	Change in n'hood % lone parents
Interaction * Move into SH n'hood before KS3	-1.770 (4.220)	-2.453 (2.128)	0.010 (0.169)	-0.016 (0.097)
Interaction* Move into SH n'hood before or after KS3	1.368 (3.790)	2.536 (1.620)	0.020 (0.127)	-0.021 (0.074)
Move into SH neighbourhood before KS3	-0.359 (1.10)	0.795 (1.462)	-0.114 (1.269)	0.324 (1.250)
Move into SH neighbourhood before or after KS3	-12.801 (0.816)**	-14.124 (1.085)**	-13.838 (0.980)**	-13.731 (0.926)**
Control for moving into social housing	Yes	Yes	Yes	Yes
Controls for effects of moving	Yes	Yes	Yes	Yes
Ethnicity-controls	No	No	No	No
School fixed effects	No	No	No	No
Output Area fixed effects (after move)	No	No	No	No

Notes: Baseline regressions is Table 4 (column 1). Interaction main effects and for non-SH movers always included (coefficients not reported here). Neighbourhoods classified as social housing if at least 80% of residents in social rented sector. Movers only move once. Only pupils who always lived in Local Authority with more than 5% of population on Social Housing waiting list. Over 596k obs., errors clustered at neighbourhood level. Standard errors in brackets. ** sig. at 1%.

Table 6: Probability of moving in the two years before versus after the KS3 test

	(1)	(2)	(3)
	Moving into SH n'hood	Moving out of SH n'hood	Non-SH n'hood move
Key Stage 2 score	-0.0097	-0.0076	0.0173
	(0.044)	(0.039)	(0.006)**
FSME eligibility year 7	-0.030	0.015	0.024
	(0.033)	(0.028)	(0.006)**
FSME eligibility year 8	0.051	0.009	-0.004
	(0.038)	(0.331)	(0.008)
FSME eligibility year 9	0.006	-0.037	-0.016
	(0.032)	(0.029)	(0.007)**
Gender (male==1)	0.011	-0.011	0.012
	(0.021)	(0.018)	(0.003)**
Ethnicity-White British Isles	0.121	0.046	0.009
	(0.075)	(0.062)	(0.009)
Ethnicity-Other White	0.051	-0.024	0.006
	(0.093)	(0.086)	(0.014)
Ethnicity-Asian	0.130	0.009	0.021
	(0.080)	(0.073)	(0.011)$
Ethnicity-Black	0.121	-0.041	0.006
	(0.073)	(0.066)	(0.012)
Ethnicity-Chinese	0.163	-0.134	0.010
	(0.127)	(0.158)	(0.030)
Ethnicity-Mixed	-0.228	0.074	-0.001
	(0.095)	(0.073)	(0.013)
Ethnicity-Other	-0.001	-0.080	0.005
	(0.006)	(0.096)	(0.018)
Teacher to pupil ratio (y7)	-0.001	0.002	-0.002
	(0.006)	(0.006)	(0.001)*
Cohort	-0.010	-0.026	-0.022
	(0.022)	(0.019)	(0.003)**
School FX	No	No	No
H0: All coefficients equal zero. Prob > chi2	0.2996	0.1461	0.000

Notes: Dependent variable equals one if pupil move before KS3 in sample where everyone move once and into Social Housing neighbourhoods, hence either before or after KS3. (a): 2,094 obs.; (b): 2,977 obs.; (c): over 105,000 obs. Probit regression, marginal effects. Standard errors in brackets and clustered at neighbourhood level. Only pupils who always lived in Local Authority with more than 5 per cent of population on Social Housing waiting list included.

Table 7: Balancing regressions by type of move

Dependent Variable: KS2 test scores	(1) Moving into SH n'hood, OLS	(2) Non-SH n'hood move, OLS	(3) Moving into SH n'hood	(4) Non-SH n'hood move
Move before KS3	-0.437 (1.082)	0.396 (0.159)*	0.092 (1.053)	0.443 (0.148)**
Move	-12.802 (0.824)**	-4.208 (0.126)**	-7.023 (0.809)**	-2.908 (0.117)**
School FX	NO	NO	YES	YES

Notes: Neighbourhood classified as Social housing if at least 80% of residents in social rented sector. Over 596k obs., errors clustered at neighbourhood level. Only pupils who always lived in Local Authority with more than 5% of population on Social Housing waiting list. Standard errors in brackets. ** sig. at 1%, * sig. at 5%.

Specifications:

For SH n'hood in-movers (column 3):

$$\begin{aligned} \textit{Test Score}_{i,n,t-1} = \alpha_i & + \kappa_1\, d(\textit{SH-Move})_{i,t,t-1} \\ & + \kappa_2\, d(\textit{SH-Move})_{i,t-1,t+1} \\ & + \boldsymbol{D}(\textit{secondary school at enrolment})_i{}'\boldsymbol{\kappa}_3 \\ & + \kappa_4\, d(\textit{cohort})_i + \mu_{i,n,t} \end{aligned}$$

For non-SH n'hood in-movers (column 4):

$$\begin{aligned} \textit{Test Score}_{i,n,t-1} = \alpha_i & + \lambda_1\, d(\textit{nonSH-Move})_{i,t,t-1} \\ & + \lambda_2\, d(\textit{nonSH-Move})_{i,t-1,t+1} \\ & + \boldsymbol{D}(\textit{secondary school at enrolment})_i{}'\boldsymbol{\lambda}_3 \\ & + \lambda_4\, d(\textit{cohort})_i + \tau_{i,n,t} \end{aligned}$$

Table 8: Sample selection

	Effect of moving into SH n'hood
Panel A: 20% vs 80% threshold:	0.400
	(1.231)
Panel B: Excluding y9-10 movers:	-0.532
	(0.857)
Panel C: Only first cohort:	-0.392
	(1.059)
Panel D: Only movers:	-0.013
	(1.063)

Notes: All regressions include Output Area (neighbourhood) fixed effects, like Table 4 (column 4). Standard errors in brackets.

Table 9: Expanding the treatment period, years 6-9 and years 9-11 movers

	(1)	(2)	(3)	(4)
Move into SH neighbourhood before KS3, between years 6 and 9.	-0.990 (1.348)	0.883 (0.825)	0.864 (0.783)	0.083 (0.992)
Move into SH neighbourhood before or after KS3, between years 6 and 11	-12.767 (1.145)**	-3.596 (0.704)**	-2.344 (0.669)**	-0.177 (0.861)
Control for moving into social housing	Yes	Yes	Yes	Yes
Controls for effects of moving	Yes	Yes	Yes	Yes
Ethnicity-controls	No	Yes	Yes	No
School fixed effects	No	No	Yes	No
Output Area fixed effects (after move)	No	No	No	No

Note: Here we compare KS3 scores for pupils who move into social housing neighbourhoods during three years prior to taking the KS3 compared to pupils who move during the two years afterwards. Neighbourhoods classified as Social housing if at least 80 per cent of residents in social rented sector. Movers only move once. Only pupils who always lived in Local Authority with more than 5% of population on Social Housing waiting list. 280k observations (based on only one cohort), 2,419 schools.

Figure 1: The English School System and identification

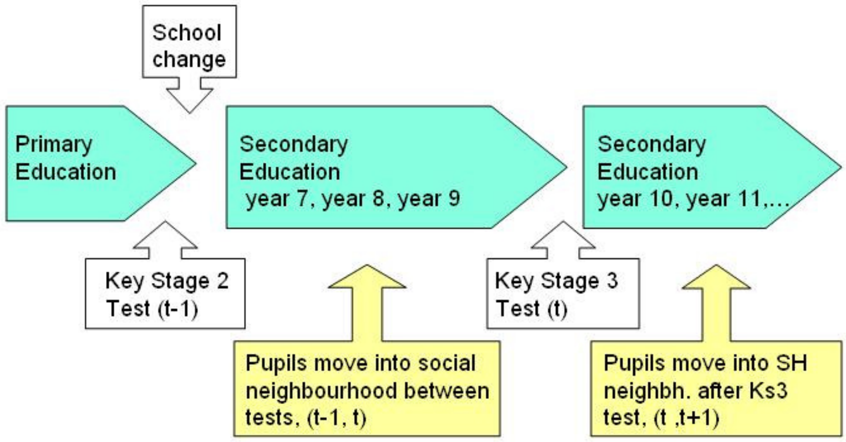
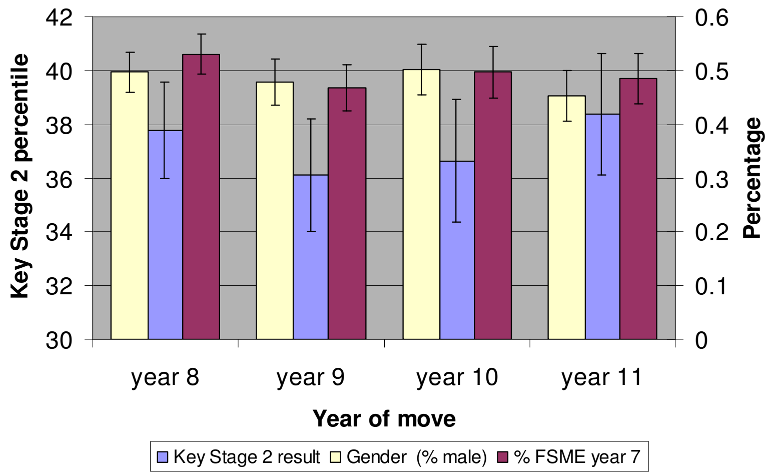


Figure 2: Balancing of pupils who move into SH neighbourhoods

Panel A: Test results, gender, FSME (income)



Panel B: Change in neighbourhood quality for SH-movers

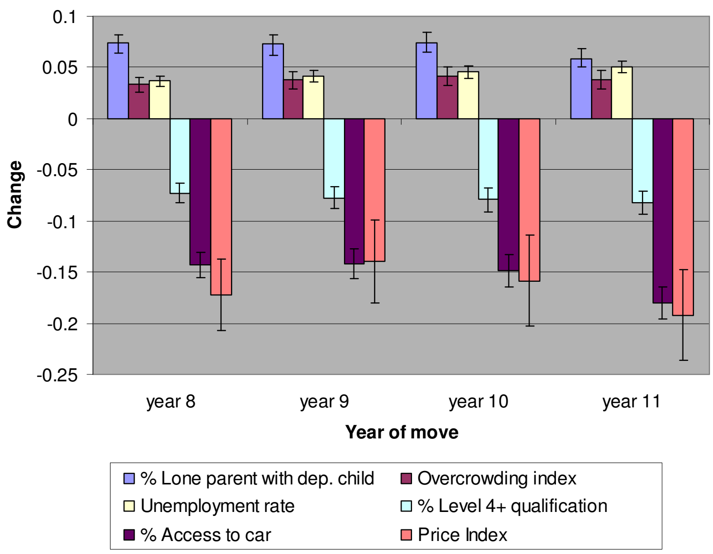
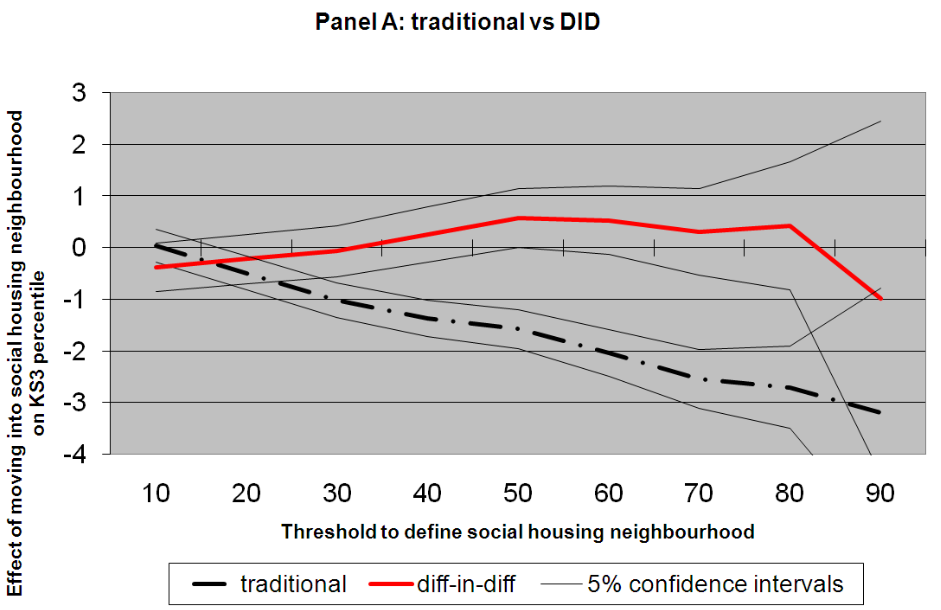
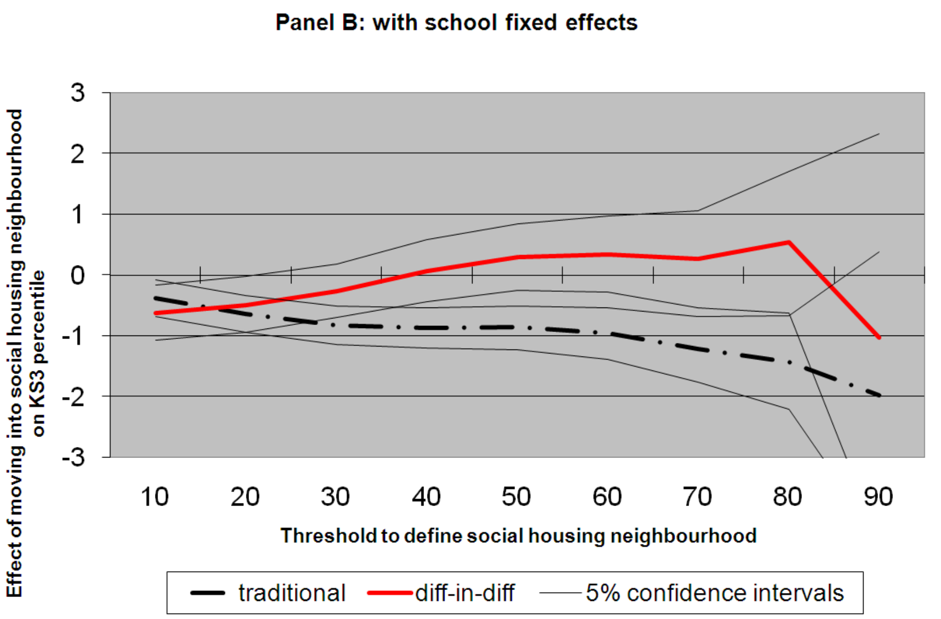


Figure 3:
Changing the threshold definition of social housing neighbourhoods





CHAPTER III

EVIDENCE ON THE IMPORTANCE OF PEERS IN THE NEIGHBOURHOOD

1 Introduction[9]

There are evidently significant disparities between the achievements, behaviour and aspirations of children growing up in different neighbourhoods (Lupton et al., 2009). These disparities have long been a centre of attention for researchers and policy makers concerned with addressing socioeconomic inequalities. Indeed, many area-based policies, including inclusionary zoning and desegregation policy in the US, and the 'Mixed Communities Initiative' in England, are predicated on the idea that individuals' outcomes are causally linked to the social interactions with others who live around them (see discussions in Currie, 2006 for the US, and Cheshire et al., 2008 for the UK). However, the question of whether differences between children's outcomes are truly causally related to the type of people amongst whom they live remains difficult to answer. Even though a large body of empirical literature has focussed on estimating 'neighbourhood effects' in residential neighbourhoods and peer effects in schools, researchers have come to varying conclusions depending on the data and methods used for the analysis.[10]

Nearly all studies in this field proceed – as does ours – by trying to learn about neighbourhood effects from the statistical associations between individual outcomes and the socioeconomic composition of the neighbourhood in which they live.[11] However, there are at least three pervasive obstacles to this endeavour. Firstly, non-random sorting of residents into different neighbourhoods means that individual and neighbours' characteristics are correlated through 'non-causal' channels. This sorting makes it hard to disentangle whether the correlations between neighbourhood composition and individual outcomes is attributable to differences in neighbourhood

[9] This chapter presents work undertaken in the preparation for a more extensive joint paper titled: "Everybody needs good neighbours? Evidence from students' outcomes in England.", which is currently published as IZA working paper (Gibbons *et al.*, 2011) and submitted to a peer-reviewed journal. I use plural pronouns to refer to the authors.

[10] Recent examples related to the school peer effects literature include Angrist and Lang (2004) on peer effects through racial integration; Hoxby (2000) and Lavy and Schlosser (2007) on gender peer effect; Gould et al. (2011) on the effect of immigrants on native students; and Gibbons and Telhaj (2008) and Lavy et al. (2011) on ability peer effects. We discuss examples from the neighbourhood effects literature in section 2.

[11] Manski (1993) refers to these as 'contextual' effects.

composition, or to differences between individuals. Secondly, neighbourhoods that differ in terms of socioeconomic composition potentially differ along other dimensions (often unobserved), so that it becomes difficult to tell whether any observed effects are due to neighbours' interactions, or to the common coincidental factors that neighbours face.[12] Lastly, there are uncertainties and practical limitations in how to define the reference groups within which individuals interact, because 'neighbourhood effects' could arise in geographical neighbourhoods, local friendship networks, or neighbourhood schools, but this operational scale is almost always unknown. This paper presents new evidence on neighbourhood peer effects on cognitive outcomes from age 11 (grade 6) through to age 14 (grade 9), using detailed administrative data on multiple cohorts of English school children.[13] We believe our methodology and data allow us to provide more satisfactory solutions to the problems outlined above than has been previously done in the literature.

As with previous research in this field, residential sorting is an issue for our study because the characteristics of children are closely interwoven with those of their parents, who choose where to live on the basis of their preferences for local amenities and services, the income at their disposal and other constraints they face. The literature on the link between school quality and house prices (e.g. Black, 1999 and Gibbons et al., 2009) shows that people are willing to pay a significant premium to access 'better' schools (as well as other amenities; see Kain and Quigley, 1975 and Cheshire and Sheppard, 1995), and suggests that neighbourhoods will be stratified along the lines of income and socio-economic background. This sorting means that one child's characteristics – both observed and unobserved – will be correlated with those of his/her neighbours, confounding the causal influence of neighbours with children's and their parents' own inherent attributes. Even without sorting of this type, the problem of unobserved differences between neighbourhoods remains important. Explicit randomisation (e.g. the 'Moving to Opportunity' experiment, MTO) is not a solution because the neighbourhoods to which individuals are assigned potentially differ not

[12] Manski (1993) refers to these as 'correlated' effects.

[13] Note, throughout the paper we use the term 'grade' to refer to a school year group. Although the term grade is not used in the English school system, there is no convenient term with equivalent meaning.

only in terms of peer group composition, but also in terms of housing stock, labour market opportunities, school quality and other factors. For some purposes it might be sufficient to estimate the combined 'black-box' effects of these coincidental factors, but this approach does not allow separate identification of the effects arising specifically through interaction among neighbours. In order to overcome these difficulties, Moffitt (2001) suggests that researchers should 'reverse-engineer' the evaluation of programmes like the MTO or the Gautreaux intervention (Rosenbaum, 1992), and study changes in the outcomes of the original residents of the areas receiving relocated households. For these people, neighbourhoods remain approximately unchanged except in so far as their composition is affected by the influx of new families.

Following this intuition, our study tackles the problems of sorting and confounding neighbourhood attributes by exploiting changes in neighbourhood composition induced by the migration of residential 'movers' in a population of school-age families. We estimate the effect of these mover-induced changes in neighbourhood composition on the evolution of educational outcomes of 'stayers' (i.e. students who do not move neighbourhoods). Using this methodology, we are able to partial out the individual fixed effects of stayers, as well as neighbourhood fixed effects, such as the presence of a library or other localised infrastructures/amenities. We are thus able to separately identify causal effects arising specifically from changes in neighbourhood peer composition, which we attribute to neighbours' interactions and role model effects. This approach is similar to Angrist and Lang (2004), who estimate peer effects from changes in peer composition due to students' mobility induced by desegregation programmes, and to Gibbons and Telhaj (2011) who study the effect of students' between-school mobility on students who do not change school. Note though, that our method differs from the literature on peer effects in schools that exploits naturally arising cohort-to-cohort variation in group composition (e.g. Hoxby, 2000, Hanushek et al., 2003, Gibbons and Telhaj, 2008, Lavy et al., 2008) because we can control for individual fixed effects without needing these individuals to move between groups.[14]

[14] Note that while using cohort-to-cohort variation can be justified in the school setting (where pupils study with same-age peers), using this variation to study neighbourhood effects would require strong assumptions, i.e. that children do not interact with peers in the neighbourhood who are not the same age.

As already stated, our identifying variation comes from the movements of residents in and out of neighbourhoods on those who stay put – i.e. it is induced by real changes in the neighbourhood experienced by stayers. To address potential sample selection concerns arising from estimation using stayers only, we conduct an additional intention-to-treat analysis that includes movers in the estimation sample, but assigns them to the neighbourhoods in which they originate (thus fixing their neighbourhood assignment, and avoiding problems induced by endogenous neighbourhood choices).

Another important feature of our data and design is that we can control for factors that simultaneously induce *changes* in movers' characteristics and stayers' outcomes within neighbourhoods over time. Firstly, the fact that we can track several cohorts of students as they progress from primary through secondary education, experiencing changes in the neighbourhood composition over a number of years, means that we can control for unobserved linear trends in neighbourhood 'quality' (e.g. 'gentrification' or deterioration in housing quality). Secondly, we can include school-by-grade-by-cohort fixed effects to allow for changes in school quality as students move between one grade and the next, and to allow for changes in the composition and quality of the group of schools represented in each neighbourhood (i.e. attended by its residents). This is feasible – and necessary in our context – because students change schools between grades, and because there is not a one-to-one mapping between residential neighbourhood and school attended, with different students in the same residential neighbourhood attending two to three different secondary schools, and secondary schools enrolling students from around sixty different residential areas.

Like other neighbourhood effects studies, we also face the problem of defining the operational reference group for a child's social interactions. In common with most other research, we have no information on actual friendship networks (which are in any case prone to problems of sorting and self-selection), so we must approximate the level at which interactions take place. However, whereas much research is limited in the way reference units can be defined (e.g. census tracts), we have precise geographical detail on residential location coupled with information on school attended and children's age. This richness in our data means we can start by defining neighbourhoods at a very small scale, and then experiment with larger groupings of contiguous neighbourhood units (similar to Bolster et al., 2007). We can also modify these groups to allow for peer interactions between students of different ages capturing interactions within the same birth-cohort and across adjacent birth-cohorts. These groups can be quite finely delineated: our smallest geographical units (Census Output Areas or OAs) contain an average of 5 students of the same age, and 8 students in adjacent birth cohorts (+1/-1 year). We can further split the reference groups into neighbours who attend the same school and neighbours who attend different schools, allowing us to separate peer effects in neighbourhoods from peer effects and other shared influences in schools.

To preview our findings, we show that the large cross-sectional correlation between young peoples' test score outcomes and neighbourhood composition – measured in terms of prior achievement, eligibility for free school meals (an indicators for low family income) and special education needs (a proxy for learning disabilities) – is dramatically reduced once we control for individual and neighbourhood fixed effects by looking at changes in the neighbourhood peer composition over time. Any remaining significant association is eliminated once we control for school-by-cohort effects and/or neighbourhood-specific time trends. Differentiating between effects for neighbours in the same school and neighbours in different schools still yields no evidence that peer composition matters either way.

The rest of the paper is structured as follows. The next section reviews the literature, while section 3 describes our empirical strategy and section 4 discusses that data that we use and the English institutional context. Next, section 5 discusses our findings and robustness checks, while section 6 provides some concluding remarks.

2 Literature Review: Previous Methods and Findings

While neighbourhood effects could arise for a number of reasons, economists have put substantial emphasis on peer group and role model effects (Akerlof, 1997 and Glaeser and Scheinkman, 2001), social networks (Granovetter, 1995 and Bayer et al., 2008), conformism (Bernheim, 2004 and Fehr and Falk, 2002) or local resources (Durlauf, 1996). Disappointingly though, it has proved very difficult to distinguish between these competing theories empirically and research has mainly concentrated on estimating a general 'contextual' effect that does not delineate the causal channels.

These studies have used a variety of methods to address biases caused by residential sorting. These methods include: *(i)* instrumental variables for neighbourhood quality (Cutler and Glaeser, 1997 and Goux and Maurin, 2007); *(ii)* institutional arguments related to social renters who have limited choice in relation to where to live, and limited mobility across social housing projects (Gibbons, 2002, Oreopolous, 2003, Jacob, 2004, Goux and Maurin, 2007, Weinhardt, 2010); *(iii)* quasi-experimental placement policies for immigrants (Edin et al., 2003 and 2011, Gould et al., 2011); and *(iv)* fixed-effects estimations to partial out individual, family and aggregate unobservables (Aaronson, 1998, and Bayer et al., 2008). Finally, there have been a number of experimental studies looking at randomised control-trial interventions, namely the 'Gautreaux' and 'Moving to Opportunity' programmes (Rosenbaum 1995, Katz et al. 2005 and 2007, Sanbonmatsu et al. 2006).

Overall, the literature tends to find negligible effects on educational attainments, but some effects on behavioural outcomes, such as involvement in criminal activities or health status (Katz et al., 2007). However, the distinction between the effects of better *neighbours* and those of better *neighbourhoods* is often blurred. Competing explanations, in particular the importance of social interactions with neighbours as opposed to local resources, infrastructures and school quality, are simply brushed aside. For example, Goux and Maurin (2007) do not control for the quality of local schools and other neighbourhood infrastructures. Similarly, most of the MTO based studies (Kling et al. 2005, 2007, Sanbonmatsu et al. 2006) treat neighbourhoods as a 'black box', although more recent work has started to unpick the contributory factors (Harding et al., 2010). Some studies have tried to distinguish between school and neighbourhood level variables. Card and Rothstein (2007) investigate the effects of racial segregation at the city level on the black-white test score gap in the US. Their results suggest that any effect is driven by neighbourhood segregation, rather than school segregation, although the authors cannot reject the null of equality between the two effects. On the other hand, Gould et al. (2004), who are primarily interested in the effect of school quality on the educational outcomes of Ethiopian immigrants in Israel, show that additional neighbourhood level variables have no explanatory power. Even then, although these studies control for school level variables, they still do not distinguish between the effects of neighbourhood peers and those of other local factors.

On a more general note, the fact that the existing empirical literature has not taken a clear stance on this issue has led to some confusion about what constitutes a 'neighbourhood effect'. Notably, it is not uniformly agreed whether differences in outcomes driven by local school quality constitute a neighbourhood effect or not, even though this distinction has important policy implications. To be clear from the outset, our study specifically aims at estimating peer effects in the neighbourhood. These represent neighbourhood effects that arise from social interactions and role models at the place of residence, and net of potential confounding effects such as differences in local school quality (e.g. school resources, teaching methods, but also quality of its intake) and other local infrastructure/resources. To this end, we exploit the richness of our data which allows us to estimate neighbourhood-peer effects, while controlling for neighbourhood fixed effects (including neighbourhood infrastructures), neighbourhood trends and school-by-cohort effects. The next section spells out our empirical strategy in detail.

3 Empirical strategy

3.1 General identification strategy: a changes-in-changes specification

Our empirical work concentrates on identifying the effect of neighbourhood peers on students' educational outcomes during secondary schooling. As outlined in the introduction, the estimation of neighbourhood peer effects is greatly complicated by the sorting of individuals across neighbourhoods in relation to both observable and unobservable local factors. This sorting implies that there will be a strong degree of correlation between the characteristics of an individual in the neighbourhood and those of his/her neighbours, and as well as potential correlation between local factors and the characteristics of its residents. Any study that aims to estimate the causal influence of neighbourhood peers must therefore eliminate the biases that arise from the fact that neighbourhood peer group quality is correlated with individual-level and neighbourhood-level unobservables, which directly affect individual outcomes. We use a changes-in-changes design that eliminates these unobserved components. A novelty of our study is that we explicitly restrict any measured neighbourhood variation to that caused by movements of students in our sample from one neighbourhood to another. Moreover, the size of our administrative population-wide data and the fact that we observe multiple cohorts means that we can control carefully for unobserved neighbourhood fixed effects, neighbourhood-specific unobserved time trends and school-by-cohort specific shocks. The rest of this section sets out our simple linear empirical model more formally, in order to elucidate in what ways these various data transformations take account of individual and neighbourhood level unobservables.

Assume that students' outcomes depend linearly on the characteristics of peers in the neighbourhood, other neighbourhood infrastructures and individual characteristics:

$$y_{insct} = z_{nct}\beta + \mathbf{x}'_i\gamma + \mathbf{x}'_i\delta t + \varepsilon_{insct} \tag{1.1}$$

where y_{insct} denotes the outcome of student i living in neighbourhood n, attending school s, belonging to birth cohort c and measured at grade or age t. Note that school grade is equivalent to age, since there is no grade repetition in England. In the empirical analysis, we look at test outcomes from grade 6 to grade 9. We observe students' test scores at grades 6 and 9 (ages 11 and 14), and attended school and place of residence for these grades as well as all those in between. In this specification, z_{nct} is a variable measuring *neighbour-peer* composition, e.g. mean prior achievements of peers in the neighbourhood or the proportion from low-income families. Our definition of these neighbour-peers is set out in sections 3.3 and 4.3 below. The vector $\boldsymbol{x}_i$ contains time-fixed predetermined observable student characteristics, which we allow to have a time-trending effect captured by δt. Furthermore, we assume that the error term has the following components:

$$\varepsilon_{insct} = \alpha_i + \phi_n + \xi_n t + \vartheta_{sct} + e_{insct} \tag{1.2}$$

where α_i represents an unobserved individual-level fixed effect that captures all constant personal and family background characteristics; Φ_n represents unobserved time-fixed neighbourhood characteristics – such as access to a good public library and other infrastructures – and $\xi_n t$ represents neighbourhood unobserved trending factors –

such as gentrification dynamics. Finally, ϑ_{sct} is a school-by-cohort-by-grade specific shock. Among other things, this term is intended to capture variation in school resources, composition and or quality of teaching that are common to students attending the same schools *s* in a given grade – e.g. grade-6 (age-11) – and belonging to the same cohort *c*. Finally, the term e_{insct} is assumed to be uncorrelated with all the right hand side variables. Endogeneity issues arise because the components a_i, Φ_n, $\xi_n t$ and ϑ_{sct} in equation (1.2) are potentially correlated with z_{nct} and x_{it} in equation (1.1).

In order to eliminate some of the unobserved components that could jointly determine neighbour-peer composition and students' outcomes, we exploit the fact that we observe students as they progress from primary through secondary education, and know their outcomes and the composition of the neighbourhood where they live at different school grades (ages). We can therefore take within-student differences between two grades and estimate the following equation:

$$\left(y_{insc1} - y_{insc0}\right) = \left(z_{nc1} - z_{nc0}\right)\beta + \mathbf{x}'_i\,\delta + (\varepsilon_{insc1} - \varepsilon_{insct0}) \qquad (2.1)$$

Where the subscripts *t=0* and *t=1* identify the initial and subsequent grade (e.g. grade 6 and grade 9), and the exact grade interval varies according to the outcome under consideration. Notice that when we estimate this model we restrict our estimation sample to students who *do not move* neighbourhood. This implies that neighbour-peer changes $(z^p_{nc1} - z^p_{nc0})$ depend on inflows and outflows of movers who are not in the estimation sample. The within-individual, between-grade differencing for stayers reduces the error term to:

$$(\varepsilon_{insc1} - \varepsilon_{insct0}) = \xi_n + (\vartheta_{sc1} - \vartheta_{sc0}) + \nu_{insct} \qquad (2.2)$$

and so eliminates both the individual (a_i) and the neighbourhood (Φ_n) unobserved components that are fixed over time for students and their residential neighbourhoods, including unobserved ability, family background and other forces driving sorting of families across different neighbourhoods. One caveat to this approach is that focussing on stayers could give rise to selectivity issues and bias our estimates of neighbourhood effects. To allay these concerns, in one of our robustness checks we include movers and stayers, and assign to movers the changes in the neighbour-peer quality they would have experienced had they not moved. In this second set-up, our estimates of the neighbourhood effects are more properly interpreted as intention-to-treat effects.

Equation (2.2) shows that this grade-differenced specification does not control for school quality factors that change between grades for a given student. The between-grade school quality change term $\vartheta_{sc1} - \vartheta_{sc0}$ in equation (2.2) is likely to be non-zero, especially because students change schools over the grade intervals that we study. In particular, students go through a compulsory school change from primary to secondary school, between grades 6 and 9. Furthermore, their secondary school 'quality' could change because of new leadership, changes in the teaching body or variation in school resources. This possibility poses a threat to our identification strategy because school quality changes for students in neighbourhood *n* might influence the inflow and outflow of students, as well as the characteristics of in/out-migrants into neighbourhood *n*, which would in turn affect changes in neighbourhood peer composition, $z_{nc1} - z_{nc0}$. Differencing between cohorts is unlikely to eliminate these school quality effects, because they are not necessarily fixed across cohorts.[15] In some specifications we therefore control for secondary-school-by-cohort fixed effects, or secondary-by-primary-school-by-cohort fixed effects (effectively school-by-grade-by-cohort fixed effects). We can, however, further control for more general unobserved neighbourhood-specific time trends ξ_n relating to general neighbourhood changes such as regeneration, gentrification or decline of some neighbourhoods relative to others, by differencing from neighbourhood means across cohorts *c*.[16]

Our identifying assumption in these models is that the remaining idiosyncratic shocks to student outcomes (after eliminating student fixed effects, neighbourhood fixed effects, school-by-cohort effects and/or neighbourhood trends) are uncorrelated with the changes in neighbourhood composition experienced by student *i* as he/she stays in the residential neighbourhood between grades *t=0* and *t=1*. Our results include

[15] Note also that the school effects may vary by cohort within the same neighbourhood not only because the quality of schools is changing, but also because different cohorts in the same neighbourhood attend a different mix of schools.

[16] Note that if we want to allow for both neighbourhood trends and school-by-cohort fixed effects in our specifications, we need to implement a multi-way fixed effects estimator. To do so, we use the Stata's routine `felsdvreg`.

a set of balancing regressions that supports the empirical validity of this assumption, showing that changes in the neighbour-peer composition are not strongly related to time-fixed neighbourhood characteristics or time-fixed average characteristics of the students living in the neighbourhood, even before we allow for neighbourhood unobserved trends or school-by-cohort effects. This lends credibility to our identification strategy.

3.2 Distinguishing neighbourhood from school peer effects

In England, there is not a one-to-one link between neighbourhood and school attended, but students in a given neighbourhood tend to attend a mixed group of local schools, their choices being influenced by travel costs and school admissions policies that tend to prioritise local residents (see section 4.1). On average, students in the same age-group and living in the same small neighbourhood (hosting five such students) attend two to three different secondary schools. Therefore, we can separately identify the effect of changes in neighbourhood peer composition for neighbours who attend the same secondary school, and for those who do not. More formally, we can estimate the following model that partitions neighbourhood peers into two groups, those that go to the same secondary school (*same*) as student *i*, and those that attend other secondary schools (*other*):

$$\left(y_{insc1}-y_{insc0}\right)=\left(z_{nc1}-z_{nc0}\right)^{same}\beta+\left(z_{nc1}-z_{nc0}\right)^{other}\gamma+\mathbf{x}'_i\delta+(\varepsilon_{insc1}-\varepsilon_{insct0}) \tag{3}$$

Most variables in equation (3) were defined above. The variable $(z_{nc1}- z_{nc0})^{same}$ refers to changes in neighbour-peer composition driven by the mobility of peers who attend the same school as *i* at grade *t=1* (e.g. at grade 9 at secondary school). These students are therefore peers *both* in the neighbourhood and at secondary school. Note however that schools are attended by students from a large number of residential areas: in our sample, on average secondary schools attract students from sixty different neighbourhoods. This implies that same-neighbourhood-same-school peers are only a small fraction of the peers that students interact with at school. On the other hand, the variable $(z_{nc1}- z_{nc0})^{other}$ captures changes to the neighbour-peer composition that are driven by neighbourhood peers who *do not* attend the same school as *i*. Any differences between the coefficients β and γ will shed light on the relative contribution of school and neighbourhood peers. More importantly, whereas peer effects (β) among neighbouring students who attend the same school might pick up interactions among students in schools, peer effects among neighbouring students who go to different schools (γ) should capture a 'pure' neighbourhood-social-interaction effect. As before, we can difference equation (3) within neighbourhoods, across cohorts to eliminate neighbourhood trends, and can control for school-by-cohort fixed effects.[17]

3.3 Defining neighbourhood geography

Research on social interactions in the neighbourhood shares many of the empirical issues that the literature on peer effects at school has had to face in terms of defining group membership and measuring peers' characteristics, but has the additional complication of having to define the 'right scale' of the neighbourhood. While there is some discussion of whether the effects of social interactions should be measured at the grade or class level in the peer effects literature (see Ammermueller and Pischke, 2009), there are no similar natural boundaries such as school or classroom that define the area of interest in the case of neighbourhoods. Consequently, what has been used to measure neighbourhood effects has varied greatly with respect to geographical size. Goux and Maurin (2007) speculate that using large neighbourhood definitions – i.e. US Census tracts containing on average 4000 people – leads to an underestimate of interaction effects. However, over-aggregation on its own will not necessarily attenuate regression estimates of neighbourhood effects since any reduction in the covariance between mean neighbours' characteristics and individual outcomes is offset by a reduction in the variance of average neighbours' characteristics. Nonetheless, it is crucial that the neighbourhood group definition includes relevant neighbours, and in this respect a larger neighbourhood definition might be better than a small one if the small group is mis-specified.

All in all, whether or not the level of aggregation matters in practice is an empirical question. We take full advantage of the detail and coverage of our population-wide

[17] Note that school-by-cohort fixed effects can still be controlled for in equation (3) because students living in the same area attend a number of different schools, and schools attract students from a large number of different neighbourhoods so that the terms $(z_{nc1} - z_{nc0})^{same}$ and $(z_{nc1} - z_{nc0})^{other}$ in equation (3) are not perfectly collinear with the term $(\vartheta_{sc1} - \vartheta_{sc0})$.

data to experiment with alternative geographical definitions, starting from a very small scale unit - Output Areas (OA) from the 2001 British Census - which contains 125 households on average and approximately five students in the same age-group (e.g. five, 6th grade, age-11 students). Notice that, since our identification approach relies on neighbourhood fixed effects to control for unobserved neighbourhood factors, a small scale neighbourhood definition minimises the risk of endogeneity of neighbourhood quality (that is, it is less likely that there are unobserved neighbourhood changes over time within-streets, than within-regions). Nevertheless, we experiment with larger geographical areas based on this underlying OA-geography. This allows us to tackle the problem of defining a suitable spatial unit in neighbourhood research in a highly flexible way.

Another advantage of our data is that we observe the population of English school children[18] and can measure neighbour-peer composition using students in a variety of school grades. Since we are interested in social interactions in the neighbourhood, we argue that these neighbour-peer variables should be constructed aggregating the characteristics of students of similar age. This neighbour definition is motivated by the idea that students of similar age are more likely to interact and/or be influenced by similar role models. For this reason, in the majority of our paper we construct neighbour-peer variables using individual level data from student who are either of the same school grade (i.e. grade 6, age 11 at the beginning of our observation window) or one year younger/older (grade 5 or grade 7, from age 10 up to age 12). However, we perform a number of checks using different grade-bands, for example by including only students in the same school grade. Note finally that the neighbour-peer variables are constructed from information on students' characteristics that pre-date the first period of our analysis, using a balanced panel of students with non-missing data in every year of the census. This set up implies that changes over time in neighbour-peer composition occur only when students within our sample move across neighbourhoods, and not when students drop out/come into our sample, or when their

[18] Our dataset is a census of multiple cohorts of all children in state-education in England. No comparable information is available for the private sector, which has a share of about 7%.

personal characteristics change. More detail on the neighbour-peer variables is provided in section 4.3 below.

The complex data that we use in order to pursue this analysis is described in the next section alongside the English institutional background.

4 Institutional Context and Data Setup

4.1 The English school system

Compulsory education in England is organized into five stages referred to as Key Stages (KS). In the primary phase, students enter school at grade 1 (age 4-5) in the Foundation Stage, then move on to KS1, spanning grades 1-2 (ages 5-7). At grade 3 (age 7-8), students move to KS2, sometimes – but not usually – with a change of school. At the end of KS2, in grade 6 (age 10-11), children leave the primary phase and go on to secondary school, where they progress through KS3, from grade 7 to 9, and KS4, from grade 10 to 11 (age 15-16), which marks the end of compulsory schooling. Importantly, the vast majority of students change schools on transition from primary to secondary education between grades 6 and 7. Students are assessed in standard national tests at the end of each Key Stage, generally in May, and progress through the phases is measured in terms of Key Stage Levels.[19] KS1 assessments test knowledge in English (Reading and Writing) and Mathematics only and performance is recorded using a point system. On the other hand, at both KS2 and KS3 students are tested in three core subjects, namely Mathematics, Science and English and attainments are recorded in terms of the raw test scores. Finally, at the end of KS4, students are tested again in English, Mathematics and Science (and in another varying number of subjects of their choice) and overall performance is measured using point system (similar to a GPA), which ranges between 0 and 8.[20]

Admission to both primary and secondary schools is guided by the principle of parental choice and students can apply to a number of different schools. Various

[19] KS3 assessments were dropped in 2009, which marks the end of our data period.

[20] Details on the weighting procedures are available from the Department for Education (formerly Department for Children, Schools and Families) and the Qualifications and Curriculum Authority.

criteria can be used by over-subscribed schools to prioritize applicants, but preference is usually given first to children with special educational needs, next to children with siblings in the school and to children who live closest. For Faith schools, regular attendance at local designated churches or other expressions of religious commitment is foremost. Because of these criteria – alongside the constraints of travel costs – residential choice and school choice decisions are linked (see some related evidence in Gibbons et al, 2008 and 2009, and in Allen et al., 2010). Even so, most households will have a choice of more than one school available from where they live. Indeed, on average students in the same-age bracket (e.g. age-14 students) living in the same Output Area (OA) – i.e. our smallest proxy for neighbourhoods sampling on average five such students – attend two to three different secondary schools every year, and each secondary school on average samples students from around sixty different OAs (out of more than 160,000 in England). As already mentioned, this feature of the institutional context allows us to measure changes in neighbourhood peer composition for students who attend the same or a different school. If school attendance was more tightly linked to residential location, we would not be able to discriminate between these two groups.

4.2 Main data source and grade 6 (KS2) to grade 9 (KS3) tests

To estimate the empirical models specified in section 3, we draw our data from the English National Student Database (NPD). This dataset is a population-wide census of students maintained by the Department for Education (formerly Department of Children Schools and Families) and holding records on KS1, KS2, KS3 and KS4 test scores and schools attended for every state-school student from 1996 to the present day. Since 2002 the database has been integrated with a Pupil Level Annual School Census (PLASC, carried out in January), which holds records on students' background characteristics such as age, gender, ethnicity, special education needs and eligibility for free school meals. The latter is a fairly good proxy for low income, since all families who are on unemployment and low-income state benefits are entitled to free school meals (Hobbs and Vignoles, 2009). Crucially for our research, PLASC also records the home postcode of each student on an annual basis. A postcode typically corresponds to

15 contiguous housing units on one side of a street, and allows us to assign students to common residential neighbourhoods and to link them to other sources of geographical data. In particular, we use data from PLASC to map every student's postcode into the corresponding Census Output Area (OA, described above).

The main focus of our analysis will be the period spanning grade 6 (age 11, end of KS2) to grade 9 (age 14, end of KS3), but we report results for other time periods and outcomes (discussed in detail later). The main advantage of concentrating on this grade interval and these outcomes is that the data provides comparable measures of performance in English, Mathematics and Science at grade 6 (KS2) and grade 9 (KS3). We exploit this feature to construct measures of students' test-score value-added which allow us to estimate the changes-in-changes specification spelled out in section 3.1. Operationally, we average each student's performance at KS2 and KS3 across the three subjects, then convert these means into percentiles of the cohort-specific national distribution, and finally create KS2-to-KS3 value-added by subtracting age-11 from age-14 percentiles. Note that we restrict our attention to students in schools that do not select students by academic ability (i.e. 'comprehensive' schools).

Given the time-span of the NPD-PLASC integrated dataset and our data requirements, we track several birth cohorts of students as they progress through education. For our main analysis, we retain students in the four 'central' cohorts, namely students in grade 6 (taking KS2 tests) in academic years 2001/2002, 2002/2003, 2003/2004 and 2004/2005, who move on to grade 9 (KS3 tests) in the years 2004/2005, 2005/2006, 2006/2007 and 2007/2008. We use other cohorts to construct the neighbour-peer variables as described in section 4.3 below. Finally, we concentrate on students who live in the same OA over the period covering grade 6 (age 11) to grade 9 (age 14), which we label as 'stayers' (we will address issues of selectivity caused by focussing on the stayers in our robustness checks). After applying these restrictions, we obtain a balanced panel of approximately 1.3 million students spread over four cohorts.

4.3 Data on neighbour-peer composition

Using NPD/PLASC information, we construct measures of neighbour-peer composition based on neighbourhood aggregates of student characteristics. These neighbour-peer characteristics are: *(i)* Average grade 3 (KS1) score in English (Reading and Writing) and Mathematics; *(ii)* Share of students eligible for free school meals (FSM); *(iii)* Share of students with special education needs (SEN) ; *(iv)* Fraction of males. FSM and SEN status are based on students' status in the first year they appear in the data. We use KS1 scores to proxy students' academic ability at the earliest stages of primary education, FSM eligibility as an indicator for low family income, and SEN as a proxy for learning difficulties and disabilities. The fraction of SEN neighbour-peers is based on students deemed by the school to have special educational needs, which includes those who have official SEN 'statements' from their local education authority. Finally, the share of males has been highlighted as important in previous research on peer effects (see Hoxby, 2000 and Lavy and Schlosser, 2007).[21] To construct these neighbour-peer aggregates, we use individual level data from all students who live in the same OA and are either in the same grade (i.e. grade 6, age 11 at the beginning of our observation window) or in the school grade above or below (from grade 5 up to grade 7).[22] Note that we keep OA neighbourhoods in our estimation sample only if there are at least 5 students in the OA in these grade/age categories. Note too that we keep a balanced panel of students with non-missing information in all years, so that neighbourhood quality changes are driven by the same students moving in and out of the local area, and not by students joining in and dropping out of our sample. Given the quality of our data, this restriction amounts to excluding approximately 2% of the initial sample.

Figure 1 provides a graphical representation of the time-window in the data and the construction of the neighbourhood peer groups. For example, Cohort 1 is the cohort of children in grade 6 and taking KS2 in 2002, who go on to secondary school in 2003 and take their KS3 in grade 9 in 2005. Neighbour-peer composition for Cohort 1 is calculated in 2002 from those in the OA who are in Cohort 1, plus those in grades 5 and 7. Neighbour composition is calculated in 2005 from Cohort 1 and grades 8 and 9.

[21] We do not observe immigrant status and so cannot perform an analysis similar to Edin et al. (2003) and (2010).

[22] We also compute these proxies separately for students who attend/do not attend the same secondary school at age 14 in order to estimate the specification detailed in equation (3).

In order to check the validity of our basic neighbourhood definition, we construct some alternatives based on: *(i)* students in the same OA and the same grade only; and *(ii)* students in the same and adjacent grades, but living in a set of contiguous OAs. Specifically, for *(ii)* we create neighbourhoods that include students' own OA plus all contiguous OAs. These extended neighbourhoods include on average 6 to 7 OAs, and approximately 80 students.[23]

5 Main Results on KS2-KS3 Test Scores

5.1 Summary statistics

Descriptive statistics for the main variables for the grade 6 (KS2) to grade 9 (KS3) dataset are provided in table 1. Starting from the top, panel A presents summary statistics for the characteristics of the 'stayers'. The KS2 and KS3 sores are percentiles in the population in our database. The KS2 and KS3 percentiles are around 50, with a standard deviation of about 25 points, and mean value-added on 1.1. Note that mean value-added is not centred on zero, and the standard deviations of KS2 and KS3 percentiles are slightly smaller than theoretically expected, because we percentalised test-score variables before: *(i)* dropping students with some missing observations (approximately 2% of the initial sample); *(ii)* disregarding students in small neighbourhood (less than 5 students in the OA in the same grade), and *(iii)* considering only students who do not change neighbourhood between grades 6 and 9 (the 'stayers'). We use figures from this table to standardize all the results in the regression analysis that follows. About 15 percent of the students are eligible for free school meals (FSM), 21 percent have special educational needs (SEN) and 50 percent are male. Average secondary school size is around 1080 students, and the rates of annual inward and outward neighbourhood mobility are similar (they are based on mobility within a balanced panel) and close to 8 percent. Note finally that these figures are similar to those obtained before dropping 'movers' and students in small neighbourhood (see appendix table 1), which suggests that students and neighbourhoods in our sample are broadly representative of the students' population and England as a whole.

[23] This computationally intense task is implemented in GeoDA using rook contiguity.

Panel B of table 1 presents the means and standard deviations (unweighted) of the neighbour-peer characteristics and their changes between grades 6 and 9 (age-11/KS2 to age-14/KS3). KS1 test scores at grade 2 are measured in points (not percentiles), and a score of 15 is in line with the national average. By construction, from our balanced panel, the levels of the shares of FSM, SEN and male students are very similar to those of the underlying population of students (see panel A) and none of the neighbour-peer characteristic means changes much between grades (any changes are due to the fact that the statistics report neighbour-group means and individuals are changing group membership). Our neighbourhoods sample on average around 5 students in the same grade and 14 students in the same or adjacent grades. This means that relative to most of the previous research in the field, we focus on small groups of neighbour-peers.

The most important point to note from table 1 is the amount of variation we have in our neighbour-peer variables once we take differences to eliminate individual and neighbourhood fixed effects. Looking at the figures, we see that the standard deviation of KS1 scores is 1.76, while the change in this variable between grades 6 and 9 has a standard deviation just over 0.86. This suggests that 24% of the variance in the average KS1 scores is within-OA over time. The corresponding percentages for the shares of FSM, SEN and male students in the neighbourhood are 16%, 31% and 41%, respectively. Figures 2a and 2b illustrate this point further by plotting the distributions of the neighbourhood mean variables: *(i)* levels (top left panels), *(ii)* between-grade differences (top right panels), *(iii)* between-grade differences, after controlling for primary-by-secondary-by-cohort school effects (bottom left panels); and *(iv)* between-grade, between-cohort differences netting out OA trends (bottom right panels). All these figures suggest that there is considerable variation over time in neighbour-peer characteristics, from which we can estimate our coefficients of interest, and that controlling for school-by-cohort or OA trends does not lead to a drastic reduction in this variation.

5.2 Neighbours' characteristics and students' test score: cross sectional and causal estimates

Table 2 presents our main regression results on the association between neighbour-peer characteristics and students' test scores for the residential 'stayers' sample. The table reports *standardised* regression coefficients, with standard errors in parentheses (clustered at the OA level). As discussed in section 4.3, neighbour-peers are defined as students in the same OA and in the same or adjacent school grades, and we report the effect of: average grade 3 (KS1) point scores (*panel A*); share of FSM students (*panel B*); share of students with SEN status (*panel C*); and share of male students (*panel D*). Each coefficient is obtained from a separate regression, i.e. we enter one neighbour-peer characteristic at a time. Clearly, some of these neighbour-peer characteristics are very highly correlated with one another, but our aim is to look for effects from any one of them – interpreted as an index of neighbour-peer quality – rather than the effect of each characteristic conditional on the other. Columns (1)-(4) present results from regressions that do not include control variables other than cohort dummies and/or other fixed effects as specified at the bottom of the table. Columns (5)-(8) add in control variables for students' own characteristics as described later in this section. The note to the table provides more details.

Column (1) shows the cross-sectional association between neighbour-peer characteristics and students' own KS3 test scores. All four characteristics are strongly and significantly associated with students' KS3 scores. A one standard deviation increase in KS1 scores is associated with a 0.3 standard deviation increase in KS3, while a one standard deviation increase in FSM or SEN students is linked to a 0.2-0.3 standard deviation reduction in KS3. The fraction of males has a small positive relation with KS3 scores.

However, these cross-sectional estimates are almost certainly biased by residential sorting and unobserved individual, school and neighbourhood factors (as discussed in sections 1 and 3). In order to tackle this problem, we first eliminate student and neighbourhood unobserved fixed effects by estimating within-student, between-grade differenced specifications as set out in equations (2.1)-(2.2). The corresponding results in column (2) show that the associations between changes in neighbour-peer characteristics and KS2-to-KS3 value-added are driven down almost to zero and only significant in two out of the four panels. The coefficients are up to 100 times smaller than in column (1). A one standard deviation change in neighbour KS1 scores and in the FSM proportion over the three-year interval is linked to a mere 0.3-0.5% of a standard deviation change in students' test-score progression. Neighbours' SEN and male proportions are no longer significantly associated with students' KS2-to-KS3 value-added, and their estimated effects are close to zero.

As discussed in section 3, it is still possible that estimates from these within-student between-grade differenced models are biased by unobserved school specific factors and neighbourhood trends. In order to control for school specific factors, columns (3) adds primary-by-secondary-by-cohort fixed effects that absorb any cohort-specific shock to changes in school quality when moving from the primary to the secondary phase. Results from these specifications show that none of the neighbour-peer characteristics are now significantly related to students' KS2-to-KS3 value-added. The loss in significance is not due to a dramatic increase in the standard errors, but to the magnitude of the coefficients shrinking towards zero. This further backs the intuition gathered from Figures 2a and 2b that *in principle* there is sufficient variation to identify significant associations between neighbourhood composition and students' achievements. In order to control for neighbourhood (OA) specific time trends, column (4) further adds OA fixed effects in the value-added specification, but the results are nearly identical to those in column (3).[24] Columns (5)-(8) repeat the analysis of columns (1)-(4), but add other characteristics as control variables in the regression (namely, students' own KS1 scores, FSM and SEN status and gender, plus school size, school type dummies and average rates of inward and outward mobility in the neighbourhood). Comparing columns (1) and (4) suggests that the cross sectional associations in column (1) are severely biased by sorting and unobserved student characteristics since adding in the control variables reduces the coefficients

[24] Note that school-by-cohort effects and neighbourhood specific time trends do not capture the same things because there is not a one-to-one mapping between neighbourhood of residence and school attended. Note also that including primary-by-secondary-by-cohort effects and OA trends proved computationally not feasible, so we replaced the former with secondary-by-cohort effects.

substantially (by a factor of three). In contrast, it is important to notice that, once we eliminate student and neighbourhood fixed effects in columns (2) and (6), adding in the control set does not significantly affect our results. The only case where there is a notable change is in the effect of neighbour-peer SEN, which becomes statistically significant (at the 5% level), even though the point estimate is virtually unchanged. The similarity of the results in columns (2)-(4) with those in columns (6)-(8) is reassuring since it suggests that changes in neighbour-peer composition are not strongly linked to students' background characteristics. This finding lends initial support to our identification strategy which relies on changes in the treatment variables to be 'as good as random' once we partial out student and neighbourhood fixed effects. The next section presents more formal evidence on this point.

Once concern might be that the attenuation in the estimates once we difference the data within-student between-grades is caused by inflation in the noise to signal ratio because of noise in our neighbour-peer variables. Although our proxies are constructed from administrative data on the population of state school children, they may still be noisy measures of the 'true' neighbour attributes that matter for students' achievements (which we cannot observe), and this noise could be exacerbated by differencing the data (in particular since there is a high degree of serial correlation in the neighbour-peer characteristics within neighbourhoods). To systematically assess this issue, we performed two robustness checks. First, we used teachers' assessment of students' performance during KS1 to construct instruments for neighbour-peer KS1 test scores on the grounds that the only common components of KS1 test scores and teacher assessments should be related to 'true' underlying neighbours' abilities. Instrumental variable (2SLS) regressions confirmed that the effect of changes in KS1 test scores of neighbour-peers is not a strong and highly significant predictor of students' KS2-to-KS3 value-added. Next, in our second robustness check, we estimated a linear predictor of students' KS2 achievement by regressing students' own KS2 achievements on own KS1 test scores, FSM eligibility, SEN status and gender. The predictions from these regressions were then aggregated across neighbour-peers to create new measures of predicted neighbour-peer KS2 at grade 6 and grade 9. This new composite indicator should be less affected by measurement error in relation to the 'true' neighbourhood quality that matters for students' achievements since it is based on the best linear combination of the individual characteristics that predicts KS2 test scores. Using this measure as a proxy for neighbour-peer 'quality' produces similar results to those in table 2, with no evidence of any sizeable, significant effect from neighbours on students' achievement.[25]

In summary, our baseline results indicate that the effects of neighbour-peers on student achievement are statistically insignificant and/or negligibly small. In the following sections we assess our identifying assumptions and present several extensions and robustness tests. Since controlling for unobserved neighbourhood trends does not affect our main estimates, once we have taken into account school-by-cohort effects, the analysis that follows only considers only the basic grade-differenced value-added specifications (like columns (2) and (6)) and specifications that further control for school cohort-specific effects (like column (3) and (7)).

5.3 Assessing our identification strategy

The validity of our empirical method rests on the assumption that changes in neighbour-peer composition between grades are not related to the unobserved characteristics of students who stay in the neighbourhood over the grade interval, nor to other unobservable attributes of the neighbourhoods. We have shown already that the results of the between-grade within-individual value-added specifications are insensitive to whether or not we include additional individual, school and neighbourhood mobility control variables, which supports the validity of the identifying assumptions. In this section, we tackle this issue more systematically by providing evidence that our treatments are balanced with respect to student and neighbourhood characteristics.

[25] The reduction in coefficients from column (2) to (3), and (6) to (7) is not simply a result of the inclusion of a large number of fixed effects, especially when controlling for primary-by-secondary-by-cohort groups. To test this, we replaced primary-by-secondary-by-cohort fixed effects with actual cohort-specific changes in school-level characteristics on transition from primary to secondary school. These included student-to-teacher ratios, fraction of students of White ethnic origin, fractions of students eligible for FSM and with SEN, number of full-time equivalent (FTE) qualified teachers, and numbers of support teachers for ethnic minorities and SEN students. These specifications confirmed that neighbourhood composition is not strongly associated with students' value-added.

The neighbourhood characteristics we consider are drawn from the GB 2001 population census at OA level. Specifically, we consider proportions of: *(i)* households living in socially rented accommodation; *(ii)* owner-occupiers; *(iii)* adults in employment; *(iv)* adults with no qualifications; *(v)* lone parents. Additional characteristics are generated by collapsing some salient student characteristics from our NPD data to OA level, based on OA of residence at grade 6 (age 11), namely: KS1 test scores, FSM and SEN status and gender, as well as the mean and the standard deviation of students' KS2 test scores. We carry out simple cross-sectional OA level regressions of these neighbourhood characteristics on the OA-specific changes in neighbour-peer characteristics that we used in the regressions in table 2 (i.e. grade 6-to-9 changes in neighbour-peer KS1 test scores, and FSM, SEN and male proportions).

Standardised coefficients and standard errors from these regressions are reported in table 3. The top panel shows the association between OA-mean student characteristics and the change in neighbour-peer composition between grades 6 and 9. These regressions have no control variables other than the proportion of students in the neighbourhood from each cohort in our data and the proportions of students represented in different school types.[26] The only significant and meaningful associations that we detect are related to the changes in neighbour-peer FSM. The sign of these estimates suggests that neighbourhoods with low KS1, high FSM and high SEN experience increases in fraction of neighbours who are FSM-registered, which would imply *upward* biases in the estimates in table 2, columns (2)-(4). However, these associations are very small in magnitude. Moreover, it should be noted that we have only imperfect controls for cohort and school effects in these balancing tests, and these factors are more effectively controlled for in the specifications in table 2 which include school-by-cohort effects and neighbourhood trends.

In the bottom panel of table 3 we regress OA-level KS2 statistics and Census variables on the neighbour-peer change variables. These regressions further include OA-level averages of the controls added in the specifications of columns (4) to (8) of

[26] School 'types' include: Community, Voluntary Aided, Voluntary Controlled, Foundation, City Technology College and Academy. The cohort and school type proportions stand in for the cohort-by-school effects in our main student level regressions, which we are unable to include in the aggregated OA-level regressions.

table 2. The intuition for this approach is based on the idea of using Census characteristics and OA KS2 statistics as proxies for additional unobservable factors in the regressions of columns (4)-(8), and testing for their correlation with the changes in neighbour-peer characteristics to see if these unobservable OA factors drive neighbourhood composition. The results present a reassuring picture: nearly all the estimated coefficients are very small and insignificant.

Overall, the balancing tests in table 3 provided no evidence of strong associations between neighbour-peer changes and other neighbourhood characteristics, and provide no evidence that the near-zero neighbour-peer effect estimates in table 2 are downward biased by student or neighbourhood unobservables.

5.4 Peers at school or peers in the neighbourhood?

In the analysis conducted so far, we have not distinguished between neighbour-peers who attend the same secondary school, and those who do not. However, this distinction could be important for a number of reasons. First, children who are at school for a large part of their day may simply not interact with neighbours, unless they know each other from school already. In this case, neighbour-peers who attend a different school may exert little or no influence on students' outcomes. Secondly, distinguishing between school and neighbourhood peers is more generally useful for uncovering a 'pure' neighbourhood level peer effect, net of interactions that happen at school (i.e. school peer effects) and other school factors that have not otherwise been effectively controlled for in our regressions.

Table 4 presents evidence on this issue by tabulating results obtained from the specifications detailed in equation (3), and including different levels of fixed effects as we move from column (1) to column (4). The sample used to estimate these specifications is slightly smaller than the one used to obtain the results presented in table 2 since we drop neighbourhoods in which all students attend the same school, or all students attend different schools. Results in column (1), panel A show that neighbour-peer KS1 has an impact on a student's achievement *only* if these neighbours also attend that student's secondary school. However, in line with our previous findings, this association vanishes as soon as we include secondary-by-cohort or primary-by-secondary-by-cohort effects. Next, results in panel B, column (1) show that FSM status of neighbour-peers matters irrespective of school attended, with a standardised coefficient of negative 0.003 (s.e. 0.001). Again, as soon as we include school-by-cohort effects to control for the school-related residential sorting during the transition between primary and secondary school, the estimated effects shrink and become insignificant. Similarly, we find no evidence of neighbour-peer effects when looking at neighbours' SEN-status and gender, irrespective of the school attended.

All in all, the evidence gathered in this section rejects the hypothesis that neighbourhood peers matter differentially depending on whether they attend the same school or not. More importantly, this evidence confirms our conclusion that neighbourhood peer effects – in particular 'pure' neighbourhood peer effects, not confounded by interactions at school – do not matter for students' test score progression.

5.5 Robustness checks: intention-to-treat estimates and alternative definitions of neighbourhoods and peers

An important issue that we already flagged in both sections 3 and 4 is that, by focussing on the sample of students who stay in the same neighbourhood between grades 6 and 9, we might induce some bias due to endogenous sample-selection. To circumvent this problem, we estimate the grade-differenced specification in equation (2.1) using both 'stayers' and students who move neighbourhood between grades 6 and 9. At grade 9, we assign to these 'movers' the grade-9 characteristics of the neighbourhood in which they lived at grade 6. Stated differently, we assign them to the changes in the neighbourhood 'quality' that they would have experienced had they not moved. Estimates obtained following this approach are more properly interpreted as intention-to-treat effects. Table 5 presents the results from specifications as in equation (2.1) both without (column (1)) and with (column (2)) primary-by-secondary-by-cohort effects (both columns include our standard control variables). The new results are almost identical to those reported in table 2 for the stayers only, allaying sample-selection concerns.

As discussed in section 3.3, there are ambiguities about the correct neighbour-peer group definition. Given we cannot know *a priori* the correct grouping, we experiment in table 5 with a different group definition as discussed in section in 4.3. Columns (3) and (4) consider neighbour-peers in the same OA and grade. Finally, column (5) and (6) change the neighbourhood definition to include, on average, 6-7 adjacent OAs (on average 80 students). In general, these re-definitions make no substantive difference to the results. In some cases, previously insignificant coefficients become more precise, although all the effects remain very small in magnitude, and most are insignificant once we include school-by-cohort effects. It is worth noting that using aggregates computed over larger residential areas in column (5) *increases* the precision and the size of our estimates. However, including school-by-cohort effects as in column (8) brings our estimates close to zero and insignificant (with the exception of the changes in the share of males). This pattern might be explained by the fact that changes in larger neighbourhood aggregates are more likely to be 'contaminated' by omitted time-varying neighbourhood factors – such as changes to neighbourhood infrastructure or household mobility dictated by school quality and access – than for smaller geographical units. This lends support to our claim made earlier that, since our identification approach relies on fixed effects to control for neighbourhood unobservables, a small scale is desirable in order to minimise the risk of endogeneity of changes in neighbourhood quality.

6 Concluding Remarks

Our study has used various detailed administrative datasets on the population of students in England to study the effect of the characteristics and prior achievements of peers in the neighbourhood on the educational achievements outcomes of secondary school students. In our main sample we track over 1.3 million students across four cohorts that go through the first three years of their secondary schooling. Our findings show that, although there is a substantial cross-sectional correlation between students' test scores and the characteristics of their residential neighbourhoods, there is no evidence that this association is causal. The 'true' effect of changes in peers in the neighbourhood on students' test-score gains between grades 6 (ages 11) and 9 (age 14) is nil.

In order to assess the robustness of this conclusion, we have extended our analysis in a number of dimensions. First, we have distinguished between peers in the neighbourhood that attend the same school and those who do not. Next, we have considered alternative definitions of neighbourhoods and different ways of identifying peers in the place of residence. All in all, our evidence leads us to conclude that neighbourhood effects are a non-significant determinant of students' test score attainments in schools.

Besides presenting new evidence on the effect of peers in the neighbourhood, our study makes a number of important methodological contributions. First, we 'drill down' to the effect of neighbourhood changes that are caused by real movements of families in an out of small neighbourhoods. We can track these changes through information on the detailed residential addresses of our census of students. This is radically different from the approach used in the literature that looks at peer effects at schools, which focuses on the year-on-year changes in school composition under the maintained assumption that students only interact with peers within their grade (or class). Moreover, the English institutional setting where secondary school attendance is not tightly linked to place of residence, allows us to distinguish between neighbours who attend the same or a different school, and to test for potential interactions between school and neighbourhood peer effects. Furthermore, by exploiting the detail and density of our data, we are able to change our definitions of neighbourhoods and peers in the place of residence, and thus address the inherent problem in the literature of pinning-down the correct definition of what constitutes 'a neighbourhood'. This allows us to exclude the possibility that our findings are stemming from data-driven incorrect levels of aggregation. Finally, exploiting the fact that we observe several cohorts of students experiencing changes in the composition of their neighbourhoods at the same as they move through the education system, we are able to partial out student and family background unobservables, neighbourhood fixed effects and time trends as well as school-by-cohort unobserved shocks. We believe this is unique in getting us close to pinning down an unbiased 'neighbourhood effect' estimate stemming solely from social interactions and role models in the place of residence as originally advocated by Moffit (2001).

7 References

Allen, R., S. Burgess and T. Key (2010). Choosing Secondary School by Moving Home: School Quality and the Formation of Neighbourhood, *CMPO Working Paper*, 10/238.

Aaronson, D. (1998). Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes, *Journal of Human Resources*, pp. 915-46.

Angrist, J.D., D. Lang and P. Oreopolous (2009). Incentives and Services for College Achievement: Evidence from a Randomized Trial, *American Economic Journal: Applied Economics*, 1.

Angrist, J.D. and V. Lavy (2009). The Effect of High-Stakes High School Achievement Awards: Evidence from a Randomized Trial, *American Economic Review*, 99(4): 1384–1414.

Akerlof, G. (1997). Social Distance and Social Decisions, Econometrica, 65(5), pp. 1005-1027.

Bayer, P., L. Ross, L. and G. Topa (2008). Place of Work and Place of Residence: Informal Hiring Networks and Labour Market Outcomes, *Journal of Political Economy*, 116(6): pp. 1150-1196.

Bernheim, B. (1994). A Theory of Conformity, *The Journal of Political Economy*, 102(5), pp. 841-877.

Black, S. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education, *The Quarterly Journal of Economics*, 114(2), pp. 577-99.

Bolster A., Burgess S., Johnston R, K. Jones, C. Propper and R. Sarker (2007). Neighbourhoods, households and income dynamics: a semi-parametric investigation of neighbourhood effects, *Journal of Economic Geography*, 7(1), pp. 1-38

Card, D., and J. Rothstein (2007). Racial Segregation and the Black-White Test Score Gap, *Journal of Public Economics*, 91(11-12): 2158-2184.

Cheshire, P. and S. Sheppard (1995). On the Price of Land and the Value of Amenities, *Economica*, New Series, Vol. 62, No. 246 (May), pp. 247-267.

Cheshire, P., S. Gibbons, S. and I. Gordon (2008). Policies for 'Mixed Communities': A Critical Evaluation, *SERC Policy Paper*, 2.

Currie, J. (2006). *The Invisible Safety Net*, Princeton University Press, New Jersey, USA.

Cutler, D. and E. Glaeser (1997). Are Ghettos Good Or Bad?, *The Quarterly Journal of Economics* 112(3), pp. 827-72.

Durlauf, S. (1996). A Theory of Persistent Income Inequality, *Journal of Economic Growth*, pp. 75-93.

Edin, P., P. Fredriksson and O. Aslund (2003). Ethnic Enclaves and the Economic Success of Immigrants – Evidence from a Natural Experiment, *Quarterly Journal of Economics*, 118, pp. 329-357.

Edin, P., P. Fredriksson, H. Gronqvist and O. Aslund (2010). Peers, Neighbourhoods and Immigrant Student Achievement – Evidence from a Placement Policy, *American Economic Journal: Applied Economics*, accepted for publication.

Fehr, E. and A. Falk (2002). Psychological Foundations of Incentives, *European Economic Review*, 46(4-5), pp. 687-724.

Gibbons, S. (2002). Neighbourhood Effects on Educational Achievement, *CEE Discussion Paper*, CEEDP0018.

Gibbons, S., S. Machin and O. Silva (2008). Choice, Competition and Student Achievement, *Journal of the European Economic Association*, 6(4), pp. 912-947.

Gibbons, S., S. Machin and O. Silva (2009). Valuing School Quality Using Boundary Discontinuities, *SERC Discussion Papers*, 18.

Gibbons, S. and O. Silva (2008). Urban Density and Student Attainment, *Journal of Urban Economics*, 63(1), pp. 631-650.

Gibbons, S and S. Telhaj (2008). Peers and Achievement in England's Secondary Schools, *SERC Discussion Paper*, 1.

Gibbons, S and S. Telhaj (2011). Mobility and School Disruption, *Journal of Public Economics*, forthcoming.

Gibbons, S., Silva, O. and F. Weinhardt (2011). Everybody Needs Good Neighbours? Evidence from Students' Outcomes in England, IZA Discussion Papers 5980, Institute for the Study of Labor (IZA).

Glaeser, E. and J. Scheinkman (2001). Measuring Social Interactions, in *Social Dynamics*, Durlauf, S. and P. Young (eds.), Boston, MA: MIT Press.

Glennerster, H. (1991). Quasi-Markets for Education, *Economic Journal*, 101, pp. 1268-76.

Gould, E., V. Lavy and D. Paserman (2004). Immigrating to Opportunity: Estimating the Effect of School Quality using a Natural Experiment on Ethiopians in Israel, *Quarterly Journal of Economics*, 119(2), pp. 489-526.

Gould, E., V. Lavy and D. Paserman (2011). Sixty Years after the Magic Carpet Ride: The Long-Run Effect of the Early Childhood Environment on Social and Economic Outcomes, *Review of Economic Studies*, 78(3), pp. 938-973.

Goux, D. and E. Maurin (2007). Close Neighbours Matter: Neighbourhood Effects on Early Performance at School, *Economic Journal*, 117(523), pp. 1193-215.

Granovetter, M. (1995). *Getting a Job*, second edition, Chicago, IL: University of Chicago Press.

Harding, D, L. Gennetian, C. Winship, L. Sanbonmatsu, J. Kling (2010). Unpacking Neighborhood Influences on Education Outcomes: Setting the Stage for Future Research, *NBER Working Paper* 16055.

Hanushek, E., J. Kain, J. Markman and S. Rivkin (2003). Does Peer Ability Affect Student Achievement?, *Journal of Applied Econometrics*, 18, pp. 527-544.

Hanushek, E., J. Kain and S. Rivkin (2004). Disruption versus Tiebout improvement: the costs and benefits of switching schools, *Journal of Public Economics*, 88, pp. 1721-1746

Hoxby, C. (2000). Peer Effects in the Classroom: Learning from Gender and Race Variation, *NBER Working Paper* 7867.

Jacob, B. (2004). Public Housing, Housing Vouchers, and Student Achievement: Evidence from Public Housing Demolitions in Chicago, *The American Economic Review*, 94(1), pp. 233-58.

Kain, J. and J. Quigley (1975). The Demand for Individual Housing Attributes, in: *Housing Markets and Racial Discrimination: A Microeconomic Analysis*, pages 231-255, National Bureau of Economic Research, Inc.

Kling, J., J. Liebman and L. Katz (2007). Experimental Analysis of Neighborhood Effects, *Econometrica*, pp. 83-119.

Kling, J., J. Ludwig and L. Katz (2005). Neighborhood Effects on Crime for Female and Male Youth: Evidence From a Randomized Housing Voucher Experiment, *Quarterly Journal of Economics*, 120, pp. 87-130.

Lavy, V., D. Paserman and A. Schlosser (2008). Inside the Black Box of Ability Peer Effects: Evidence from Variation in Low Achievers in the Classroom, *NBER Working Paper* 14415.

Lavy, V., and A. Schlosser (2007). Mechanisms and Impacts of Gender Peer Effects at School, *NBER Working Paper* 13292.

Lavy, V., O. Silva and F. Weinhardt (2012). The Good, The Bad and The Average: Evidence on Ability Peer Effects in Schools, *Journal of Labor Economics*, forthcoming.

Lupton, R. (2009). Growing up in Social Housing in Britain, Report: Joseph Rowntree Foundation, York.

Manski, C. (1993). Identification of Endogenous Social Effects: The Reflection Problem, *Review of Economic Studies*, 60, pp. 531-42.

Moffitt, R. (2001). Policy interventions, Low-level Equilibria, and Social Interactions, in *Social Dynamics*, Durlauf, S. and P. Young (eds.), Boston, MA: MIT Press.

Oreopolous, P. (2003). The Long-Run Consequences of Living in a Poor Neighborhood, *Quarterly Journal of Economics*, 118, pp. 1533-1575.

Rosenbaum, J. (1995). Changing the Geography of Opportunity by Expanding Residential Choice: Lessons from the Gautreaux Program, *Housing Policy Debate*, 6, pp. 231-69.

Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates, *Quarterly Journal of Economics*, 116(2), 681-704.

Sampson, R., J. Morenoff and T. Gannon-Rowley (2002). Assessing Neighbourhood Effects: Social Processes and New Directions in Research, *Annual Review of Sociology*, 26, pp. 443-478.

Sanbonmatsu, L., J. Kling, G. Duncan and J. Brooks-Gunn (2006). Neighborhoods and academic achievement: Results from the Moving to Opportunity Experiment, *Journal of Human Resources*, 41(4), p. 649.

Weinhardt, F. (2010). Moving into the Projects: Social Housing Neighbourhoods and School Performance in England, *SERC Discussion Paper*, 44.

Tables:

Table 1: Descriptive statistics of the main dataset

Variable	Mean	Standard Deviation
Panel A: Students' characteristics, 'stayers' only		
KS2 percentiles, average English, Maths and Science	50.125	25.236
KS3 percentiles, average English, Maths and Science	51.253	25.819
KS2 to KS3 value-added	1.127	13.598
KS1 score, average English and Maths	15.122	3.611
Student is FSM eligible	0.155	0.362
Student is SEN	0.213	0.409
Student is Male	0.508	0.499
Average rate of outward mobility in n'hood over four years	0.081	0.057
Average rate inward mobility in n'hood over four years	0.083	0.062
Secondary school size (in grade 7)	1083.9	384.9
Panel B: Characteristics of students in the neighbourhood – Output Area		
KS1 score, average English and Maths – At grade 6	15.017	1.762
KS1 score, average English and Maths – At grade 9	14.981	1.760
KS1 score, average English and Maths – Change grade 6 to 9	-0.036	0.863
Share FSM – At grade 6	0.165	0.196
Share FSM – At grade 9	0.170	0.199
Share FSM – Change grade 6 to 9	0.005	0.081
Share SEN – At grade 6	0.215	0.154
Share SEN – At grade 9	0.217	0.153
Share SEN – Change grade 6 to 9	0.002	0.087
Share Male – At grade 6	0.509	0.153
Share Male – At grade 9	0.509	0.157
Share Male – Change grade 6 to 9	0.000	0.103
Number of students in OA, 'central cohort' +1/-1, grade 6	13.878	6.317
Number of students in OA, 'central cohort' +1/-1, grade 9	13.865	6.186
Number of students in OA, 'central cohort' only, grade 6	5.173	2.612
Number of students in OA, 'central cohort' only, grade 6	5.169	2.639

Note: Descriptive statistics refer to: *(i)* students who do not change Output Area (OA) of residence in any period between grade 6 and 9; *(ii)* students in OAs with at least five students belonging to the 'central cohort' +1/-1 in every period between grade 6 and grade 9; *(iii)* students in the non-selective part of the education system. These restrictions were operated after computing OA aggregate information (see panel B). Number of 'stayers': approximately 1,310,000 (evenly distributed over four cohorts). Number of OAs: approximately 134,000. Average inward mobility and outward mobility in neighbourhood refer to (cohort-specific) OA mobility rates averaged over the period grade 6 to 9. KS1 refers to the average test score in Reading, Writing and Mathematics at the Key Stage 1 exams (at age 7); FSM: free school meal eligibility; SEN: special education needs (with and without statements). Secondary school type attended in grade 7: 66.7% Community; 14.9% Voluntary Aided; 3.1% Voluntary Controlled; 14.5% Foundation; 0.3% Technology College; 0.5% City Academy.

Table 2: Characteristics of young peers in the neighbourhood: the effect on students' achievements

	Dependent Variable/Timing is:							
	No controls				With controls			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	KS3/ Grade 9	KS3-KS2/ Grade 6 to 9	KS3-KS2/ Grade 6 to 9	KS3-KS2/ Grade 6 to 9	KS3/ Grade9	KS3-KS2/ Grade 6 to 9	KS3-KS2/ Grade 6 to 9	KS3-KS2/ Grade 6 to 9
Panel A: N'hood Average KS1								
KS1 score – Level (Grade 9) or Change (Grade 6 or 9)	0.279	0.003	-0.000	0.001	0.079	0.003	-0.000	-0.000
	(0.001)**	(0.001)**	(0.001)	(0.001)	(0.001)**	(0.001)**	(0.001)	(0.001)
Panel B: N'hood Share of FSM								
Share FSM – Level (Grade 9) or Change (Grade 6 or 9)	-0.289	-0.005	-0.001	0.001	-0.101	-0.005	-0.001	0.001
	(0.001)**	(0.001)**	(0.001)	(0.001)	(0.001)**	(0.001)**	(0.001)	(0.001)
Panel C: N'hood Share of SEN								
Share SEN – Level (Grade 9) or Change (Grade 6 or 9)	-0.191	-0.002	-0.000	-0.001	-0.055	-0.002	-0.001	-0.001
	(0.001)**	(0.002)	(0.001)	(0.001)	(0.001)**	(0.001)*	(0.001)	(0.001)
Panel D: N'hood Share of Males								
Share Males – Level (Grade 9) or Change (Grade 6 or 9)	0.004	0.001	0.001	0.002	0.000	0.001	0.001	0.002
	(0.001)**	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Controls	No	No	No	No	Yes	Yes	Yes	Yes
Secondary by Cohort FX	No	No	No	Yes	No	No	No	Yes
Second. by Primary by Cohort FX	No	No	Yes	No	No	No	Yes	No
OA FX (trends)	No	No	No	Yes	No	No	No	Yes

Note: Table reports standardised coefficients and standard errors. Number of observations ~1,310,000 in ~134,000 Output Areas. All regressions include cohort dummies. Controls include: student own KS1 test scores; student is FMSE; student is SEN; student is male; school size (refers to school attended in grade 7); school type dummies (refers to school attended in grade 7 and includes: Community, Voluntary Aided, Voluntary Controlled, Foundation, CTC and Academy); average rates of outward and inward mobility in n'hood over four years. Secondary by cohort effects: 12,273 groups (refer to school at grade 7 when student enters secondary education). Secondary by primary by cohort school effects: 191,245 groups. OA effects (trends): 134,000 groups. Standard errors clustered at the OA level in round parenthesis. **: 1% significant or better; *: at least 5% significant.

Table 3: Balancing of changes in neighbourhood characteristics

	Treatment is:			
	(1)	(2)	(3)	(4)
Dependent Variable is:	KS1 score – Change, grade 6 to 9	Share FSM – Change, grade 6 to 9	Share SEN – Change, grade 6 to 9	Share Male Change, grade 6 to 9
Panel A: Individual Characteristics (unconditional)				
KS1 score, average English and Maths	0.007	-0.019	-0.006	-0.001
	(0.004)	(0.004)**	(0.004)	(0.003)
Student is FSM eligible	0.000	0.026	-0.006	0.003
	(0.004)	(0.004)**	(0.004)	(0.003)
Student is SEN	-0.000	0.008	-0.005	0.002
	(0.004)	(0.004)*	(0.003)	(0.003)
Student is Male	-0.004	0.005	-0.002	0.009
	(0.004)	(0.004)	(0.004)	(0.004)*
Panel B: Neighbourhood Characteristics (conditional on controls)				
Average KS2 of students living in OA	0.005	-0.004	-0.004	-0.004
(PLASC/NPD)	(0.002)*	(0.002)	(0.003)	(0.002)*
Std.Dev. of KS2 across students living in OA	-0.000	0.001	-0.002	-0.003
(PLASC/NPD)	(0.004)	(0.004)	(0.004)	(0.004)
Share of households living in socially rented	0.002	0.002	-0.003	0.000
accommodation (Census 2001)	(0.002)	(0.003)	(0.002)	(0.002)
Share of households owning place of residence	-0.002	-0.002	0.002	0.001
(Census 2001)	(0.002)	(0.003)	(0.002)	(0.002)
Share of adults in employment	0.003	0.002	-0.001	-0.003
(Census 2001)	(0.003)	(0.003)	(0.003)	(0.002)
Share of adults with no educational qualifications (Census 2001)	0.004	-0.001	0.001	0.002
	(0.003)	(0.003)	(0.003)	(0.002)
Share of lone parents in the population (Census 2001)	-0.001	-0.003	0.001	0.000
	(0.002)	(0.003)	(0.002)	(0.002)

Note: Table reports standardised coefficients and standard errors from regressions of one of the dependent variables (first column) on each of the treatments separately. Census characteristics recorded at the OA level in 2001. All other data was collapsed at the OA level and the regression analysis was performed at this level. Number of observations: approximately 134,000. Regressions in the top panel only control for cohort effects and school-type effects (refers to school attended in grade 7). Regressions in the bottom panel include cohort effects, OA-averaged student KS1 test scores; OA-averaged student eligibility for FMSE; OA-averaged student SEN status; OA-averaged student male gender; OA-averaged school size (refers to school attended in grade 7); school-type effects (refers to school attended in grade 7); OA-averaged rates of outward and inward mobility in neighbourhood. Standard errors clustered at the OA level in round parenthesis. **: 1% significant or better. *: at least 5% significant.

Table 4: The impact of neighbourhood peers attending the same/different school

	Dependent Variable/Timing is: KS3-KS2 value-added/Grade 6 to 9		
	(1)	(2)	(3)
Panel A: N'hood Average KS1			
KS1 score – Same school	0.003	0.001	0.001
Change, Grade 6 to 9	(0.001)*	(0.001)	(0.001)
KS1 score – Other school	0.001	-0.001	0.000
Change, Grade 6 to 9	(0.001)	(0.001)	(0.001)
Panel B: N'hood Share of FSM			
Share FSM – Same school	-0.003	-0.001	-0.001
Change, Grade 6 to 9	(0.001)**	(0.001)	(0.001)
Share FSM – Other school	-0.003	0.000	-0.001
Change, Grade 6 to 9	(0.001)**	(0.001)	(0.001)
Panel C: N'hood Share of SEN			
Share SEN – Same school	-0.001	0.000	-0.001
Change, Grade 6 to 9	(0.001)	(0.001)	(0.001)
Share SEN – Other school	-0.002	-0.001	-0.001
Change, Grade 6 to 9	(0.002)	(0.001)	(0.001)
Panel D: N'hood Share of Males			
Share Male – Same school	0.000	0.001	0.001
Change, Grade 6 to 9	(0.001)	(0.001)	(0.001)
Share Male – Other school	0.000	0.001	-0.001
Change, Grade 6 to 9	(0.001)	(0.001)	(0.001)
Controls	Yes	Yes	Yes
Secondary × Cohort FX	No	Yes	No
Second. × Prim. × Cohort FX	No	No	Yes

Note: Table reports standardised coefficients and standard errors. Number of observations approximately 970,000 in approximately 122,000 Output Areas. The smaller sample size and number of Output Areas is driven by the restriction that Output Areas must have both a subset of students going to the same school and a subset of students going to different schools. All regressions include cohort dummies. Controls include: student own KS1 test scores; student is FMSE; student is SEN; student is male; school size (refers to school attended in grade 7); average rate of outward mobility in neighbourhood over four years; average rate inward mobility in neighbourhood over four years.. Secondary by cohort effects: approximately 12,000 groups. Secondary by primary by cohort school effects: 134,000 groups. Standard errors clustered at the OA level in round parenthesis. **: 1% significant or better; *: at least 5% significant.

Table 5: Robustness to alternative estimation samples and peer-group definition

	Dependent Variable/Timing is:					
	Movers 'ITT' set-up		'Central cohort' only		Adjacent OA n'hoods' only	
	(1)	(2)	(3)	(4)	(6)	(7)
	KS3-KS2/ Grade 6 - 9	KS3-KS2/ Grade 6 to 9	KS3-KS2/ Grade 6 to 9	KS3-KS2/ Grade 6 to 9	KS3-KS2/ Grade 6 to 9	KS3-KS2/ Grade 6 to 9
KS1 score –	0.003	0.000	0.001	-0.000	0.005	-0.001
Change (Grade 6 or 9)	(0.001)**	(0.001)	(0.001)	(0.001)	(0.001)**	(0.001)
Share FSM –	-0.005	-0.001	-0.003	-0.001	-0.003	0.001
Change (Grade 6 or 9)	(0.001)**	(0.001)	(0.001)**	(0.001)	(0.001)**	(0.001)
Share SEN –	-0.002	-0.001	-0.001	-0.000	-0.004	-0.000
Change (Grade 6 or 9)	(0.001)*	(0.001)	(0.001)	(0.001)	(0.001)**	(0.001)
Share Males –	0.001	0.001	0.000	0.001	0.002	0.002
Change (Grade 6 or 9)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)*	(0.001)*
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Second. * Primary * Cohort FX	No	Yes	No	Yes	No	Yes

Note: Table reports standardised coefficients and standard errors. Number of observations approximately 1,310,000 in approximately 134,000 Output Areas. All regressions include cohort dummies. Controls include: student own KS1 test scores; student is FMSE; student is SEN; student is male; school size (refers to school attended in grade 7); school type dummies (refers to school attended in grade 7 and includes: Community, Voluntary Aided, Voluntary Controlled, Foundation, CTC and Academy); average rate of outward mobility in n'hood over four years; average rate inward mobility in n'hood over four years. Secondary by primary by cohort effects: 191,245 groups. Standard errors clustered at the OA level in round parenthesis. **: 1% significant or better; *: at least 5% significant.

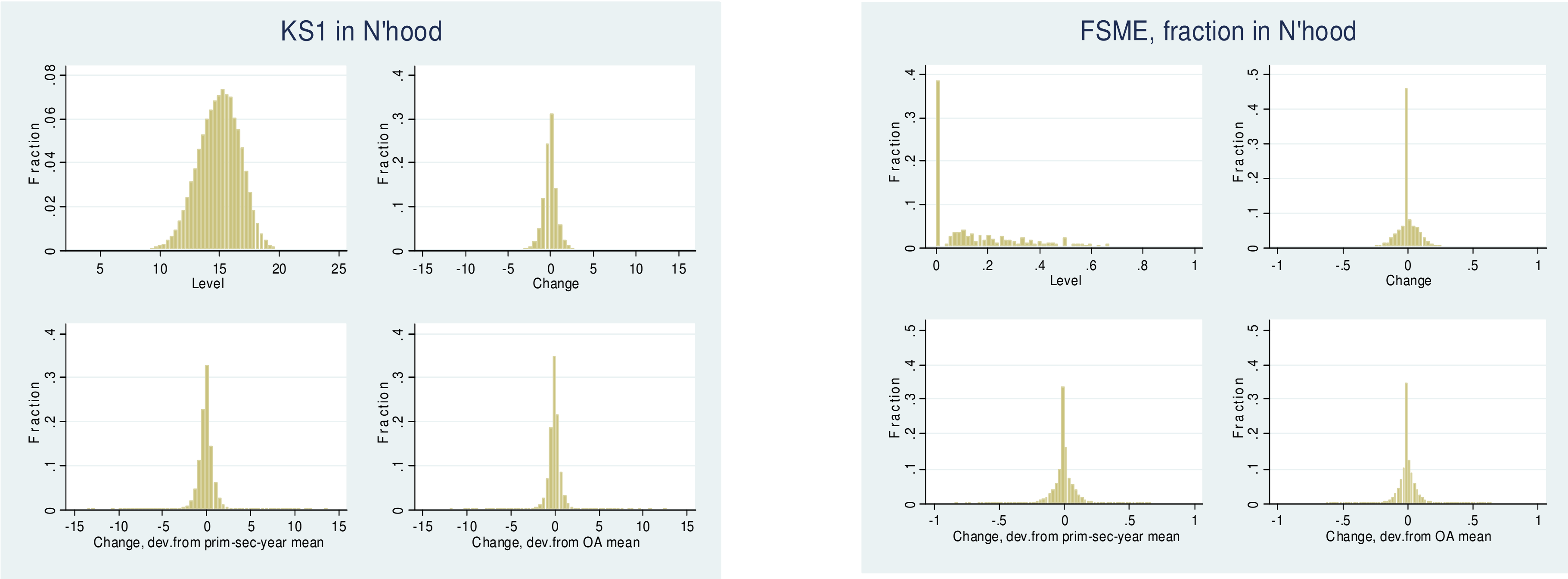
Figures:

Figure 1: Main dataset construction; four 'central cohorts' and adjacent cohorts

	PLASC 2002	PLASC 2003	PLASC 2004	PLASC 2005	PLASC 2006	PLASC 2007	PLASC 2008
				Grade 5			Grade 8
Cohort 4			Grade 5	Grade 6/KS2		Grade 8	Grade 9/KS3
Cohort 3		Grade 5	Grade 6/KS2	Grade 7	Grade 8	Grade 9/KS3	Grade 10
Cohort 2	Grade 5	Grade 6/KS2	Grade 7	Grade 8	Grade 9/KS3	Grade 10	
Cohort 1	Grade 6/KS2	Grade 7		Grade 9/KS3	Grade 10		
	Grade 7			Grade 10			

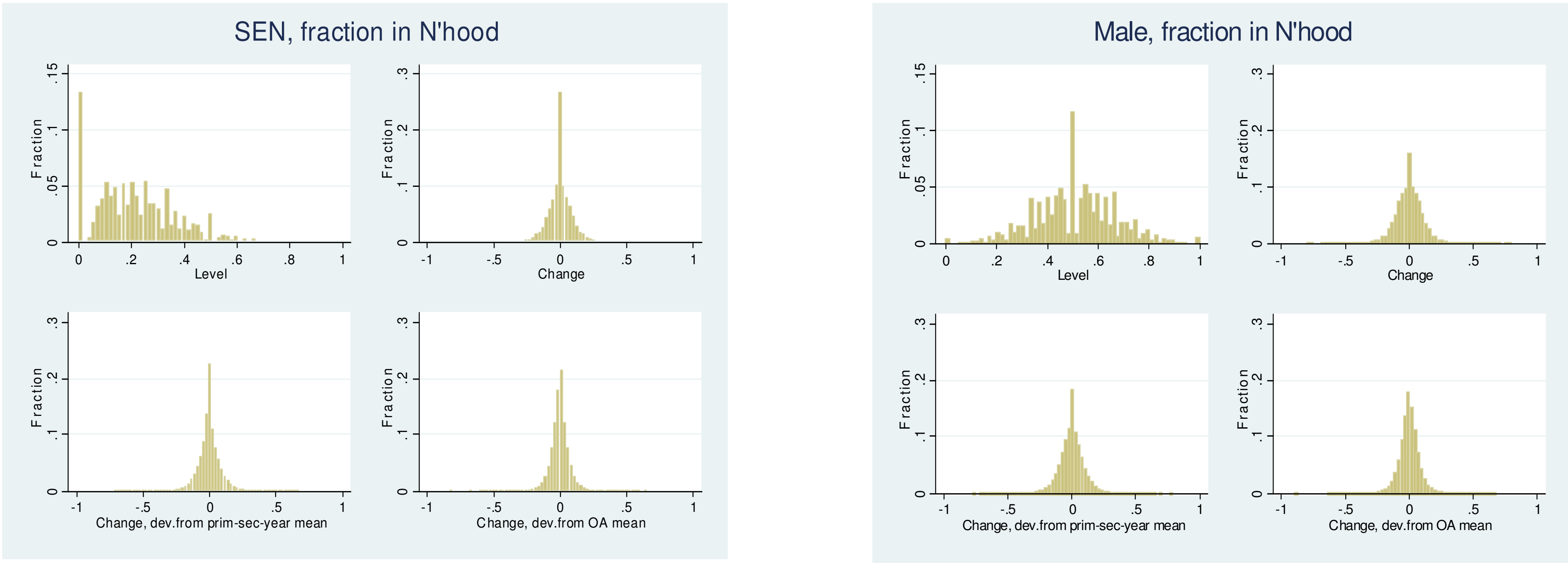
Note: Shaded cells refer to the estimation sample; immediately adjacent non-shaded cohorts represent the additional set of students used to construct measures of quality of neighbourhood. PLASC refers to the Student Level Annual School Census. Students finish their primary school in grade 6 when they sit for their Key Stage 2 (KS2) at age 11. Thick border indicates end of primary school. Students enter secondary education in grade 7 and complete their Key Stage 3 exams in grade 9 when aged 14.

Figure 2a: Characteristics of students in the neighbourhood and amount of variation: prior achievements (KS1) and free school meal eligibility (FSM)



Note: Descriptive statistics of deviations from primary-by-secondary-by-cohort mean changes are as follows. Average KS1, mean 0.000; std.dev. 0.778. Fraction of FSM students: mean 0.000, std.dev. 0.073. Descriptive statistics of deviations from Output Area mean changes as follows. Average KS1, mean 0.000; std.dev. 0.632. Fraction of FSM students: mean 0.000, std.dev. 0.061. Descriptive statistics for the level and change in these variables are reported in table 1, panel B.

Figure 2b: Characteristics of students in the neighbourhood and amount of variation: special education needs (SEN) and share of male students



Note: Descriptive statistics of deviations from primary-by-secondary-by-cohort mean changes are as follows. Fraction of SEN students: mean 0.000, std.dev. 0.078. Fraction of Male students: mean 0.000, std.dev. 0.093. Descriptive statistics of deviations from Output Area mean changes as follows. Fraction of SEN students: mean 0.000, std.dev. 0.065. Fraction of male students: mean 0.000, std.dev. 0.076. Descriptive statistics for the level and change in these variables are reported in table 1, panel B.

Appendix

Appendix Table 1:

Descriptive statistics before dropping mobile students and small n'hoods

Variable	Mean	Standard Deviation
Panel A: Students' characteristics, 'stayers' only		
KS2 percentiles, average English, Maths and Science	50.207	25.915
KS3 percentiles, average English, Maths and Science	49.308	25.251
KS2 to KS3 value-added	0.898	13.770
KS1 score, average English and Maths	15.004	3.647
Student is FSM eligible	0.171	0.377
Student is SEN	0.220	0.414
Student is Male	0.507	0.500
Average rate of outward mobility in n'hood over four years	0.098	0.075
Average rate inward mobility in n'hood over four years	0.089	0.073
Secondary school size (in grade 7)	1081.6	385.0
Panel B: Characteristics of students in the neighbourhood – Output Area		
KS1 score, average English and Maths – At grade 6	14.968	1.857
KS1 score, average English and Maths – At grade 9	14.966	1.854
KS1 score, average English and Maths – Change grade 6 to 9	-0.002	1.407
Share FSM eligible – At grade 6	0.172	0.205
Share FSM eligible – At grade 9	0.172	0.206
Share FSM eligible – Change grade 6 to 9	-0.001	0.140
Share SEN – At grade 6	0.218	0.166
Share SEN – At grade 9	0.218	0.166
Share SEN – Change grade 6 to 9	0.000	0.139
Share Male – At grade 6	0.509	0.174
Share Male – At grade 9	0.509	0.176
Share Male – Change grade 6 to 9	0.000	0.128
Number of students in Output Area, 'central cohort' +1/-1, grade 6	13.212	6.562
Number of students in Output Area, 'central cohort' +1/-1, grade 9	12.884	6.628

Note: Descriptive statistics refer to students in the non-selective part of the education system. The data *includes (i)* students who change OA of residence between grade 6 and 9; and *(ii)* students in Output Areas with less than five students belonging to the 'central cohort' +1/-1 in every period between grade 6 and grade 9. Number of observations: approximately 1,850,000, almost evenly distributed over four cohorts. Number of Output Areas: approximately 158,000. Secondary school type attended in grade 7: 66.6% Community; 14.9% Voluntary Aided; 3.1% Voluntary Controlled; 14.5% Foundation; 0.4% Technology College; 0.5% City Academy. See note to table 1 for further details on the variables.

CHAPTER IV

EVIDENCE ON THE IMPORTANCE OF PEERS AT SCHOOL

1 Introduction[27]

The estimation of peer effects in the classroom and at school has received intense attention in recent years. Several studies have presented convincing evidence about race, gender and immigrants' peer effects[28], but important questions about the scale and nature (i.e. the 'origins') of *ability* peer effects in schools remain open, with little conclusive evidence.[29] In this paper we[30] study ability peer effects in educational outcomes between schoolmates in secondary schools in England. Our aims are both to investigate the size (i.e. the 'scale') of ability peer effects on the outcomes of secondary school students, and to explore which segments of the ability distribution of peers drive the impact of peer quality on pupils' achievements (i.e. the 'nature'). In particular, we study whether the extreme tails of the ability distribution of peers – namely the exceptionally low- and high-achievers – as opposed to the average peer quality drive any significant peer effect on the outcomes of other students.

To do so, we use data for all secondary schools in England for four cohorts of age-14 (9th grade) pupils entering secondary school in the academic years 2001/2002 to 2004/2005 and taking their age-14 national tests in 2003/2004-2006/2007. We link this information to data on pupils' prior achievement at age-11, when they took their end-of-primary education national tests, which we exploit to obtain pre-determined proxy measures of peer ability in secondary schools. In particular, we construct measures of average peer quality based on pupils' age-11 achievements, as well as proxies for the very high- and very low-achievers, obtained by identifying pupils who are in the highest or lowest 5% of the (cohort-specific) national distribution of cognitive achievement at age-11. The way in which we measure peer ability is a major

[27] This chapter presents work undertaken in the preparation for a more extensive joint paper titled: "THE GOOD, THE BAD AND THE AVERAGE: EVIDENCE ON ABILITY PEER EFFECTS IN SCHOOLS", which is currently published as NBER working paper Lavy *et al.* (2009), and a revised version is accepted for publication by the Journal of Labor Economics.

[28] Recent examples include Angrist and Lang (2004) on peer effects through racial integration; Hoxby (2000) and Lavy and Schlosser (2007) on gender peer effect; and Gould et al. (2009a) on the effect of immigrants on native students.

[29] One exception is Sacerdote (2001), who presents evidence on ability peer effects in college based on co-residence of randomly paired roommates in university housing.

[30] Since this is joing work I will again use plural pronouns to refer to the authors.

improvement over previous studies. The vast majority of previous empirical evidence on ability peer effects in schools arises from studies that examine the effect of average background characteristics, such as parental schooling, race and ethnicity on students' outcomes (e.g. Hoxby, 2000 for the US, and Ammermueller and Pischke, 2009 for several European countries). A limitation of these studies is that they do not directly measure the academic ability of students' peers, but rely on socio-economic background characteristics as proxies for this. Additionally, our measures of peer quality are immune to refection problems (Manski, 1993) for two reasons. First, we identify peers' quality based on pupils' test scores at the end of primary education, before students change school and move on to the secondary phase. As a consequence of the large reshuffling of pupils in England during this transition, on average secondary school students meet 87% new peers at secondary school, i.e. students that do not come from the same primary. Secondly and crucially, we are able to track pupils during this transition, which means that we can single out new peers from old peers, and construct peer quality measures separately for these two groups. In these respects, our strategy follows Gibbons and Telhaj (2008), also on English secondary schools. In our analysis, we focus on the effect of new peers' ability on pupil achievement (controlling for old peers' quality), thus by-passing reflection problems.[31]

Our results show that a large fraction of 'bad' peers at school as identified by students in the bottom 5% of the ability distribution negatively and significantly affect the cognitive performance of other schoolmates. Importantly, we find that it is only the very bottom 5% students that (negatively) matter, and not 'bad' peers in other parts of the ability distribution. On the other hand, we uncover little evidence that the average peer quality and the share of very 'good' peers as identified by students in the top 5% of the ability distribution affect the educational outcomes of other pupils. However, these findings mask a significant degree of heterogeneity along the gender dimension. Indeed, we show that girls significantly benefit from interactions with very bright peers, and the more so if they are in the bottom half of the ability distribution. In

[31] Note that this does not imply that we are able to separate endogenous from exogenous peer effects (see Manksi, 1993; Moffitt, 2001). We see this as a further and separate issue from reflection problems that arise from previous/simultaneous interactions among students that affect measures of peers' ability (Sacerdote, 2001).

marked contrast, boys are negatively affected by a larger proportion of academically outstanding peers at school, with this adverse effect being more evident for male students in the top part of the ability distribution. On the other hand, we find that the negative effect of the very weak students does not significantly vary by the ability of regular students, nor along the gender dimension. Finally, the effect of the average peer quality on pupil cognitive achievement is estimated to be zero for boys and girls, and for students of different abilities. Although we cannot pin down the exact mechanisms that give rise to these peer effects, we draw on some theoretical explanations and related evidence put forward by the economics literature (e.g. Lazear, 2001; Hoxby and Weingarth, 2005; Jackson, 2009), as well as by psychological and educational research (e.g. Cross and Madson, 1997; Eagly, 1978; Marsh, 2005) to rationalize to our findings.

Besides providing some novel insights about the nature of ability peer effects, our paper presents a new identification approach that allows us to improve on the (non-experimental) literature in the field and to identify the effects of peers' ability while avoiding biases due to endogenous selection and sorting of pupils, or omitted variables issues. Indeed, the distribution of pupils' characteristics in secondary schools in England, like in many other countries, reflects a high degree of sorting and selection by ability. For example, using pupils' age-11 nationally standardized test scores as an indicator of ability we find that the average ability of peers and pupil's own ability in secondary school are highly correlated. This is so despite the fact that most students change school when moving from primary to secondary education and that on average pupils meet 87% new peers. Similarly, there is a high correlation between pupils' and their peers' socioeconomic background characteristics, which is further evidence that students are not randomly assigned to secondary schools and that the very top and very low achievers are typically clustered in high- and low-achieving schools. More surprisingly, these correlations survive even when we look at the within-secondary-school variation over time of pupils' and their peers' ability (i.e. conditional on secondary school fixed-effects)[32]. This suggests that some sorting/selection might be taking place, with pupils and schools being affected by and/or responding to cohort-specific unobserved shocks to students' and schools' quality. Identification strategies

[32] A similar result is documented by Gibbons and Telhaj (2008) and Black et al. (2009).

that rely on the randomness of peers' quality variation within-schools over time find little justification against this background.

In order to overcome this selection problem, we rely on within-pupil regressions (i.e. specifications including pupil fixed-effects) that exploit variation in achievements across the three compulsory subjects (English, Mathematics and Science) tested at age-14. We further exploit the fact that students were tested on the same three subjects at age-11 (at the end of primary schools), so that we can measure peers' ability separately by subject. We then study whether subject-to-subject variation in outcomes for the same student is systematically associated with the subject-to-subject variation in peers' ability. To the best of our knowledge, we are the first to use pupil fixed-effects and inter-subject differences in achievement to address identification issues of peer effects in schools.[33]

One significant advantage of this approach is that by including pupil fixed-effects we are able to control for pupil own unobservable average ability across the three subjects, as well as for unmeasured family background characteristics. Additionally, we can partial out in a highly flexible way school-by-cohort fixed-effects and other more general cohort-specific unobserved shocks that might affect pupils' outcomes and peers' quality similarly across the three subjects. This seems particularly important given the evidence of year-on-year secondary school sorting highlighted here above. On the other hand, one potential threat to our identification strategy is the possibility that sorting occurs along the lines of subject-specific abilities, so that within-student across-subject variation in ability is correlated with the variation in peers' ability across subjects. However, as we shall see below, conditional on pupil fixed-effects, our results are virtually identical irrespective of whether or not we control for pupils' own age-11 test scores, a proxy for students' subject-specific prior academic ability. This is because there is neither a sizeable nor a significant correlation between the within-student across-subject variation in age-11 achievements, and the variation in peers' ability across subjects. This suggests that specifications that include pupil fixed-effects

[33] Lavy (2009) uses the same approach to investigate the effect of instructional time on academic achievements, while Bandiera et al. (2009) use within-student across-subjects variation to study class size effects at university and Bandiera et al. (forthcoming) exploit within-worker over-time variation to analyse social incentives at work.

effectively take care of most of the sorting of pupils and their peers into secondary education, and provide reliable causal estimates of ability peer effects. To further support this claim, we provide an extensive battery of robustness checks and falsification exercises that lend additional credibility to the causal interpretation of our results.

The rest of the paper is organized as follows. The next section reviews the recent literature on peer effects, while section 3 describes the identification strategy. Section 4 describes the institutional background and our dataset. Section 5 reports our main estimates and robustness checks, while section 6 presents some heterogeneity in our findings. Finally, section 7 provides some concluding remarks.

2 Related literature

For a long time social scientists have been interested in understanding and measuring the effects of peers' behavior and characteristics on individual outcomes, both empirically (e.g. Coleman, 1966) and theoretically (e.g. Becker, 1974). The basic idea is that group actions or attributes might influence individual decisions and outcomes, such as educational attainment. Despite its intuitiveness, the estimation of peer effects is fraught with difficulties and many of the related identification issues have yet to find a definitive answer. In particular, Manski (1993) highlights the perils of endogenous group selection and the difficulty of distinguishing between contextual and endogenous peer effects. In practice, most studies have ignored this distinction and focused on reduced form estimation as outlined by Moffit (2001), where peer group characteristics are used to explain differences in individual outcomes. Even then, the literature has had to by-pass a variety of biases that arise because of endogenous sorting or omitted variables and has not yet reached a consensus regarding the size and importance even of these reduced form effects.

In particular, two main issues have taxed researchers interested in the identification of the causal effect of peer quality in education. Firstly, it is widely recognized that a pupil's peer group is evidently self-selected and hence the quality of peers is not exogenous to pupil's own quality and characteristics.[34] Failing to control for all observable and unobservable factors that determine individual sorting and achievements would result in biased estimates of ability peer effects. Secondly, peer effects work in both directions, so that peer achievements are endogenous to one pupils' own quality if students have been together for a while. This mechanical issue, known as the 'reflection problem', is particularly difficult to undo unless the researcher is able to reshuffle group formation and belonging and measure peers' quality in ways that are predetermined to interactions within the group.

To account for these difficulties, recent years have seen a variety of identification strategies. Different studies have exploited random group assignments (Sacerdote, 2001; Zimmerman, 2003; Duflo et al., 2008; De Giorgi et al., 2009; Gould et al., 2009b), within-school random variation (Hoxby, 2000; Hanushek et al., 2003; Ammermueller and Pischke, 2009; Gould et al, 2009a), instrumental variables (Goux and Maurin, 2007) or sub-group re-assignments (Katz et al., 2001; Sanbonmatsu et al., 2006).[35] Only recently, Lavy and Schlosser (2007), Lavy et al. (2008) and Duflo et al. (2008) have tried to enter the 'black box' of ability peer effects in Israel and Kenya, respectively, and have explicitly focused on understanding the mechanisms through which interactions could exert their effects. Duflo et al. (2008) exploits random assignment of pupils in primary schools in Kenya to classes by ability in order to identify peer effects. The authors find improvements from ability-tracking in primary schools and attribute this result to the fact that more homogeneous groups of students might be taught more effectively. Lavy et al. (2008) present related evidence of significant and negative effect of a high fraction of low ability students in the class (repeaters) on the outcomes of other pupils, which might arise through classroom disruption and decrease in attention paid by the teacher.

The study that is closest to ours in terms of context and data is Gibbons and Telhaj (2008) who also estimate peer effects for pupils in English secondary schools. The

---

[34] There is a well-established literature on the link between school quality and house prices (e.g. Black, 1999; Gibbons et al., 2009; Kane et al., 2006), suggesting that pupils are segregated into different neighborhoods and schools by socio-economic status.

[35] Other examples include: Aizer (2008), Bifulco et al. (2008), Burke and Sass (2008), Carrell and Hoekstra (2008), Figlio (2007), Lefgren (2004), Nechyba and Vidgor (2007) and Vidgor and Nechyba (2004).

authors attempt to control for the endogenous sorting of pupils to secondary schools by allowing for primary and secondary school fixed-effect interactions and trends. However, this approach does not fully eliminate the correlation between pupils' own ability and peer quality, and their results provide little evidence of sizeable and significant peer effects.

To the best of our knowledge, our study is the first one to rely on pupil fixed-effects and inter-subject differences in achievement to address identification issues of peer effects in schools. A similar approach has been previously used in Lavy (2009) to investigate the effect of instructional time on academic achievements; Bandiera et al. (2009) to study class size effects at university; Bandiera et al. (forthcoming) to analyse social incentives at work; and Dee (2007) to study the effect of teacher gender on students' attainments. As already mentioned, the within-student approach allows to control for pupil unobservable average ability, unmeasured family background characteristics, school-by-cohort fixed-effects and other more general cohort-specific shocks that are common to the three subjects. We believe this approach helps us to achieve a clean identification of the causal effect of peers' ability. In the next section we spell out in more details our empirical strategy.

3 Empirical strategy

3.1 General identification strategy: within-pupil regressions

The main problem with identifying the effect of the ability composition of peers on pupil educational achievements is that peer quality measures are usually confounded by the effects of unobserved correlated factors that affect students' outcomes. This correlation could arise if there is selection and sorting of students across schools based on ability differences, or if there is a relation between average students' ability in one school and other characteristics of that school (not fully observed) that might affect students' outcomes. The approach commonly used in several recent studies relies on within-school variations in the ability distribution of students across adjacent cohorts or across different classes (e.g. Ammermueller and Pischke, 2009; Hoxby, 2000; Gibbons and Telhaj, 2008; Gould et al., 2009a; Lavy et al., 2008; Lavy and Schlosser, 2007). This method potentially avoids both sources of confounding factors, although the identifying assumption is that the variation of peer quality over time (or across classes) is idiosyncratic and uncorrelated with students' potential outcomes and background.

In this paper, we suggest an alternative approach for overcoming the potential selection/sorting and omitted variable biases, namely we examine subject-to-subject variation in outcomes for the same student and investigate if this is systematically associated with the subject-to-subject variation in peers' ability. The ability peer effects that we study here are therefore subject-specific. Stated differently, in this paper we question whether pupils who have school peers that have on average higher ability in subject *j* (e.g. Mathematics) than in subject *i* (e.g. Science), have better cognitive performance in subject *j* than in subject *i*.

More formally, using test scores in multiple subjects and four cohorts of 9th graders taking their age-14 national tests in the academic years 2003/2004-2006/2007, we estimate the following pupil fixed-effect equation:

$$A_{iqst} = \alpha_i + \beta_q + \gamma_{st} + \beta_q \times Gender_i + \delta_1 P_{qst} + \delta_2 P^h_{qst} + \delta_3 P^l_{qst} + \varepsilon_{iqst} \quad (1)$$

Where i denotes pupils, q denotes subjects (English, Mathematics and Science), s denotes schools and t denotes pupils' cohort. A_{iqst} is an achievement measure for student i in subject q at school s in cohort t. In our analysis, we focus on test scores in the three compulsory subjects (English, Mathematics and Science) assessed at age-14 during the national tests; these are denoted in England as Key Stage 3 (KS3; more details are presented in section 4). Additionally, a_i is a student fixed-effect, β_q is a subject-specific effect, and γ_{st} is a school × cohort effect. We also include an interaction term between pupil's gender and subject-specific effects which is meant to control for the well-documented gender disparities in achievements in different subjects (see Ellison and Swanson, 2009, and Fryer and Levitt, forthcoming), and for the effects that these might have on pupils' and their peers' sorting into secondary schooling.[36] Next, P_{qst} captures the average ability of peers in subject q in secondary school s in cohort t as measured by test scores in a given subject in the national tests taken by students at age-

[36] We also tried specifications where we interact other pupil characteristics (e.g. eligibility for free school meals) with subject-specific dummies, and found virtually identical results. However, we prefer the more parsimonious specification in equation (1).

11 at the end of primary school (denoted as Key Stage 2, or KS2). On the other hand, P^h_{qst} and P^l_{qst} capture the fraction of very high-ability and the very low-ability peers in one students' cohort. More precisely, we choose the top and bottom 5% in the (cohort-specific) national distribution of KS2 test scores as the cut off points to determine P^h_{qst} and P^l_{qst} (more details in the data section). Finally, ε_{iqst} is an error term, which is composed of a pupil-specific random element that allows for any type of correlation within observations of the same student and of the same school.

The coefficients of interest are δ_1, which captures the effect of the average ability of peers on students' achievement; δ_2, which measures the effect of the proportion of peers in the cohort who are in the top 5% of the national distribution of KS2 test scores; and δ_3, which identifies the effect of the fraction of students who are in the bottom 5%. As discussed above, we are interested in the relative strength and significance of these three coefficients to determine which segments of the peer ability distribution drive any ability peer effect that we will document.

Note that one significant advantage of this approach is that pupil fixed-effects 'absorb' students' own unobservable average ability across subjects as well as unmeasured family background characteristics. Moreover, this specification allows to partial out in a very flexible way school-by-cohort fixed-effects (e.g. unobserved changes in school resources or head teacher), and other more general cohort-specific unobserved shocks (e.g. changes in the quality of primary schooling or in the quality of childcare facilities) that might affect pupils' outcomes and peers' quality similarly across the three subjects. This seems particularly important given the issues discussed in Arcidiacono et al. (2009) and given that, as highlighted in the Introduction, we find evidence of a significant correlation between pupils' characteristics and ability, and the characteristics and ability of their peers even conditional on secondary school fixed-effects. This suggests that some form of parental sorting based on school-by-year specific considerations might be taking place, or that cohort-specific shocks to pupil and school quality might have occurred.

Before moving on, three remarks are worth being made. First, one necessary assumption for our identification strategy is that peer effects are the same for all three subjects; stated differently, we cannot interact the δ parameters with β_q in equation (1).

Although this restriction does not seem untenable, in the analysis that follows we will provide some evidence to support this conjecture. Second, our peer effects are 'net' measures of peer influences, that is net of ability spillovers across subjects (e.g. peers' ability in English might influence pupils' test scores in Mathematics). If spillovers are very strong such that subject-specific abilities do not matter, then we are bound to find zero peer effects. Third, results are broadly unaffected if we use the absolute number of very weak and very good peers instead of their proportion.

3.2 Dealing with potential threats to identification

Although the strategy described so far allows us to effectively control for pupils' average ability across subjects, unobservable family background characteristics and school-by-cohort unobservable shocks, this setup does not preclude the possibility that selection and sorting of students in different schools is partly based on subject-specific ability and considerations. In particular, there might be some residual correlation between the within-student across-subject variation in age-11 prior achievements, capturing students' subject-specific abilities, and the variation in peers' quality across subjects.

Our main approach to account for such potential sorting is to control for pupils' KS2 test scores in the within-pupil estimation. The underlying assumption is that the lagged test scores effectively capture any subject-specific abilities, and therefore within-subject peer assignment is as good as random conditional on primary school test scores. Stated differently, there is no sorting based on other unobserved factors that are not correlated with KS2 scores. To our advantage, we can control for lagged test scores in a very flexible way by including in our specification at the same time *same*-subject lagged test scores (e.g. looking at KS3 English test score for pupil i controlling for his/her age-11 English achievement), as well as *cross*-subject test scores (e.g. looking at pupil i's age-14 English test score controlling for his/her age-11 attainments in Mathematics and Science). This allows us to partial out the effect of one pupil's own ability in a specific subject, as well as any cross-subject effects. Additionally, we can interact lagged test scores with subject-specific dummies, so that age-11 achievements can exhibit different effects on age-14 outcomes in different subjects. Under our most flexible (and preferred) specification, we estimate the following model:

$$A_{iqst} = \alpha_i + \beta_q + \gamma_{st} + \beta_q \times Gender + \delta_1 P_{qst} + \delta P^h_{qst} + \delta P^l_{qst} + \lambda_q a_{iqst} + \theta_q a_{iq(-1)st} + \kappa_q a_{iq(-2)st} + \varepsilon_{iqst} \quad (2)$$

where now a_{iqst} represents *same*-subject lagged test scores, $a_{iq(-1)st}$ and $a_{iq(-2)st}$ are the two *cross*-subjects lagged test scores, and λ_q, θ_q and κ_q are subject-specific parameters that capture the effects of lagged test scores in the same- and cross-subjects.[37] Anticipating our findings below, we find that results from within-pupil specifications are virtually unaffected by whether or not we control for pupils' own age-11 test scores. This is because there is neither a sizeable nor a significant correlation between the within-student across-subject variation in prior achievements, and the variation in peers' ability across subjects. Stated differently, conditional on pupil fixed-effects, peers' subject-specific quality measures are almost perfectly balanced with respect to pupils' own age-11 test scores, and specifications that include pupil fixed-effects effectively take care of the sorting of pupils and their peers into secondary education.

We further complement our core strategy with a set of robustness checks and alternative specifications that allow us to gauge the importance of subject-specific school selection and pupil sorting. For example, we include in some of our specifications school-by-subject fixed-effects to control for the sorting of pupils and their peers into schools based on subject-specific school unobservables. All these exercises provide strong support to the causal interpretation of our estimates.

3.3 Measuring peers' ability

A key requirement for our empirical approach is that the proxies of peer ability are based on pre-determined measures of students' ability that have not been affected by the quality of his/her peers and thus do not suffer from reflection problems. As already discussed, the longitudinal structure of the administrative data that we use allows us to link peers' KS2 test scores taken at the end of primary school (6th grade) to students'

[37] Note that conditional on pupil fixed effects, the same-subject and two cross-subjects lagged test scores cannot be simultaneously identified. Therefore, in our within-pupil empirical specification, we only include the same-subject lagged test score and one of the two cross-subject lagged outcomes.

KS3 achievements four years later, that is 9th grade in secondary school. Additionally, by following individuals over time, we are able to point out which secondary school students come from the same primary and identify who the new peers and the old peers are. On average, about 87% of pupil *i*'s peers in secondary school did not attend the same primary institution as student *i*, and therefore their KS2 test scores could not have been affected by this pupil. Following Gibbons and Telhaj (2008), in our analysis, we construct peer quality measures separately for new peers and old peers, and focus on the effect of the former on pupil achievement to avoid reflection problems. Nevertheless, in most of our empirical work we also include measures of the quality of old peers as additional controls. These help us to control for primary-school × cohort × subject effects that might persist on age-14 test scores and that are shared by pupils coming from the same primary school and cohort. Note however that our estimates are not sensitive to the inclusion of these variables.

Two additional remarks are worth being made. First, we use information about the school that a pupil is attending at age-12 (7th grade), when he/she enters secondary education, to define our base population. Similarly our three measures of peer quality 'treatment' (the 'good', the 'bad' and the average peer quality) are based on 7th-grade enrollment. This is because any later definition of these proxies, for example as recorded at KS3, might be endogenous. Second, in implementing this methodology, we use peers' ability measured at the grade and *not* at the class level because our data does not include class identifiers. We do not see this as a particularly restrictive compromise since the majority of schools do not strictly group pupils with different subject-specific abilities in different classes at the early stages of secondary education (see more details in the next section). Therefore, the quality of peers within a grade is likely to be strongly correlated with the quality of peers within classes. On the other hand, if some degree of subject-specific streaming takes place so that our peer quality measures capture the peer quality actually experienced by pupils with some noise, our estimates will be downward biased and more properly interpreted as 'intention-to-treat' effects.[38]

[38] Note that our study does not suffer from measurement error due to incomplete information on pupils' schoolmates as in Ammermueller and Pischke (2009).

4 Institutions, data and descriptive statistics

4.1 Schooling in England: institutional background

Compulsory education in England is organized into five stages referred to as Key Stages. In the primary phase, pupils enter school at age 4-5 in the Foundation Stage, then move on to Key Stage 1 (KS1), spanning ages 5-6 and 6-7 (these would correspond to the 1st and 2nd grade in other educational system, e.g. in the US). At age 7-8 pupils move to KS2, sometimes – but not usually – with a change of school. At the end of KS2, when they are 10-11 (6th grade), children leave the primary phase and go on to secondary school where they progress through KS3 (7th to 9th grade) and KS4 (10th to 12th grade). Importantly, the vast majority of pupils changes schools on transition from primary to secondary education, and move on to the school of their choice.

Indeed, since the Education Reform Act of 1988, the 'choice model' of school provision has been progressively extended in the state-school system in England (Glennerster, 1991). In this setting, pupils can attend any under-subscribed school regardless of where they live and parental preference is the deciding factor. All Local Education Authorities (LEAs) and schools must organize their admissions arrangements in accordance with the current statutory Governmental Admissions Code of Practice. The guiding principle of this document is that parental choice should be the first consideration when ranking applications to schools. However, if the number of applicants exceeds the number of available places, other criteria which are not discriminatory, do not involve selection by ability and can be clearly assessed by parents, can be used to prioritize applicants. These vary in detail, but preference is usually given first to children with special educational needs, next to children with siblings in the school and to those children who live closest. For Faith schools, regular attendance at local designated churches or other expressions of religious commitment is foremost. As a result, although choice is the guiding principle that schools should use to rank pupils' applications, it has long been suspected that schools have some leeway to pursue some forms of covert selection based on parental and pupil characteristics that are correlated with pupil ability (see West and Hind, 2003).

As for testing, at the end of each Key Stage, generally in May, pupils are assessed on the basis of standard national tests (SATS), and progress through the phases is measured in terms of Key Stage Levels, ranging between W (working towards Level 1) up to Level 5+ during primary education and Level 7 at KS3. Importantly for our research, at both KS2 (6^{th} grade) and KS3 (9^{th} grade) students are tested in three core subjects, namely Mathematics, Science and English, and their attainments are recorded in terms of the raw test scores, spanning the range 0-100, from which the Key Stage Levels are derived. We will use these test scores to measure pupils' attainments at KS3 and identify the quality of their peers as measured by their KS2.

Finally, regarding the organization of teaching and class formation, two important issues are worth mentioning. First, the concept of 'class' is a rather hollow one in English secondary schools since students tend to be grouped with different pupils for different subjects. A second important aspect that characterizes English secondary education is the practice of 'ability setting', i.e. subject-specific streaming. Under these arrangements, secondary school pupils are initially taught in mixed-ability groups for an observation and acclimatization period of around a year, and then eventually educated in different groups for different subjects according to their aptitude in that specific topic. Subject-specific ability is often gauged using end-of-primary education (KS2) test scores; these are only available to schools several months after they have admitted pupils. However, teachers and school staff have some discretion in determining the ability set that is most appropriate for their students in different subjects (see DfES, 2006; Kutnick et al., 2006). Note that despite some explicit support from the Government, the practice of ability setting has not been fully adopted by secondary schools in England. Kutnick et al. (2005) reports that about 80% of secondary schools have ability sets for Mathematics at some point between 7^{th} grade and 9^{th} grade, but only 53% from grade 7. These figures are much lower for English and Science respectively at: 46% (at some stage between 7^{th} and 9^{th} grade) and 34% (from 7^{th} grade); and 59% (sometimes between 7^{th} and 9^{th} grade) and 44% (from 7^{th} grade). In conclusion, two important features emerge from this brief discussion. First, because of the lack of clearly defined and stable classes during secondary education, students will predominantly interact with different peers in different subjects. Second, since ability setting is not strictly implemented, pupils will face a variety of class-mates with a heterogeneous range of abilities during instruction time even for the same subject.

4.2 Data construction

The UK's Department for Children, Schools and Families (DCSF) collects a variety of data on all pupils and all schools in state education[39]. This is because the pupil assessment system is used to publish school performance tables and because information on pupil numbers and pupil/school characteristics is necessary for administrative purposes – in particular to determine funding. Starting from 1996, a database exists holding information on each pupil's assessment record in the Key Stage SATS described above throughout their school career. Additionally, starting from 2002, the DCSF has also carried out the Pupil Level Annual School Census (PLASC), which records information on pupil's gender, age, ethnicity, language skills, any special educational needs or disabilities, entitlement to free school meals and various other pieces of information, including the identity of the school attended during years other than those when pupils sit for their Key Stage tests. The PLASC is integrated with the pupil's assessment records in the National Pupil Database (NPD), giving a large and detailed dataset on pupil characteristics, along with their test histories. Furthermore, various other data sources can be merged in at school level using the DCSF Edubase and Annual School Census, which contain details on school institutional characteristics (e.g. religious affiliation), demographics of the enrolled students (e.g. fractions of pupils eligible for free school meals) and size (e.g. number of pupils on roll).

The length of the time series in the data means that it is possible for us to follow the academic careers of four cohorts of children from age-11 (6^{th} grade) through to age-14 (9^{th} grade), and to join this information to the PLASC data for every year of secondary schooling (7^{th} to 9^{th} grade). The four cohorts that we use include pupils who finished primary education in the academic years 2000/2001 to 2003/2004, entered secondary school in 2001/2002 to 2004/2005, and sat for their KS3 exams in 2003/2004 to 2006/2007.

---

[39] The private sector has a market share of about 6-7%. However, very little consistent information exist for pupils and schools in the private domain. For this reason, we do not consider private schooling in our analysis.

We use information on these four cohorts as our core dataset because this is the only time window where we can identify the secondary school where pupils *start* their secondary education, and not only the one where they take their KS3 tests. As explained above, this is crucial to our analysis because we want to be able to measure peer exposure at the beginning of secondary schooling (in 7th grade), and not after three years (in 9th grade). The data also allows us to gather information about the primary school where pupils took the KS2 exams, which implies that we are able to single-out secondary schoolmates that are new peers from those who instead came from the same primary school (i.e. old peers).

Using this set of information we construct a variety of peer quality measures based on pupil achievements at KS2 in the three core subjects. In order to do so, we use the KS2 test scores, separately by subject and cohort, to assign each pupil to a percentile in the cohort-specific and subject-specific national distribution. We then go on to create three separate measures of peer quality. First, we compute the average attainments of peers in the grade at school. Next, we create two measures that are meant to capture peer effects coming from very bright and very worst students at school, namely: the fraction of peers (in the grade at school) below the 5th percentile or above the 95th percentile of the cohort-specific national distribution of KS2 test scores.

We have imposed a set of restrictions on our data in order to obtain a balanced panel of pupil information in a balanced panel of schools. First, we have selected only pupils with valid information on their KS2 and KS3 tests for whom we can also match individual background characteristics and the identity of the school where they start their secondary education using PLSAC. Given the quality of our data, this implies that we drop less than 2.5% of our initial data. Next, we have focused on schools that are open in every year of our analysis, and have further dropped secondary schools that have a year-on-year change of entry-cohort size of more than 75% or enrolments below 15 pupils. While the former restriction excludes schools that were exposed to large shocks that might confound our analysis, the latter excludes schools that are either extremely small or had many missing observations. These restrictions imply that we lose less than 2.5% of our observations.[40] Furthermore, we apply some restriction based on the fraction of bottom 5% and top 5% pupils, in order to exclude schools with particularly high or low shares of 'good' and 'bad' peers. In particular, we drop schools where the fractions of pupils below the 5th percentile or above the 95th percentile of the cohort-specific KS2 national distribution exceeds 20%, and schools that do not have any variation over the four years in these fractions. This last restriction predominantly trims schools that have no students in either the top or bottom 5% of the ability distribution in any year in any subject and would not contribute to the identification of peer effects. The two combined restrictions imply that we drop an additional 10% of our sample. Since this seems a large share, we checked that our main results are not affected when we omit these restrictions.

Our final dataset includes a balanced panel of approximately 1,300,000 pupils for whom we can observe complete information in terms of KS2 and KS3 test scores, individual and family background characteristics, and both primary and secondary school level information from age-11 to age-14. In the next section, we present some descriptive statistics for our core sample.

4.3 Some descriptive statistics

In table 1 we present descriptive statistics for the main variables of interest for the sample of 'regular' students, defined as pupils with age-11 test scores in the three core subjects above the 5th percentile and below the 95th percentile of KS2 test score distribution (column 1). The regression analysis that follows is mostly based on these pupils, which we sometimes refer to as 'treated' students. In the same table, we also presents descriptive statistics for pupils in either the top 5% or bottom 5% tails of the ability distribution – that is 'good' and 'bad' peers – which we also label as 'treatments' and which are excluded from the estimation sample. The rationale for this exclusion is to keep the distinction between treated students and pupils that 'form' our treatments clean. However, we are aware that there are valid reasons to include the top/bottom 5% pupils in the sample, e.g. to capture the extent of peer effects for pupils in the tails of

[40] We have also excluded selective schools (e.g. Grammar schools) from our analysis, as these schools can actively choose their pupils based on their ability (about 8% of our original sample).

the ability distribution. Therefore we also discuss some results that are based on a sample that further appends top/bottom 5% pupils, and show that our general findings and conclusions are totally unaffected by the inclusion of these pupils in the estimation sample.

In the top panel of the table we describe pupils' test scores at KS2 and KS3. Unsurprisingly, the first column shows that for regular students test score percentiles are centered just below 50, for all subjects and at both Key Stages. The correlations of pupils' KS2 test scores across subjects are 0.60 for English and Mathematics; 0.63 for English and Science; and 0.68 for Science and Mathematics. At KS3 these correlations increase to 0.64, 0.68 and 0.80, respectively. Appendix table 1 further shows that the within-pupil variations of KS2 and KS3 test scores across the three subjects are respectively 11.9 and 11.2. Overall, this provides evidence that test scores are not perfectly correlated across subjects for the same student, although they tend to be more closely associated in Science and Mathematics, in particular at KS3.

The remaining two columns of the table illustrate how pupils with at least one subject in either the top 5% or the bottom 5% of the ability distributions score at their KS2 and KS3 tests. By construction, pupils in top 5% of the KS2 test score distribution perform much better than any other pupil in their KS2 exams, while the opposite is true for pupils in the bottom 5% tail. We get a very similar picture if we look at pupils' KS2 test scores in one subject (e.g. English) imposing that at least one of the other two subjects (e.g. Mathematics or Science) is above the 95^{th} percentile or below the 5^{th} percentile of the test score distribution.[41] More interestingly, this stark ranking is not changed when we look at KS3 test scores, for all subjects, with little evidence of significant mean reversion in the achievements of very good and very bad peers between age-11 and age-14. To further substantiate this point, we have thoroughly analyzed the KS3 percentile ranking of pupils in the top 5% and bottom 5% of the KS2 achievement distribution. For all subjects, about 80% of the pupils ranking in the bottom 5% at KS2, still rank in the bottom 20% of the KS3 distribution, with approximately 70% concentrated in the bottom 10%. At the opposite extreme, around

[41] For example, the KS2 percentiles in English for pupils with at least Mathematics or Science in the top 5% and bottom 5% are 83.8 and 9.8, respectively.

80% of pupils ranking in the top 5% at KS2 remains in the top 20% of the KS3 achievement distribution, with the vast majority still scoring in the top 10%. This reinforces the idea that our 'good' and 'bad' peers are consistently amongst the brightest and worst performers.

The second panel of table 1 presents more information on pupil background characteristics. The figures in the first column reveal that our sample is fully representative of the population of secondary school pupils in England. On the other hand, pupils with at least one subject in the bottom 5% are less likely to have English as their first language and to be of White British ethnic origins, and more likely to be eligible for free school meals (a proxy for family income). The opposite is true for pupils with at least on subject in the top 5%. However, the differences in family background are much less evident than those in terms of academic ability presented in panel A. Peer ability measures defined in terms of pupil background would therefore severely underestimate differences in peers' academic quality.

Finally, in panel C we report school characteristics for the various sub-groups. The average cohort size at the start of secondary school in 7th grade is approximately 200, and around two thirds of all pupils attend Community schools, while about 16% of the pupils attend a religiously affiliated state-school. Pupils with at least one subject in the top 5% of the ability distribution are less likely to attend a Community school, and more likely to be in a faith school, than pupils in the central part of the ability distribution and students with at least one subject in the bottom 5%. However, these differences are not remarkable.

In table 2, we present some descriptive statistics of our 'treatments'. Statistics are presented for new peers only. Note once again that on average pupils face 87% new schoolmates, although the distribution of new peers is highly right-skewed, with many more pupils facing 100% new schoolmates than zero. Panel A summarizes the average peer quality, computed as the average KS2 percentile rank of peers in a given subject (excluding the pupil under consideration). Unsurprisingly, this is centered on 50 for all subjects. Panel B and panel C, instead, present descriptive statistics for our proxies for 'good' and 'bad' new peers. By construction, the fractions of top 5% and bottom 5% 'new peers' in the incoming cohort are smaller than the corresponding fractions including all peers (at around 5% each in every subject). Note that all peer quality measures display quite a wide range of variation, although this mainly capture differences across schools. Nevertheless, appendix table 1 shows that the same pupil faces considerably different fractions of academically bright and weak students across different subjects, as well as a significant amount of within-pupil across subject dispersion in average peer's age-11 test scores. This is the variation that our pupil fixed-effect regressions will exploit to identify the effect of peer quality.

5 Results

5.1 Effects of peers' ability: main findings

We begin the discussion of our results by presenting estimates of the impact of the peer quality on pupil outcomes at KS3 obtained using the sample of 'regular' pupils and controlling for potential subject-specific sorting by including lagged test scores. Results are reported in table 3. Columns (1) and (2) present OLS and within-pupil estimates of the effect of average peer quality. Next, columns (3) and (4) present OLS and within-pupil estimates of the effect of the percentage of bottom 5% peers, while columns (5) and (6) present estimates of the effect of the percentage of top 5% peers. The estimates presented in the four rows of the table come from a variety of specifications, which differ in the way they control for lagged test scores. In the first two rows, we report estimates unconditional on age-11 achievements, while the third row presents estimates where we include pupils' own KS2 attainment in the same subject in interaction with subject dummies. This allows pupils' lagged outcomes to affect age-14 test scores differently in different subjects. Finally, in the last row, we include pupils' own KS2 test scores in the *same*-subject and *cross*-subject (as detailed in section 3.2) in interaction with subject effects to control for pupils' own subject-specific ability, as well as cross-subject spillovers. Note that the results in the first row are obtained from different regressions entering either the average quality of peers, or the fraction of top 5% and bottom 5% peers in the grade. Results in the remaining three rows come from regressions that include all three treatments together.

Starting from the first two rows, OLS estimates in columns (1) show a high and positive partial correlation between average peer quality and students' KS3 achievements. The estimated coefficient is approximately 0.36 when only the average peer quality is entered in the regression, and it drops to 0.19 when the quality of top and bottom peers is further appended to the specification. This suggests that the tails of the ability distribution potentially capture most of the partial correlation between average peer ability and KS3 achievements.[42] A similar picture emerges when looking at columns (3) and (5), which display OLS estimates of the effect of top 5% and bottom 5% peers at schools: the estimated coefficient on 'good' peers is large, between 0.75 and 0.33, while the estimated effect of 'bad' peers is significantly negative and in the order of -0.6/-1.0.

However, a markedly different picture emerges when looking at columns (2), (4) and (6). These come from specifications that include pupil fixed-effects as described by equation (1) in section 3.1, and rely on within-pupil variation in age-14 test scores and peer quality to identify ability peer effects. Column (2) shows that the positive impact of average peer quality completely disappears upon inclusion of pupil fixed-effects: this is now estimated to be around 0.02, and not statistically different from zero. Similarly, column (6) shows that the within-pupil estimates of the effect of the most academically talented peers are small and not statistically different from zero. Only the effect of the bottom 5% peers remains sizeable and significantly negative after including pupil fixed-effects. As shown in column (4), this is estimated to be -0.12 in the first row, and -0.09 in the second row, where all three treatments are included simultaneously. Focusing on the latter, this is approximately one sixth of the corresponding OLS estimate. Although one reason why within-pupil estimates of peer effects might be smaller than OLS is because they net out overall effects that might arise through cross-subject interactions, this dramatic reduction is more likely due to the fact that within-pupil estimates control for pupil own unobserved average ability, unmeasured family background characteristics and school-by-cohort unobserved effects.

[42] To avoid double counting, we have also computed and experimented with measures of the average peer quality that exclude the top 5% and bottom 5% tails, and have come to identical conclusions.

Nevertheless, pupil fixed-effects estimates presented in the first two rows are unconditional of KS2 achievements, and thus potentially contaminated by subject-specific pupil sorting. Therefore, in the last two rows of table 3, we go on to include lagged test scores as an attempt to control for any residual pupil subject-specific ability and sorting. Note again that the specification in the third row presents estimates from specifications where we include pupils' own KS2 attainment in the same subject in interaction with subject dummies, while in the last row we include pupils' own KS2 test scores in the *same*-subject and *cross*-subject in interaction with subject effects. This 'control function approach' follows the strategy described in section 3.2.

Comparing the second to the third and fourth rows, we find that OLS estimates of ability peer effects are now between 15% and 30% smaller than before. However, even when controlling for lagged test scores in the OLS specification in a very flexible way as in row (4), we are unable to reduce our estimates of the effect of peers' quality to values close to the within-pupils estimates. This strongly speaks in favor of within-pupil regressions, which allow us to control non-parametrically for pupils' unobservable average ability and school-by-cohort unobservable shocks. On the other hand, the within-pupil estimates are essentially *unaffected* by the inclusion of pupils' age-11 test scores. The effects of the average peer quality and of the share of bright students remain small and insignificant. More interestingly, the effect of the bottom 5% peers only marginally drops to -0.091 (from -0.095), when we only include KS2 attainment in the same subject, and to -0.089, when we further include cross-subject lagged test scores.[43,44] This finding is particularly reassuring especially considering that the same-subject lagged test score enters the within-pupil regressions with a large coefficient (of

[43] We have also tried some specifications where we further include age-7 test scores. These are available for only three out of out four cohorts, and students are not tested in science at age 7, so that we had to impute test scores in this subject using the average between mathematics and English. Even then, our findings were fully confirmed, with no effects coming from average peer quality and top students, and strong negative (same size) effects from the fraction of bottom 5% peers.

[44] Note that the negative effect of 'bad' peers is slightly larger if we focus on students with a very high percentage of new peers at school. For example, considering the sample of pupils with at least 95% new peers (corresponding to the top 25% percent of students with the largest fraction of new peers) we still find that only the fraction of bad peers has a significant impact, now estimated to be at -0.093 (s.e. 0.040).

about 0.35, for example, in the third row), and is highly significant. In fact, the reason why lagged test scores hardly affect within-pupil estimates of effect of the bottom 5% new peers is that there is neither a sizeable nor a significant correlation between the within-student across-subject variation in own age-11 achievements, and the variation in 'bad' peers' ability across subjects. Stated differently, conditional on pupil fixed-effects, the fraction of bottom 5% peers in one subject is completely balanced with respect to pupils' own age-11 test scores in that subject. To assess this more formally, we re-ran the regression in equation (1) replacing age-14 with age-11 pupil test scores as the dependent variable. Results are presented in the appendix table 2. The coefficient on the fraction of bottom 5% peers was small at about -0.024 (with a standard error of 0.013), and not significant at conventional levels. On the other hand we found some degree of positive selection on the average peer quality, and some negative selection on the fraction of top 5% peers. This small residual sorting however does not substantially affect our results, which are steadfastly anchored at zero as soon as we include pupil fixed-effects.[45] All in all, these findings suggest that within-pupil specifications effectively take care of the endogenous sorting of pupils and their peers into secondary education, and that any residual subject-specific sorting is too small to confound out estimates.

The final set of results that we discuss in this section comes from regressions that further include the top 5% and bottom 5% peers in the sample that we use to estimate peer effects. As discussed in the data section, we prefer not to add 'good' and 'bad' peers in the estimation sample in order to keep the distinction between treated and treatments clean. However, our general conclusions are not affected at all by the inclusion of these pupils in the estimation sample. For example, based on the specification of row 4 of table 3 and considering the fully-inclusive sample, our estimate for the effect of average peer quality is 0.010 (s.e. 0.011), the one for the bottom 5% peer

[45] Regarding the effect of average peer quality being zero, we further looked into this issue by using the specification of row (4) but including in the regression *only* the average peer quality variable. When doing this, the within-pupil estimates goes from 0.010 (s.e. 0.12) to 0.018 (s.e. 0.012). This suggests that the reason why average peer quality does not have a sizeable impact when we include proxies for peers in the ability tails is because these capture much of the relevant 'empirical action', and not because we estimate net peer effects.

is -0.090 (s.e. 0.030), while the impact of the top 5% peers is 0.004 (s.e. 0.024). These coefficients are virtually identical to the corresponding estimates presented in table 3.

5.2 Robustness checks to potential threats to identification

In this section, we present a set of robustness checks that further support the causal interpretation of our findings. Results from these exercises are presented in table 4. Throughout the table, estimates come from within-pupil specifications that control for same- and cross-subject KS2 test scores interacted with subject specific dummies as described by equation (2). Further details are provided in the note to the table.

As discussed in section 4, parental choice is the guiding principle that education authorities should adopt when ranking pupils' applications to schools. However, some forms of covert selection might still take place, based on pupil and family characteristics that are associated to students' academic ability, overall or in a specific subject. Such cases might arise for example for pupils attending 'specialist' schools, i.e. schools with a stated 'specialism' in a given subject. This is because specialist schools are allowed to introduce admissions priority rules for up to 10% of their intake for pupils who demonstrate a particular aptitude in the subject of their expertise. In our sample, about 8.5% of the students attend a specialist school. Some common areas of specialism include: language; mathematics and computing; science; technology; business and enterprise; and arts. In the first row of table 4, we present estimates of the effects of the three measures of peers' quality obtained excluding from the sample pupils in specialist schools. These within-pupil estimates are largely identical to those discussed in table 3 for all peer quality measures.

Next, in the second row of the table, we look into whether results are driven by the fact that the school is above capacity (over-subscribed) or not (at capacity or under-subscribed). As highlighted in section 4, over-subscribed schools have some discretion in prioritizing pupils for admissions. The concern is that popular schools, receiving more admissions requests than they can accommodate, might covertly select students with characteristics that are particularly suited to their teaching expertise and other school infrastructures *specific* to one of the three core subjects under analysis. On the other hand, we are not concerned with potential selection based on pupil overall ability, as this is fully taken care of in the within-pupil specifications. To allay these concerns, row (2) of table 4 presents results obtained by further excluding over-capacity schools (accounting for approximately 40% of pupils in non-specialist schools). The within-pupil estimates of the effects of peers' quality are similar to those obtained before, in particular for the impact of the fraction of bottom 5% peers, which is now slightly larger at -0.100 (s.e. 0.040). Results (not tabulated, but available upon requests) further show that our findings are similar for non-specialist secular schools and non-specialist schools with a religious affiliation. All in all, the evidence suggests that neither *school-side* selection of pupils with unobservables potentially correlated with ability in a given subject, nor other school institutional features are driving our main results.

Another robustness check assesses whether *parental choice* of schools with an 'expertise' in a given subject might confound our estimates of peer effects. To do so, we examine whether our findings are driven by sorting of students who choose to attend a school with peers that excel in the same subject. More precisely, we identify two groups of students: *(i)* those who excel in subject q (say English) and go to schools where, on average over the four years of our analysis, new peers also excel in that subject; and *(ii)* those who excel in subject q (say, again, English) and go to schools where, on average over the years, new peers excel in a different subject (either Mathematics or Science). We label these two groups as 'sorted' and 'mixed' pupils, respectively.[46] We then re-run our analysis only including 'mixed' students to understand whether our results are driven by sorting of pupils with similar unobservables that are conducive to excellence in subject q (e.g. English) in the same school. Results from this exercise are reported in row (3) of table 4 and support our previous findings. Even when considering only 'mixed' pupils, we find no significant effects from peers of average quality and from the fraction of new peers in the top 5% of the ability distribution. On the other hand, we still find a sizeable and statistically significant negative effect from the bottom 5%

[46] Note that peers' excellence in a subject is defined using new peers' average KS2 test scores. Our results are unaffected if we use the fraction of new peers in the top 5% of the ability distribution.

peers. The estimated impact is at -0.100 (s.e. 0.034), which fully confirms our results so far.

To provide further evidence of the validity of our specifications, we next perform a robustness check based on replicating our results for increasingly selected subsets of students with increasingly small within-pupil standard deviation of KS2 test scores across the three subjects. While we move towards more 'limited' samples, we reduce the possibility that there is any correlation between one pupil's subject-specific *observed* ability and that of his/her peers. This is because the within-pupil variation of age-11 test scores across subjects is forced to become progressively close to zero. In the empirical application, we perform this exercise by selecting students with the within-pupil standard deviation of KS2 test scores across the three subjects below increasingly smaller thresholds (e.g. s.d.≤4, s.d.≤3.5, s.d.≤3, etc.).[47] Following the reasoning in Altonji et al. (2005), any residual sorting on *unobservable* subject-specific attributes most likely tracks and is upward bounded by the amount of selection on *observable* subject-specific characteristics, in particular lagged tests scores. Thus, by focusing on progressively 'limited' samples of students with little or no within-pupil variation in age-11 achievements and by studying how our estimates of the ability peer effects change, we are able to assess whether any residual subject-specific sorting might bias our estimates.

We present our findings graphically in Figure 1, where we focus on the bottom 5% new peers.[48] The plots present regression coefficients and 95% confidence intervals (standard errors clustered at the school level) coming from 23 different regressions estimated separately for progressively smaller subsets of pupils with variation across subject in KS2 test scores falling below predefined thresholds of the within-pupil standard deviation of age-11 attainments. These spanned the interval s.d.≤3 to s.d.≤11.5, in steps of 0.5, and then std.dev.≤15; std.dev.≤17.5; std.dev.≤23; std.dev.≤26; and full sample. Note that the estimates presented in the top panel come from specifications as

[47] Note that identifying the 'limited' samples by imposing a restriction on the variation in lagged test scores within-pupil is analogue to within-pupil non-parametric 'matching' based on the three lagged test scores observed for each student. That is we match within-pupil on α_{iqst}, $\alpha_{iq(-1)st}$, and $\alpha_{iq(-2)st}$ in equation (2), and only keep pupils with a 'close-enough' match to themselves across subjects.

[48] Results for the other two treatments lead us to identical conclusions. They are not reported for space reasons, but are available from the authors.

in equation (1), where the dependent variable is pupil age-11 achievement, and therefore present the balancing of this treatment with respect to pupils' own KS2 test scores. On the other hand, the estimates displayed in the bottom panel are obtained from specifications as in equation (2), and show how our the treatment effect varies across different groups of pupils.

The top panel shows that the share of 'bad' peers is not significantly related to within-pupil variation in KS2 test scores almost throughout the various sub-samples. Even when this relation reaches some statistical significance, the degree of unbalancing is very small. Expectedly, as we move to more restricted samples of pupils, the balancing gets closer to perfect with estimated coefficients of -0.005, -0.002 and exactly zero for pupils with s.d.≤4, s.d.≤3.5 and s.d.≤3, respectively. However, the most remarkable findings from this exercise appear in the bottom panel: even as we shrink the within-pupil standard deviation of KS2 test scores towards zero, we still find negative and significant estimates of the effect of the bottom 5% new peers. More importantly, these estimates are stable at approximately -0.09 throughout the plot. For example, they takes values of -0.084 (s.e. 0.032) and -0.109 (s.e. 0.040) for the sets of pupils with s.d.≤11.5 and s.d.≤3, respectively. Furthermore, the confidence intervals throughout the figure are largely overlapping, clearly allowing us to reject the hypothesis that the estimates are different.

This last piece of evidence reinforces our main finding (evident in tables 3 and 4) that any residual subject-specific sorting based on unobservable considerations must be sufficiently small not to confound our estimates of the effect of peers' quality conditional on pupil fixed-effects. In fact, any bias due to confounding subject-specific unobservables should have a very special pattern so as to lead to the same or slightly larger point estimates of the effects of 'bad' peers in samples of pupils with progressively shrinking degrees of cross-subject variation in lagged test scores. In particular, selection on unobservables should be uncorrelated or negatively related to lagged test scores in order to explain these results. This is highly implausible since KS2 test scores are reliable proxies of pupils' subject-specific abilities, and pupils with similar subject-specific abilities or preferences tend to sort into the same schools.

5.3 Extending the group of bottom and top peers beyond the 5% threshold

One issue that we have so far left un-assessed is our choice of the 5% threshold to define the very good and markedly poor peers. Different cut-off points could have been chosen, potentially affecting our results. In Figure 2, we tackle this issue directly by looking at whether peers in other parts of the ability distribution significantly affect pupils' age-14 cognitive outcomes. The figure presents treatment effect estimates and associated 95% confidence intervals for different measures of the bottom and top new peers, and coming from specifications as in equation (2). For the bottom treatment, we define the following five groups: bottom 5%; 5 to 10%; 10 to 15%; 15 to 20% and 20 to 25%. For the top group, we define the following five peer measures: top 5%; 90 to 95%; 85 t0 90%; 80 to 85% and 75 to 75%. Note that the sample of 'treated' pupils now only includes students in the range from 25th to the 75th percentiles of KS2 test scores.

Figure 2 reveals a markedly asymmetric pattern. All five bottom peer groups have a negative effect on other pupils, but this effect is clearly significant only for the first group, and it declines sharply in scale as we move away from the very bottom group. On the other hand, the effect of the top peers at school is small and insignificant throughout. This suggests that our choice of top 5% and bottom 5% peers is not arbitrary and provide clear evidence that: *(i)* it is only the very bottom 5% of news peers that are strongly and negatively associated with pupils' own age 14 test scores, and not 'bad' peers in other parts of the ability distribution; and *(ii)* there is no evidence that 'good' peers in other parts of the ability distribution affect students' cognitive outcomes.

To conclude this section, we provide an assessment of the magnitude of the negative effect of the bottom 5% peer treatment based on the estimates presented in table 3. To do so, we begin by scaling it according to the minimum and maximum values of the bottom treatment variable observed in the data, at zero and 20% respectively (see table 2). A pupil who moves from 20% to 0% of the bottom quality peer group would experience an improvement in KS3 test scores of about 1.8/2 percentiles, which amounts to 0.08/0.09 of the standard deviation of KS3 test score, or 0.16/0.17 if we consider the standard deviation of the within-pupil KS3 distribution. Note that these are rather sizeable experimental changes, as they correspond to about 20 standard deviation changes in the within-pupil peer quality distribution. More modest changes of a 10 percentage point decline in the share of weak peers would imply an improvement of around 0.08 of the within-pupil standard deviation in the KS3 distribution. Relative to other studies that focus on school inputs and interventions, our estimates of the effect of academically weak peers capture a medium-to-small sized effect. For example, Lavy (2009) estimates the effect of instructional time in secondary schools using the PISA 2006 data and reports an average effect for OECD countries of 0.15 of the within-pupil standard deviation of test scores across subjects for an additional hour of classroom instruction. These estimates imply that the ability peer effects that we estimate here for a 10 percentile decrease in the percentage of 'bad' peers quality is equivalent to the effect of half an hour of weekly instruction time. Another possible comparison is to the effect size of peer quality estimated in Ammermueller and Pischke (2009) across-classes within-schools in six European countries. This study reports that one standard deviation change in their student background measure of peer composition leads to a 0.17 standard deviation change in reading test scores of fourth graders. Finally, Bandiera et al. (2009) study class size effects at university using a within-pupil specification similar to ours. Their results show that a one standard deviation of the within-pupil class size distribution improves test scores by 0.11 of the within-pupil standard deviation of outcomes.

5.4 Estimates of peer effects in small schools

We next turn to analyze whether the within-pupil estimates of the peer effects are significantly different in small schools. As explained in section 3.3, the possibility that schools implement subject-specific ability grouping (setting) means that we might underestimate the full extent of the scale of peer effects. By focusing on smaller schools and analyzing how our estimates change, we can partly allay these concerns. This is because schools with a smaller pupil intake will have fewer classes. Therefore students will be more mixed with peers of heterogeneous abilities in smaller schools than pupils in larger ones, where more classes can be created to group students according to their abilities. Notice that these arrangements stems from the fact that schools receive funding based on pupil number and have clear incentives to run classes at maximum capacity (approximately 30/35 students).

To perform this check, we focus on schools with pupil intake below the median of the year-7 cohort-size distribution. Stated differently, we consider (approximately) the 50% smallest schools with incoming cohort size of at maximum 180 pupils, and with on average of 136 students. Results are reported in the last three rows of table 4. Rows (4) and (5) present estimates that come from specifications as detailed in equations (1) and (2), respectively. Once more, we find that controlling for pupil age-11 test scores in a very flexible way does not affect the within-pupil estimates. These are still clearly zero for the effect of the average peer quality and negative significant at around -0.10 for the fraction of new peers in the bottom 5% of the ability distribution. On the other hand, we find more positive effects for the fraction of top 5% new peers. This points into the direction of better peers interacting more with regular students in smaller schools and exercising more positive externalities. However, note that neither of the estimates in rows (4) and (5) is significant at conventional levels. To further explore this issue, we also looked at the results for the smallest 25% schools and found that the effect of the very bright new peers is in the order of 0.060, but still not statistically significant, with an associated standard error of 0.044 (on the other hand, the effect of the bottom 5% new peers rises slightly to -0.12 with a standard error of 0.052).

One further advantage of focusing on small schools is that peers' subject-specific quality is more likely to display significant year-on-year variation due to random subject-specific cohort shocks (recall that general cohort-specific unobserved effects are account for by our regression). We can exploit this fact to further augment our specifications with school-by-subject fixed-effects that account for subject-specific school unobservables – such as teachers' expertise in a given field – which might drive pupils' and their peers' sorting. We estimate this specification using only the first and last cohorts in our data in order to maximize the variation over time that we can exploit to estimate peer effects. In fact, this approach is very demanding since conditional on pupil fixed-effects our data shows very little within-school-subject variation over time, in particular in terms of students' age-14 outcomes. This is because the 'spread' of pupils' KS3 test scores around their average is not significantly widening or vanishing over time within schools. This is perhaps not surprising given that we are considering standardized test scores, and that schools' composition does not dramatically fluctuates over four years. Even then, our results (presented in row 6 of table 4) broadly support our previous conclusions. The effects of the average peer quality and the fraction of top 5% new peers are still estimated to be small and insignificant. On the other hand, the peer effect from the very weak students is estimated to be a significant -0.070 (s.e. 0.021), only between 20-30% smaller than our main estimates.[49]

5.5 Additional findings: peer effects estimates by subject-couples

We mentioned in section 3 that one of the underlying assumption of the identification strategy is that peer effects are constant across different subjects. Although this assumption is difficult to test, some of the studies that have previously investigated peer effects separately for Reading and Maths have found similar estimates (e.g. Vigdor and Necyba, 2004; Hoxby and Weingarth, 2005), while others have documented small differences across subjects (see the review in Epple and Romano, 2010). To shed some light on this issue, we first ran OLS regressions by subjects separately, and found that the estimated effects of average peer quality and of 'bad' peers are virtually the same in the three subjects. These estimates are presented in the appendix table 3. On the other hand, there is slightly more variation for the effect of top peers across the three subjects, but the confidence intervals of the various estimates are largely overlapping. Even though we know OLS estimates are biased, these findings are informative and suggest that our assumption is not too unrealistic. In another related robustness check, we ran within-pupil regressions separately for couples of subjects, i.e. by pooling observations for: English and Mathematics only; English and Science only; and Mathematics and Science only. Results are not tabulated for space reasons, but are available from the authors.

[49] Importantly, when using the school-by-subject fixed-effects (coupled with pupil fixed-effects) specification on the full sample of schools, we find results that are very similar to the estimates based on the sample of small schools. In particular, the effect of average peer quality is 0.004 (s.e. 0.005), the impact of bottom peers is -0.057 (s.e. 0.013), and the effect of top peers is -0.001 (s.e. 0.010).

Our previous findings for the average quality of peers and the fraction of top 5% new peers were confirmed for all pairs of subjects. On the other hand, we found stronger peer effects from students in the bottom 5% of the ability distribution coming from the comparison of English with Mathematics and English with Science, than when only pooling Mathematics and Science. For the former two couples of subjects, estimates of effect of 'bad' peers were -0.102 (s.e. 0.059) and -0.116 (s.e. 0.057) respectively, whereas the comparison of Science and Mathematics yielded a smaller estimate of -0.049 (s.e. 0.040). This is perhaps unsurprising given that, as discussed in section 4, pupils' KS3 test scores are much more correlated for Science and Mathematics (0.80), than for English and Mathematics (0.64) or English and Science (0.68). As a result, there is less within-pupil across-subject variation in age-14 test scores to estimate peer quality effects. Indeed, the within-pupil variations for English-Mathematics and English-Science are 10.8 and 10.2, respectively 35% and 27.5% higher than the within-pupil variation for Mathematics-Science, at about 8.0. Moreover, the institutional details discussed above suggest that 'ability setting' is more common in Mathematics and Science than in English. Given the high correlation between pupil's attainments in these two subjects, it is likely that the one student will be 'set' at a similar level in these two subjects, thus facing peers of similar quality in both Science and Mathematics. Stated differently, both the within-pupil variation of the peers that the student *actually* interacts with, and the within-pupil variation in age-14 test scores might be too small to identify any peer effect. All in all, however, we believe the findings presented in this section broadly support our assumption that peer effects are similar across subjects.

6 Allowing for Heterogeneous Effects

6.1 Heterogeneity by students' ability

In this section, we test for the presence of heterogeneous effects along a variety of dimensions. We first examine if the very 'good', the very 'bad' and the average peers differentially affect students with different academic abilities. For this purpose, we stratify the sample into six groups according to the distribution of pupils' *average* KS2 percentiles across subjects. The percentile-ranges that define the six non-overlapping groups are as follows: 5-20; 20-35; 35-50; 50-65; 65-80; and 80-95. Our regression models now simultaneously include interaction terms of the percentages of top 5% peers, bottom 5% peers and average peer quality (separately for old and new peers) with dummies indicating to which of the six KS2 ability groups a pupil belongs to. Note that the effect of KS2 achievements in the *same*- and *cross*-subject is controlled for by interacting pupils' own KS2 percentiles with the dummies indicating his/her rank in the ability distribution (as well as subject dummies).

These findings are reported in table 5. The estimates presented in column (1) reveal that the quality of average peers does not affect regular pupils' age-14 test scores at any point of the ability distribution. This evidence further speaks against the 'linear-in-means' model that assumes that pupil achievement are a linear function of the peers' average ability, and is consistent with the findings of Hoxby and Wiengarth (2005). On the other hand, column (2) shows that the negative effect of the bottom 5% peers is roughly constant across various ability groups of regular students. In fact, there is some variation in the point estimates, with larger negative effects for pupils in the 50^{th} to 80^{th} percentiles of the ability distribution (at around -0.11), and insignificant negative effects for the most able pupils (of about -0.04). However, an F-test on the hypothesis that all coefficients are equal clearly accepts the null. These results are consistent with a 'bad-apple'-type model of peer effects in which a small number of weak students (in a specific subject) adversely affects the learning of all other pupils throughout the ability distribution (Lazear, 2001; Hoxby and Weingarth, 2005).

Results for the effect of peers in the top 5% of the ability distribution reveal a more interesting pattern; these are presented in column (3). On the one hand, they confirm our main finding, namely that there is no significant peer effect from very academically bright pupils on other regular students. An F-test for the joint significance of the treatment at the various parts of the ability distribution clearly accepts the null of no effect. On the other hand, while the impact of the top 5% peers is positive (insignificant) in the bottom two-thirds of the ability distribution of regular students, it turns negative (insignificant) for the most able pupils with average age-11 test scores between the 65^{th} and 90^{th} percentile. Consistently, an F-test on the null that all coefficients are equal rejects the hypothesis at the 10% level of confidence with a p-value of 0.080.

Since this finding is rather unexpected, we have assessed its robustness along a number of directions. For example, we have tested that it survives when we restrict our attention to pupils with less potential for subject-specific sorting, as identified by students with a limited standard deviation of KS2 across subjects (i.e. pupils with s.d.$\leq$3; see the discussion in section 5.3). Similarly, we tested that this pattern is not driven by the inclusion of specialist schools or over-subscribed schools. Finally, another possible and rather mechanical explanation for why pupils who are good on average marginally suffer from having many top 5% peers might be related to mean-reversion. In general, average test scores reveal some mean reversion. Pupils in the 5th-20th percentile at KS2 experience a 4 percentile point average improvement in their average KS3 test score, while students in the 80th-95th KS2 percentile have an average 5 percentiles deterioration in their mean KS3. However, the within-pupil standard deviations of KS2 for students in the same ability group must be similar by construction. This means that all pupils within the same ability group, in particular those in the 80th-95th KS2 percentile, would be similarly affected by mean-reversion *irrespective* of how many good peers they interact with. Moreover, if mean reversion was to explain our findings, we would expect this to affect both the top and the bottom of the ability distribution. However, we do not observe any interaction between either the top 5% peers or the bottom 5% peers and the fact that a student ranks low in the KS2 ability distribution. To shed further light on this issue, we formally checked whether the pure effect of belonging to the top-group in the average KS2 ability distribution (80th-95th percentile) is related to the KS3 outcomes of students, but failed to find any evidence. In a nutshell, mean reversion does not appear to be a likely explanation for these patterns. Anticipating our findings, we find that this result is completely driven by the negative and significant response of boys to a large fraction of top 5% peers, which becomes particularly strong for the most able male students. We carefully investigate these issues in the next section.[50]

---

[50] Note that we also tested how sensitive our results are to the inclusion of the bottom 5% and top 5% peers in the estimation sample. To do so, we have included pupils in the bottom 5% ('bad' peers) in the 'ability group 5-20', and peers in the top 5% ('good' peers) in the 'ability group 80-95'. Results using this extended sample are almost identical to the ones presented in table 5 (available upon request).

6.2 Gender heterogeneity in treatment effects

In this section, we analyze the heterogeneity of peer effects by gender. This is particularly interesting given that a growing body of evidence shows that girls are more affected than boys by education inputs and intervention.[51] Moreover, peer effects might work in significantly different ways for male and female students during secondary education, a time when both the identification with and the social interactions between the two genders intensify. We report our first set of results in table 6. The top panel looks at boys (columns (1) and (2)) and girls (columns (3) and (4)) separately, but pooling pupils of all abilities. The bottom panel of the table instead further ranks students by their KS2 average ability. More details about the specifications are provided in the note to the table. Note that all regressions further include the average quality of peers. However, since this treatment did not reveal any significant pattern, we have not tabulated these coefficients (results available upon request).

Results in the panel A of the table show that the effect of the bottom 5% peers is negative and significant in both gender groups, although it is slightly smaller for boys (at -0.076) than for girls (at -0.098). On the other hand, the effect of the top 5% peers is positive, significant and sizeable at 0.066 for girls, but negative for boys at -0.052, and significant at better than the 10% level (p-value: 0.068). These patterns are not easily explained by differential subject-specific sorting for boys and girls into schools with peers of different quality. In fact, we find no significant relation between the within-pupil across-subject variation in age-11 achievement and the variation in the fraction of top 5% new peers in different subjects for boys, and a small negative relation for girls (with coefficient of -0.064 and a standard error of 0.015), indicating some degree of negative sorting for female students. This clearly suggests that selection can hardly be driving our results (unless this occurs on subject-specific unobservables that are

[51] For example, Anderson (2008) shows that three well-known early childhood interventions (namely, Abecedarian, Perry and the Early Training Project) had substantial short- and long-term effects on girls, but no effect on boys. Likewise, the Moving to Opportunity randomized evaluation of housing vouchers generated clear benefits for girls, with little or even adverse effects on boys (Katz et al., 2001). Some recent studies also show a consistent pattern of stronger female response to financial incentives in education, with the evidence coming from a variety of settings (see Angrist and Lavy, 2009; Angrist et al., 2009).

negatively correlated with age-11 test scores)[52]. Note that we also checked whether our results are driven by the inclusion of single-sex schools. These enroll approximately 2% of the boys in our sample, and slightly more than 4% of the female students. Although results obtained after excluding these pupils were slightly weaker, they provided a similar picture: the effect of the bottom 5% peers is negative for both boys and girls, but the effect of the most academically talented peers is positive for female students and negative for males.

To shed further light on these patterns, we next study the sign and size of ability peer effects separately of boys and girls, and in interaction with students' own ability. Results are presented in panel B of table 6, and replicate the structure of table 5. For both boys and girls, we find that the effect of many 'bad' peers at school is relatively stable throughout the ability distribution of regular students. The negative impact of bottom 5% peers is slightly stronger for pupils in the 50^{th}-80^{th} percentile range of the ability distribution. However, there is little evidence that these differences are statistically significant: an F-test on the hypothesis that all coefficients are equal accepts the null with p-values of 0.2597 and 0.6809 for boys and girls, respectively.

A more interesting pattern of results emerges when we focus on the effect of the top 5% new peers. Looking at girls first, we find that the impact of academically bright peers is positive throughout the ability distribution, although this effect is more pronounced and statistically significant for female students with KS2 achievements below the median of the ability distribution. On the other hand, the impact of top 5% peers becomes smaller and loses significance for the most talented girls, in particular for those with age-11 achievement above the 80^{th} percentile, where the estimated coefficient is small and insignificant at 0.011 (s.e. 0.039). In sharp contrast, we find that the impact of having many 'good' peers at school is negative for males throughout the ability distribution, although this adverse effect is only statistically significant for the most able boys. The estimated impact for males with average KS2 test scores in the 65^{th} to 80^{th} percentile is -0.079 (s.e. 0.037), and further increases to -0.096 (s.e. 0.043) for those in the 80^{th} to 95^{th} percentile bracket. Note that we checked once again whether our

[52] Results unconditional on pupil's age-11 test scores confirmed these heterogeneous patterns for boys and girls.

results are driven by mean-reversion or ceiling effects. However, this does not seem the case. We also pondered whether one possible explanation for this result is that there are too few boys relative to girls at the top of the ability distribution to properly estimate separate effects for boys and girls in different ability groups, but this does not seem to be the case.[53] Thus, a natural conclusion is that these effects are 'real', and the main question is what could explain them.

One possible explanation is based on 'crowding-out' effects: if we shift the ability distribution so as to have more of the very best top 5% students at school, this might crowd-out students who are in the next ability groups (65^{th}-80^{th} and 80^{th}-95^{th} percentiles) from advanced activities, such as Science and Mathematics 'clubs', or special field trips because of limited space available in such activities. To clarify this, consider that there is usually only a limited number of places available in top-tier activities/clubs for each subject in each school irrespective of cohort size. Under this scenario, having many 'good' peers *in that subject* has two competing effects for regular pupils, in particular for those in the top part of the ability distribution. On the one hand, there could be a positive effect that works either *directly* through interaction of students during instructional time, or *indirectly* via the teaching body (e.g. instructors' motivation). On the other hand, a large share of outstanding peers would reduce one student's chances of getting into the top extra-activities and participating in advanced level learning, thus depressing his/her motivation and ultimately potentially harming achievement. This counter-balancing effect should be more pronounced for the next-to-the-most able students, i.e. pupils in the 65^{th} to 95^{th} percentile of the ability distribution.

One implication of this line of reasoning is that these negative effects should be mitigated in smaller schools. In these schools the positive effect of having many top 5% peers should prevail, since there is at the same time more room for interactions of

---

[53] We further assessed the robustness of our results to the inclusion of bottom 5% and top 5% peers in the estimation sample. Using this extended set of students, the estimates of table 6, row (1) for boys only are -0.080 (s.e. 0.032) for the effect of 'bad' peers, and -0.052 (s.e. 0.026) for the impact of 'good' peers. For girls only, the estimated effect of 'bad' peers is -0.096 (s.e. 0.034), while the impact of top peers is 0.058 (s.e. 0.027). These results are not different from those presented in table 6. We also checked whether the findings in panel B of table 6 are affected by the inclusion of 'good' and 'bad' peers. Results on the extended sample were also virtually identical to the ones presented in table 6, panel B (available upon request from the authors).

pupils of different abilities and less scope for crowding-out of good students from top-tier activities. To check for this possibility, we re-run the analysis displayed in the bottom panel of table 6 on the sample that only includes the 50% smallest schools. Our findings show that for schools in the bottom half of the cohort-size distribution the positive impact of the top 5% peers for girls is positive and roughly constant throughout the ability distribution of regular students. Moreover, the effect of good peers is larger than before for girls in the top one-third of the ability distribution at approximately 0.075, although this estimate is not statistically significant. As for boys, we still find that a large share of top 5% peers at school has a negative impact on regular students, although the effects are now insignificant throughout the ability distribution and smaller in the top percentiles, at approximately -0.052.

This evidence suggests that a crowding-out explanation of our findings might bear some relevance. However, this hypothesis alone cannot easily account for the still markedly different results that we document for males and female, and alternative more subtle explanations discussed in the educational and psychological literature should not be dismissed. In particular, research in these areas has highlighted marked gender differences in behavioral responses to settings that should lead to reciprocity, suggesting that females are more positively and 'cooperatively' influenced by peers and social interactions than males (Cross and Madson, 1997; Eagly, 1978). Additionally, perverse 'big-fish-small-pond' mechanisms have been shown to be more pronounced for males (see Marsh, 2005). Hoxby and Weingarth (2005) refer to these mechanisms as the 'Invidious Comparison Model', where the presence of high performing students depresses the performance of other male students close-by in the ability distribution, possibly through envy, reduced motivation and self-esteem. On the other hand, a 'Shining Light'-type model of peer effects – where a few outstanding students can inspire all others to raise their achievement – might explain the response of female students to a larger fraction of 'good' peers (see again Hoxby and Weigarth, 2005). In a very recent paper, Jackson (2009) finds gender heterogeneity in peer effects along the same lines discussed here.

To conclude this section, we look at whether peer effects for boys and girls differ according to the gender of their peers. To do so, we re-compute the fraction of top 5%

and bottom 5% new peers separately for male and female students, and re-run regressions similar to those in panel A of table 6, but including: the fraction of top 5% boys; the fraction of bottom 5% boys; the fraction of top 5% girls; and the fraction of bottom 5% girls. The average quality of peers is controlled for in these regressions, but not split along the gender dimension. This is because we found little evidence that peers of average quality matter for age-14 test scores of boys and girls. Note that the fractions of bottom 5% and top 5% new peers are now computed on very small number of students. Therefore, the statistical significance of our results is less indicative than the sign and magnitude of the coefficients.

These findings are presented in table 7. Panel A tabulates results for boys, whereas panel B deals with girls. Considering first the effect of the bottom 5% students, we find that boys are similarly affected by bad peers of both genders. Although the point estimates are slightly different across peers' gender, a test on the equality of the two coefficients accepts the null. Moreover, the estimated effect sizes are very close, at 0.405 and 0.452 for male and female peers respectively. These capture the percentage effect of one within-pupil standard deviation change in either treatment on the within-pupil standard deviation in age-14 test scores. As for girls, evidence suggests that they are negatively affected by academically weak peers of both genders, although the adverse impact of bad female peers is more marked. Even though an F-test on the equality of the two coefficients accepts the null, the effect size of the bottom 5% female peers is almost twice as big as the one for bad male peers.

At the opposite end of the ability spectrum, we find that boys react more negatively to a large share of academically bright male peers, with an estimated coefficient is -0.073 (s.e. 0.039) corresponding to an effect size of negative 0.600. On the other hand the coefficient on the proportion of outstanding female peers coefficient is -0.034 (s.e. 0.044), with an effect size of negative 0.294. Remarkably, the opposite is true for girls, who respond more positively to bright peers of the same gender. The estimated effect of the top 5% female peers on other girls is 0.077 (s.e. 0.043) with an associated effect size of 0.797, whereas the effect of top 5% boys is as small as 0.037 (s.e. 0.042) with an effect size of 0.286.

6.3 Additional findings: heterogeneity by pupils' eligibility for free school meals

In this section, we briefly examine the heterogeneity of ability peer effects by pupils' eligibility for free school meals (FSM), a proxy for family income. To do so, we follow that same approach used to look at gender differences in treatment effects. Namely, we first look at estimates obtained by pooling pupils of all ability groups, and then further break down peer quality estimates by pupils' own ability. Results are not shown for space reasons, but are available upon request.

Broadly speaking, results do not highlight any significant heterogeneity. Irrespective of pupils' eligibility for FSM, the fraction of 'bad' peers has a significantly negative impact on students' KS3 attainments. This is estimated to be -0.11 (s.e. 0.040) for FSM-eligible students, and -0.08 (s.e. 0.035) for pupils from richer background. On the other hand, we find that average quality of peers and the fraction of 'good' peers do not have any significant effect on students' performance irrespective of their FSM status. Similarly, we find no evidence of heterogeneous effects when we further allow our estimates to vary along the dimension of pupils' ability. The negative effect of bad peers is sizeable and significant throughout for pupils of all aptitudes and irrespective of their FSM eligibility, except for students with KS2 average test scores in the 80^{th}-95^{th} percentile bracket, where the estimated impact remains negative, but turns insignificant. On the other hand, the percentage of top 5% new peers has no significant impact on students' achievements irrespective of their ability and eligibility for free meals. Finally, we do not detect any interesting pattern for the effect of average peer quality. All in all, we find no evidence of heterogeneous peer effects along the dimension of family income.

7 Concluding remarks and some policy implications

In this paper, we have estimated ability peer effects in schools using data for all secondary schools in England for four cohorts of age-14 (9^{th} grade) pupils and measuring peers' quality by their academic ability as recorded by test scores at age-11 (6^{th} grade). In order to shed some light on the nature of peer effects, we have estimated both the effect of average peer quality, as well as the effect of being at school with a high proportion of very low-ability and very high-ability pupils, on the cognitive outcomes of regular students. Our analysis is highly relevant because of its strong external validity: our data includes over 90 percent of four cohorts of pupils in England that transit from primary school through to the third year of secondary schooling, and sit for two crucial standardized national tests, namely the Key Stage 2 (6^{th} grade) and Key Stage 3 (9^{th} grade). Additionally, our sample is large enough to allow us to recover a variety of estimates about the heterogeneity of our treatment effects.

From a methodological perspective, we view our main contribution as twofold. Firstly, we measure peer ability by test scores that directly capture the cognitive ability of pupils and that are pre-determined with respect to peer interactions in secondary schools, since they are measured at the end of primary education before pupils change schools to start their secondary education. Moreover, by focusing only on peer quality measures based on new peers in secondary schools we by-pass reflection problems. Secondly, we offer a new approach to measuring peer effects, by focusing on within-pupil variation in performance across multiple subjects in a setting where peers' quality is also measured by the variation in their ability across subjects. By using student fixed-effects estimation we are simultaneously able to control for family, school-by-cohort and other cohort-specific unobserved shocks, as well as for pupil ability that is constant across subjects. Our findings strongly suggests that the within-pupil specifications take care of most of the sorting of pupils and their peers into secondary education, and provide reliable causal estimates of ability peer effects. However, to further support this claim, we have provided an extensive battery of robustness checks and falsification exercises that lend additional credibility to the causal interpretation of our results.

In terms of findings, our results clearly show that a large fraction of 'bad' peers at school as identified by students in the bottom 5% of the ability distribution is detrimental to other pupils' learning. On the other hand, we uncover little evidence that the average peer quality and the fraction of very 'good' peers as identified by students in the top 5% of the ability distribution affects the educational outcomes of other pupils across the board. However, these findings mask a significant degree of heterogeneity along the gender dimension. One striking result is that the very brilliant peers at school negatively impact the academic performance of boys, and in particular those who are among the highest groups at school in terms of ability. On the other hand, girls benefit more from having high achievers at school, although there is some evidence that the most able ability girls among regular students at school benefit the least from these interactions.

More in details, we have shown that a 10 percentile decrease in the proportion of 'bad' peers at school implies an improvement of approximately 0.07/0.09 of the within-pupil standard deviation of age-14 test scores for both boys and girls. On the other hand, a 10 percentage point increase in the percentage of 'good' peers would imply an improvement of 0.06 in the within-pupil standard deviation of KS3 achievements for girls, and a nearly symmetrical negative effect for boys of 0.05. These differences become more remarkable if we consider boys and girls of different abilities. For the most talented males, a 10 percentage point increase in the proportion of top 5% peers implies up to 0.09 decrease in the within-pupil standard deviation of age-14 test scores. In sharp contrast, this same increase would boost achievements by more than 0.11 for the least able girls.

These heterogeneous patterns allow us to perform two concluding thought 'policy experiments'. To begin with, suppose that our regular students were exposed to the following two treatments simultaneously: a reduction in the percentage of top 5% and bottom 5% new peers from 20% (the maximum in our data) to zero (the minimum in our data). This change can be viewed as a move towards class homogeneity in terms of ability, i.e. a sort of tracking. This shift would unambiguously improve students' achievements by up to 20% of the standard deviation in the within-pupil distribution of KS3.

Interestingly, this effect is not dissimilar for the most and least able boys, and is only slightly larger than the findings in Duflo et al. (2008) who document a 0.14 standard deviation improvement in the test score of pupils in primary schools in Kenya after 18 months of random assignment to homogenous 'tracked' classes. On the other hand, our thought-experiment would give more heterogeneous results for girls. On average, the shift would improve female students' age-14 achievements by about 0.06 of a (within-pupil) standard deviation. This overall positive effect is the sum of the positive impact of not interacting with academically weak peers (at +0.18) and the adverse effect of reduced interactions with the best peers (-0.12). However, this overall effect would turn negative for girls in the bottom half of the ability distribution, with regular students in the ability bracket of 20th-35th percentile loosing out as much as 0.10 of a (within-pupil) standard deviation. At the other extreme, the most talented girls could gain more than 0.20 of a (within-pupil) standard deviation of age-14 achievements from being educated in homogeneous environments with negligible fractions of 'bad' and 'good' peers.

Similarly, this finding is not dissimilar from Carrell et al. (2011) who provide experimental evidence that 'middle ability' students assigned to more homogeneous classes experience improvements in their university GPA by around 11% of a standard deviation. However, our results mask some significant heterogeneity along the gender dimension. Whereas male students' achievements unambiguously improve by up to 25% of one standard deviation, this policy experiment would not improve female students' age-14 achievements on average: the positive impact of not interacting with academically weak peers – at 24% of a standard deviation – would be almost perfectly matched by the adverse effect of reduced interactions with the 'good' peers at -23%. Nevertheless, the effect would clearly turn negative for girls in the bottom half of the ability distribution, and substantially positive for the most talented girls.

Another policy-relevant experiment would be to simulate the effects of tracking by grouping all students – including the bottom 5% and top 5% – into two classes perfectly segregated along the lines of student's ability. The first group would include pupils who are above the median of the ability distribution, and the second those below the median. In this case, the lower ability group will experience a doubling of the proportion of bottom 5% pupils, on average from 4% to 8%, and a decline of the proportion of top 5% pupils from about 4% to zero. For the high ability class, the opposite will occur as the proportion of top 5% pupils doubles to about 8% and the proportion of bottom 5% falls to zero. These shifts would unambiguously worsens students' KS3 achievements in the low ability group, with a negative impact of about -0.03 in the within-pupil standard deviation of KS3 for boys, and -0.06 (-0.04-0.02) for girls. On the other hand, the changes experienced in the high ability group would improve boys' KS3 achievements by at most 0.01 (0.03-0.02) of a within-pupil standard deviation of KS3, while girls would benefit by up to 0.06 (0.04+0.02).

Do our results lend overall support to tracking of students by ability? Besides any equity consideration, we have shown that there is no simple answer to this question from an efficiency-of-learning point of view: our results are clearly heterogeneous in relation to one pupils' ability and gender, and vary according to the exact details of the tracking-experiment being carried out. Nevertheless, although we are fully aware of the difficulties of using reduced-form estimates to make out-of-sample policy predictions (see Carrell at al., 2011), we believe our data is rich enough – and our findings robust enough – to provide a solid ground for insightful interventions targeting students' ability mix as a means to improve learning standards.

8 References

Aizer Anna, (2008). Peer Effects and Human Capital Accumulation: The Externalities of ADD, NBER Working Paper No. 14354.

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools, *Journal of Political Economy*, 113, 151-184.

Ammermueller, Andreas and Jörn-Steffen Pischke (2009). Peer Effects in European Primary Schools: Evidence from PIRLS, *Journal of Labor Economics*, 27, 315-348.

Anderson, Michael L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects, *Journal of the American Statistical Association*, 103(484), 1481-1495.

Angrist, Joshua D. and Kevin Lang, (2004). Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program, *American Economic Review,* 94(5), 1613-1634.

Angrist, Joshua D., Daniel Lang and Philip Oreopolous (2009). Incentives and Services for College Achievement: Evidence from a Randomized Trial, *American Economic Journal: Applied Economics*, 1(1), 136-163.

Angrist, Joshua D. and Victor Lavy, (2009). The Effect of High-Stakes High School Achievement Awards: Evidence from a Randomized Trial, *American Economic Review*, 99(4), 1384-1414.

Arcidiacono, Peter and Sean Nicholson (2005). Peer Effects in Medical School, *Journal of Public Economics*, 89(2-3), 327-350.

Arcidiacono, Peter, Gigi Foster, Natalie Goodpaster and Josh Kinsler (2009). Estimating Spillovers with Panel Data, with and Application to the Classroom, mimeo, Duke University.

Bandiera, Oriana , Iwan Barankay, Imran Rasul (forthcoming). Social Incentives in the Workplace, *Review of Economic Studies*.

Bandiera, Oriana, Valentino Larcinese and Imran Rasul (2009). Heterogeneous Class Size Effects: New Evidence from a Panel of University Students, IZA Discussion Paper No. 4496.

Becker, Gary S. (1974). A Theory of Social Interactions, *Journal of Political Economy*, 82(6), 1063-1093.

Bifulco, Robert, Jason M. Fletcher and Stephen L. Ross (2008). The Effect of Classmate Characteristics on Individual Outcomes: Evidence from the Add Health, University of Connecticut, Department of Economics Working Paper No. 2008-21.

Black, Sandra E. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education, *Quarterly Journal of Economics,* 114(2), 577–99.

Black, Sandra .E., Paul J. Devereux and Kjell G. Salvanes (2009). Long After They're Gone? The Effects of Peers on Outcomes of Young Adults, paper presented at the CEPR/IZA European Summer Symposium in Labour Economics (ESSLE), Buch/Ameersee.

Burke, Mary A. and Tim R. Sass (2008). Classroom Peer Effects and Student Achievement, Federal Reserve Bank of Boston Working Paper No. 08-5.

Carrell, Scott E. and Mark L. Hoekstra (2008). Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids, NBER Working Paper No. 14246.

Carrell, Scott E., B. I. Sacerdote, and J. E. West (2011) "From Natural Variation to Optimal Policy? The Lucas Critique Meets Peer Effects." NBER Working Papers 16865.

Coleman, James S. (1966). Equality of Educational Opportunity, Washington D.C., U.S., Department of Health, Education and Welfare, Office of Education, OE-38001.

Cross, Susan and Laura Madson (1997). Models of the Self: Self-Construals and Gender, *Psychological Bulletin*, 12, 5-37.

Dee, Thomas S. (2007). Teachers and Gender Gaps in Student Achievement, *Journal of Human Resources*, 42(3), 528-554,

De Giorgi, Giacomo, Michele Pellizzari and Silvia Redaelli (2009). Be As Careful of the Company that You Keep As of the Books You Read: Peer Effects in Education and on the Labor Market, NBER Working Paper No. 14948.

DfES (2006). Grouping Pupils for Success, *Primary and Secondary National Strategies*, Department for Education and Skills, London.

Duflo Esther, Pascaline Dupas and Michael Kremer (2008). Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya, CEPR Discussion Paper No. DP7043.

Eagly, Alice H. (1978). Sex Differences in Influenceability, *Pshycological Bulletin*, 85, 86-116.

Ellison, Glenn and Ashley Swanson (2009). The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions, NBER Working Paper No. 15238.

Epple, Dennis and Richard Romano (2010). Peer Effects in Education: A Survey of the Theory and Evidence, manuscript prepared for the *Handbook of Social Sciences*.

Figlio, David N. (2007). Boys Named Sue: Disruptive Children and Their Peers, *Education Finance and Policy*, 2(4), 376-394.

Fryer, Roland G. and Steven D. Levitt (forthcoming). An Empirical Analysis of the Gender Gap in Mathematics, *American Economic Journal: Applied Economics*.

Gibbons, Stephen, Stephen Machin and Olmo Silva (2009). Valuing School Quality Using Boundary Discontinuity Regressions, London School of Economics, mimeo.

Gibbons, Stephen and Shqiponja Telhaj (2008). Peers and Achievement in England's Secondary Schools, SERC Discussion Paper 1.

Glennerster, Howard (1991). Quasi-Markets for Education, *Economic Journal*, 101, 1268-76.

Gould, Eric D., Victor Lavy and Daniele M. Paserman (2009a). Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence, *Economic Journal*, 119, 1243-1269.

Gould Eric D., Victor Lavy, and Daniele M. Paserman (2009b). Sixty Years after the Magic Carpet Ride: The Long-Run Effect of the Early Childhood Environment on Social and Economic Outcomes, NBER Working Paper No. 14884.

Goux, Dominique and Eric Maurin (2007). Close Neighbours Matter: Neighbourhood Effects on Early Performance at School, *Economic Journal*, 117, 1-24.

Hanushek, Eric A., John F. Kain, Jacob M. Markman and Steven G. Rivkin (2003). Does Peer Ability Affect Student Achievement?, *Journal of Applied Econometrics*, 18(5), 527-544.

Hoxby, Carloine M. (2000). Peer Effects in the Classroom: Learning from Gender and Race Variation, NBER Working Paper No. 7867.

Hoxby, Caroline M., and Gretchen Weingarth (2005). Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects, mimeo, [online] http://www.ksg.harvard.edu/inequality/Seminar/Papers/Hoxby06.pdf.

Jackson, K. (2009). Peer Quality or Input Quality? Evidence from Trinidad and Tobago, mimeo, Cornell University.

Kane, Thomas J., Stephanie K. Riegg and Douglas O. Staiger (2006). School Quality, Neighborhoods and Housing Prices, *American Law and Economics Review*, 8(2), 183-212.

Katz, Lawrence F., Jeffrey R. Kling and Jeffrey B. Liebman (2001). Moving to Opportunity in Boston: Early Results from a Randomized Mobility Experiment, *Quarterly Journal of Economics*, 116(2), 607-654.

Kutnick, Peter, Judy Sebba, Peter Blatchford, Maurice Galton and Jo Thorp (2005). The Effects of Pupil Grouping: Literature Review, Department for Education and Skills Research Report 688.

Kutnick, Peter, Steve Hodgekinson, Judy Sebba, Sara Humphreys, Maurice Galton, Susan Steward, Peter Bltachford and Ed Baines (2006). Pupil Grouping Strategies and Practices at Key Stage 2 and 3: Case Studies of 24 Schools in England, Department for Education and Skills Research Report 796.

Lavy, Victor (2009). The Causal Effect of Instructional Time on Achievements in Math, Science and Reading: International Evidence, paper presented at the *Economics of Education and Education Policy in Europe* (EEEPE) Network Final Conference.

Lavy, Victor and Analia Schlosser (2008). Mechanisms and Impacts of Gender Peer Effects at School, NBER Working Paper No. 13292.

Lavy, Victor, Daniele M. Paserman and Analia Schlosser (2008). Inside the Black Box of Ability Peer Effects: Evidence from Variation in Low Achievers in the Classroom, NBER Working Paper No. 14415.

Lavy, Victor, Olmo Silva and Felix Weinhardt (2009). The Good, the Bad and the Average: Evidence on the Scale and Nature of Ability Peer Effects in Schools, NBER Working Papers 15600, National Bureau of Economic Research, Inc.

Lazear, Edward P. (2001). Educational Production, *Quarterly Journal of Economics*, 116(3), 777-803.

Lefgren, Lars (2004). Educational Peer Effects and the Chicago Public Schools, *Journal of Urban Economics*, 56(2), 169-191.

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

Marsh, Herbert W. (2005). Big Fish Little Pond Effect on Academic Self-concept: Cross-cultural and Cross-Disciplinary Generalizability, paper presented at the AARE Conference Paper.

Moffit, Robert A. (2001). Policy Interventions, Low-Level Equilibria, and Social Interactions, in *Social Dynamics*, S. N. Durlauf and H. Young (eds.), Cambridge MA: MIT, 45-82.

Nechyba, Thomas and Jacob Vigdor (2007). Peer Effects in North Carolina Public Schools, in *Schools and the Equal Opportunity Problem*, Ludger Woessman and Paul E. Peterson (eds.), Cambridge MA: MIT, 73-102.

Sacerdote, Bruce (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates, *Quarterly Journal of Economics*, 116(2), 681-704.

Sanbonmatsu, Lisa, Jeffrey R. Kling, Greg J. Duncan and Jeanne Brooks-Gunn (2006). Neighborhoods and Academic Achievement: Results from the Moving to Opportunity Experiment, *Journal of Human Resources*, XLI(4). 649-691.

Vigdor, Jacob and Thomas Nechyba (2004). Peer Effects in Elementary School: Learning from 'Apparent' Random Assignment, Duke University, mimeo.

West, Anne and Audrey Hind (2003), Secondary school admissions in England: Exploring the extent of overt and covert selection, Report for Research and Information on State Education Trust, www.risetrust.org.uk/admissions.html.

Zimmerman, David J. (2003). Peer Effects in Academic Outcomes: Evidence from a Natural Experiment, *Review of Economics and Statistics*, 85(1), 9-23.

Table 1 – Descriptive statistics: pupils' outcomes, pupils' background and school characteristics

Variable	Regular students	At least 1 subject top 5%	At least 1 subject bottom 5%
Panel A: Pupils' outcomes			
KS2 percentile, English	49.3 (24.3)	87.1 (14.8)	8.5 (12.5)
KS2 percentile, Mathematics	49.4 (24.3)	87.0 (14.1)	9.4 (13.6)
KS2 percentile, Science	48.9 (24.3)	87.7 (13.1)	10.9 (15.5)
KS3 percentile, English	48.9 (26.0)	81.2 (18.6)	15.3 (18.2)
KS3 percentile, Mathematics	49.2 (25.3)	84.5 (16.3)	14.8 (17.6)
KS3 percentile, Science	49.2 (25.5)	84.4 (16.2)	16.0 (17.9)
Panel B: Pupils' characteristics			
First language is English	0.93 (0.253)	0.95 (0.21)	0.89 (0.31)
Eligible for free school meals	0.13 (0.337)	0.05 (0.22)	0.30 (0.46)
Male	0.50 (0.500)	0.48 (0.50)	0.55 (0.50)
Changed school between Year 7 and KS3	0.11 (0.313)	0.09 (0.29)	0.14 (0.35)
Ethnicity: White British	0.85 (0.35)	0.88 (0.32)	0.81 (0.39)
Ethnicity: White other	0.02 (0.12)	0.02 (0.13)	0.02 (0.14)
Ethnicity: Asian	0.05 (0.22)	0.03 (0.18)	0.07 (0.26)
Ethnicity: Black	0.03 (0.16)	0.01 (0.11)	0.04 (0.19)
Ethnicity: Chinese	0.00 (0.05)	0.00 (0.07)	0.00 (0.04)
Ethnicity: Other	0.05 (0.22)	0.07 (0.21)	0.06 (0.24)
Panel C: School characteristics (Year 7)			
Cohort size	201.7 (57.2)	204.1 (56.3)	198.8 (58.5)
Community school	0.67 (0.47)	0.63 (0.48)	0.73 (0.44)
Religiously affiliated school	0.16 (0.37)	0.19 (0.39)	0.11 (0.32)

Note: Table report means of the listed variables and standard deviation in parenthesis. Number of regular pupils: approximately 1,200,000. The sample of regular students only includes pupils with KS2 achievement in each subject above the 5th percentile and below the 95th percentile of KS2 cohort-specific national distribution. Number of pupils with at least one subject in top 5% (≥95th percentile of KS2 cohort-specific national distribution): approximately 170,000. Number of pupils with at least one subject in bottom 5% (≤5th percentile of KS2 cohort-specific national distribution): approximately 130,000. Year 7 refers to the first year in secondary school after transition out of primary. KS3 refers to Year 9 when pupils sit for their KS3 assessment. Community schools include only secular comprehensive state schools. Religiously affiliated schools include only schools in the state sector with some religious affiliation. Fractions may not sum to 1; this is due to rounding or partially missing information.

Table 2 – Descriptive statistics of treatments: average KS2 achievements and percentages of pupils in top 5% and bottom 5% of KS2 ability distribution – *new* peers only

Variable	Mean	Std. dev.	Min	Max
Panel A: Average KS2 percentile treatment (new peers)				
Average peer achievement at KS2 in English	49.79	8.71	1	98
Average peer achievement at KS2 in Math	49.94	8.06	1	100
Average peer achievement at KS2 in Science	49.68	8.35	1	100
Panel B: Top 5% treatment (new peers)				
Percentage, top 5% in English	4.22	3.03	0	19.56
Percentage, top 5% in Maths	3.77	2.60	0	19.87
Percentage, top 5% in Science	3.91	2.75	0	19.86
Panel C: Bottom 5% treatment (new peers)				
Percentage, bottom 5% in English	3.79	2.78	0	19.30
Percentage, bottom 5% in Maths	3.81	2.67	0	19.86
Percentage, bottom 5% in Science	3.78	2.90	0	19.78
Percentage of new peers for pupils in Year 7	87.56	22.66	0	1

Note: Treatment measured in Year 7 when students start secondary school after transition from primary. New peers refers to students in Year 7 in a given cohort that do not come from the same primary school.

Table 3 – Impact of peer quality on KS3 educational attainments: main results

	Average peer KS2		*Percentage of bottom 5% pupils*		*Percentage of top 5% pupils*	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable is:	OLS	Within-pupil	OLS	Within-pupil	OLS	Within-pupil
KS3 percentiles, unconditional on KS2; treatments entered separately	0.359 (0.012)**	0.022 (0.012)	-0.958 (0.029)**	-0.120 (0.033)**	0.750 (0.028)**	0.003 (0.003)
KS3 percentiles, unconditional on KS2; all treatments together	0.191 (0.013)**	0.018 (0.013)	-0.592 (0.032)**	-0.095 (0.033)**	0.332 (0.029)**	-0.021 (0.026)
KS3 percentiles, controlling for KS2 same-subject interacted with subject dummies; all treatments together	0.161 (0.011)**	0.012 (0.012)	-0.566 (0.030)**	-0.091 (0.033)**	0.244 (0.027)**	0.005 (0.026)
KS3 percentiles, controlling for KS2 same- and cross-subject interacted with subject dummies; all treatments together	0.146 (0.011)**	0.010 (0.012)	-0.511 (0.030)**	-0.089 (0.033)**	0.227 (0.027)**	0.008 (0.026)

Note: The table reports regression coefficients and standard errors in round brackets from regressions of the dependent variable on treatments. Standard error clustered at the school level. **: at least 1% significant. Treatment effects in the first row estimated from two different sets of regressions: one including the average peer achievement at KS2 only (columns (1) and (2)); and one including the percentage of top 5% pupils and the percentage of bottom 5% pupils in the cohort only (columns (3) to (6)). All other regressions include all three treatments together. The table displays the coefficients on treatments based on new peers only. All regressions control for quality of old peers, and include subject and subject-by-gender dummies. Pupil characteristics controlled for in columns (1), (3) and (5); absorbed in columns (2), (4) and (6). Number of observations: approx. 3,600,000 (1,200,000 pupils), in 2193 schools.

Table 4 – Impact of peer quality on KS3 educational attainments: robustness to potential threats to identification and results for small schools only

	Within-pupil estimates		
	Average peer KS2	*Percentage of bottom 5% pupils*	*Percentage of top 5% pupils*
Dependent variable is:	(2)	(4)	(6)
KS3 percentiles, controlling for KS2: excluding specialist schools	0.010 (0.013)	-0.091 (0.034)**	0.013 (0.027)
KS3 percentiles, controlling for KS2: undersubscribed schools (excluding specialist)	0.013 (0.017)	-0.100 (0.040)**	0.011 (0.037)
KS3 percentiles, controlling for KS2: sample of pupils whose best subject is different from the best subject of new peers (*mixed*)	0.008 (0.013)	-0.100 (0.034)**	0.014 (0.027)
KS3 percentiles, unconditional on KS2: pupils in 50% smallest schools	0.005 (0.013)	-0.109 (0.045)**	0.028 (0.038)
KS3 percentiles, controlling for KS2: pupils in 50% smallest schools	-0.002 (0.013)	-0.104 (0.044)**	0.063 (0.038)
KS3 percentiles, controlling for KS2: pupils in 50% smallest schools; including pupils fixed effects and school × subject fixed effects	0.003 (0.006)	-0.070 (0.021)**	0.006 (0.017)

Note: All specifications as in row (4) of table 3, except in row (4) where the specification does not control for lagged test scores. Specification in rows (6) further includes school-by-subject fixed effects. Specialist schools account for about 8.5% of the pupil sample. Undersubscribed schools enrol approximately 60% of pupils in non-specialist schools. Sample of pupils with different best subject from new peers in school account for about 60% of the full sample. Sample of pupils in 50% smallest schools includes pupils in schools with less than 181 students in the year 7 cohort (approx. 6 classes of max 30 students). Regression with school × subject fixed effect (row (6)) only considers the first cohort (year 7 in 2002) and last cohort (year 7 in 2005). Standard error clustered at the school level, except rows (6) where they are robust. **: at least 1% significant.

Table 5 – Impact of peer quality on KS3 attainments: by pupil's ability

Dependent variable is: KS3, controlling for KS2	Within-pupil estimates		
	Average peer KS2	*Percentage of bottom 5% pupils*	*Percentage of top 5% pupils*
	(1)	(2)	(3)
Effect for percentiles 5-20	0.011	-0.081	0.016
	(0.012)	(0.029)**	(0.026)
Effect for percentiles 20-35	0.010	-0.068	0.044
	(0.015)	(0.035)*	(0.030)
Effect for percentiles 35-50	0.012	-0.074	0.023
	(0.015)	(0.041)§	(0.032)
Effect for percentiles 50-65	0.011	-0.118	0.020
	(0.016)	(0.043)**	(0.033)
Effect for percentiles 65-80	0.005	-0.114	-0.027
	(0.015)	(0.043)**	(0.031)
Effect for percentiles 80-95	0.016	-0.038	-0.049
	(0.017)	(0.050)	(0.032)
F-Test: all coeffs. jointly equal to zero (p-value)	*0.9317*	*0.0321*	*0.1311*
F-Test: all coefficients are equal (p-value)	*0.9816*	*0.2564*	*0.0801*

Note: The table reports regression coefficients and standard errors in round brackets from regressions of the dependent variable on treatments. Treatment effects estimated from one single regression including all three treatments together. The table displays the coefficient on treatments based on new peers. All regressions control for the quality of old peers. Interaction terms obtained by interacting the peer quality measures (separately for old and new peers) with a dummy indicating where the pupil ranks in terms of his/her KS2 percentiles *on average across subjects*. Ability blocks are defined using original KS2 percentiles computed out of the cohort-specific national distribution. The effect of KS2 achievement (same- and cross-subject) is controlled for semi-parametrically by interacting pupil KS2 percentiles with the dummies indicating his/her rank in the ability distribution (and in interaction with subject dummies). Specifications further include subject and subject-by-gender dummies. Number of observations: approximately 3,600,000 (1,200,000 pupils), in 2193 schools. Standard error clustered at the school level. **: at least 1% significant; *: at least 5% significant; §: at least 10% significant.

Table 6 – Impact of peer quality on KS3 attainments, by pupil's ability and gender

Dependent variable is: KS3, controlling for KS2	Within-pupil estimates			
	Boys only		Girls only	
	Percentage of bottom 5% pupils	*Percentage of top 5% pupils*	*Percentage of bottom 5% pupils*	*Percentage of top 5% pupils*
	(1)	(2)	(3)	(4)
Panel A: Pupils of ability pooled (overall effect)				
Overall effect	-0.076	-0.052	-0.098	0.066
	(0.035)*	(0.028)§	(0.037)**	(0.029)*
Panel B: Ability blocks defined on original KS2 percentiles				
Effect for percentiles 5-20	-0.093	-0.013	-0.080	0.066
	(0.032)**	(0.029)	(0.038)*	(0.035)§
Effect for percentiles 20-35	-0.057	-0.037	-0.072	0.126
	(0.039)	(0.033)	(0.044)§	(0.037)**
Effect for percentiles 35-50	-0.068	-0.059	-0.066	0.088
	(0.046)	(0.036)	(0.047)	(0.039)*
Effect for percentiles 50-65	-0.106	-0.036	-0.113	0.062
	(0.048)*	(0.038)	(0.050)*	(0.038)§
Effect for percentiles 65-80	-0.089	-0.079	-0.139	0.023
	(0.051)§	(0.037)*	(0.050)**	(0.036)
Effect for percentiles 80-95	0.036	-0.096	-0.116	0.011
	(0.065)	(0.043)*	(0.060)*	(0.039)
F-Test: all coeff. jointly equal to zero (p-value)	*0.0425*	*0.2642*	*0.1042*	*0.0334*
F-Test: all coefficients are equal (p-value)	*0.2597*	*0.4281*	*0.6809*	*0.0766*

Note: Specifications in panel A as in row (4) of table 3; specifications in panel B as in table 5. Separate regressions run for boys and girls. Number of observations for boys: approx. 1,800,000 (600,000 pupils) in 2101 schools. Number of observations for girls: approx. 1,800,000 (600,000 pupils) in 2134 schools. Standard error clustered at the school level. **: at least 1% significant; *: at least 5% significant; §: at least 10% significant.

Table 7 – Impact of peer quality on KS3 attainments: treatments separately defined by pupils' gender

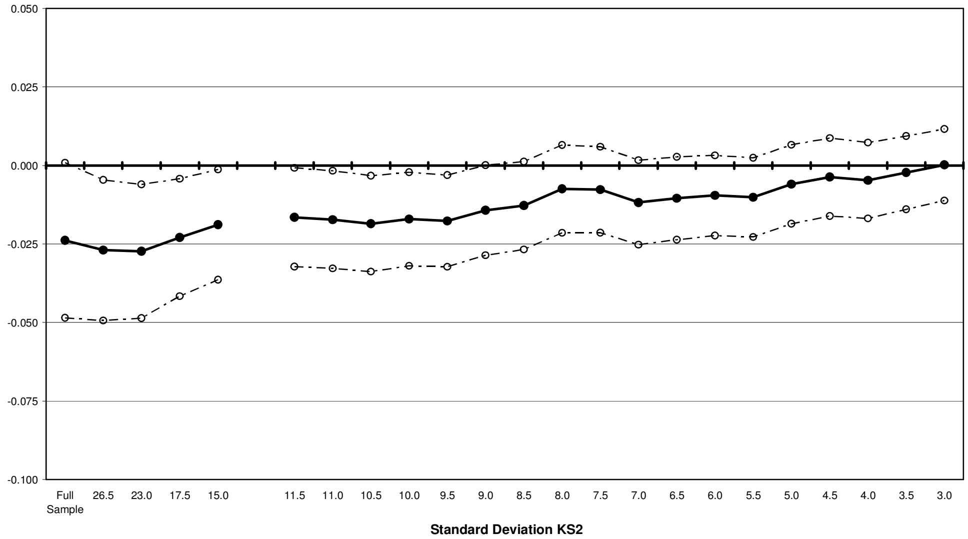
	Within-pupil estimates			
	Percentage of bottom 5% pupils		*Percentage of top 5% pupils*	
	Counting male pupils only	Counting female pupils only	Counting male pupils only	Counting female pupils only
Dependent variable is:	(1)	(2)	(3)	(4)
<u>*Panel A: Boys only*</u>				
KS3 percentiles, controlling for KS2	-0.065	-0.090	-0.073	-0.034
	(0.049)	(0.060)	(0.039)§	(0.044)
Effect size	*0.405*	*0.452*	*0.600*	*0.294*
F-Test: coefficients are equal (p-value)	*0.7685*		*0.5364*	
F-Test: coeffs. jointly equal to zero (p-value)	*0.0929*		*0.1121*	
<u>*Panel B: Girls only*</u>				
KS3 percentiles, controlling for KS2	-0.068	-0.124	0.037	0.077
	(0.053)	(0.058)*	(0.042)	(0.043)§
Effect size	*0.414*	*0.755*	*0.286*	*0.797*
F-Test: coefficients are equal (p-value)	*0.4980*		*0.5259*	
F-Test: coeffs. jointly equal to zero (p-value)	*0.0303*		*0.1168*	

Note: The table reports regression coefficients and standard errors in round brackets from regressions of the dependent variable on treatments. Treatment effects estimated from one single regression including both treatments. The table displays the coefficient on treatments based on new peers and computed separately for male and female pupils. All regressions control for the quality of old peers computed separately for male and female pupils, and for the average quality of new and old peers. Controls further include KS2 percentiles in same- and cross-subject in interaction with subject dummies included, as well as subject dummies. Effect size (in *italics*) refer to the effect of a one standard deviation of the within-pupil distribution of peers as a percentage of one standard deviation of the within-pupil distribution of KS3 percentiles. Number of observations: approximately 1,800,000 (600,000 pupils) in each panel. Number of schools: 2101 in panel A; 2134 in panel B. Standard error clustered at the school level. *: at least 5% significant; §: at least 10% significant.

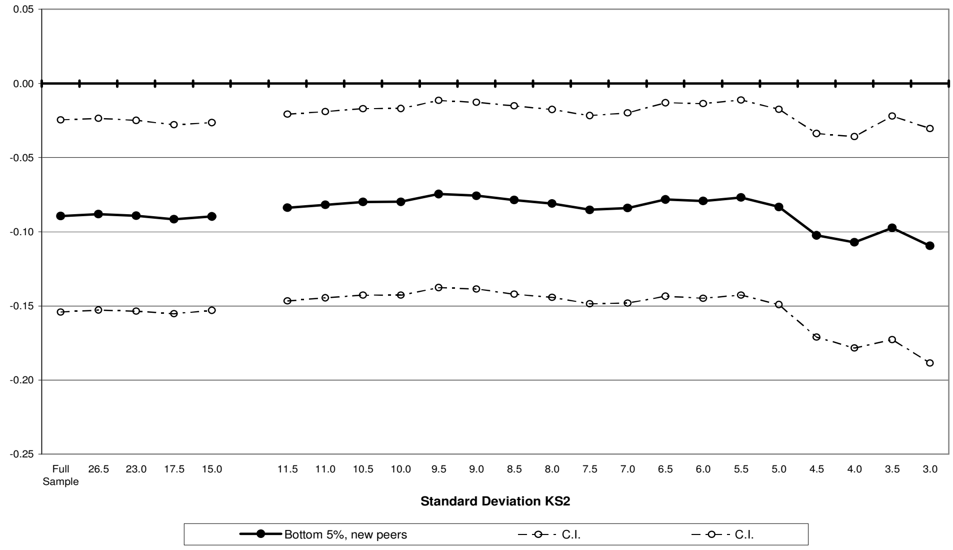
Figures

Figure 1 – Balancing and treatment effects of bottom 5% peers; by cumulative bands of the within-pupil standard deviation of KS2 scores

Balancing of KS2

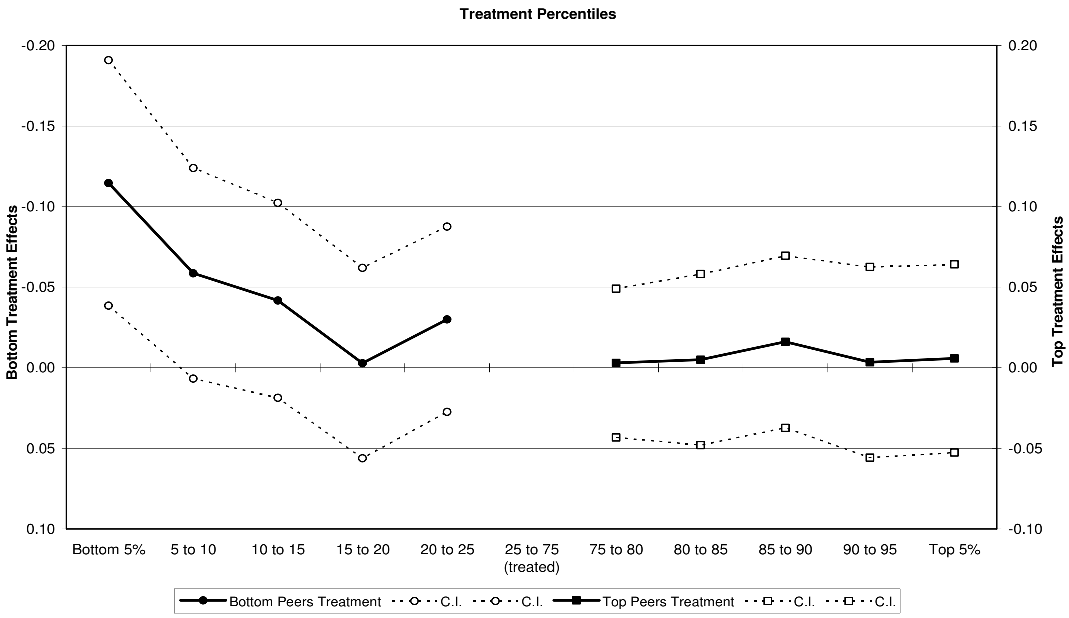


Treatment effects on KS3



Note: The figure plots regression coefficients and 95% confidence intervals (standard errors clustered at the school level) obtained regressing pupil KS2 achievements (top panels) and KS3 achievements (bottom panels) on the percentage of bottom 5% new peers. Regressions include: pupil fixed-effects; subject and subject-by-gender dummies; fraction of top 5% new peers and average quality of new peers; control for old peer quality. Regressions in the bottom panel further include pupil KS2 achievement in same- and cross -subject interacted with subject dummies. 23 different regressions were estimated over different cumulative bands of the standard deviation of KS2 attainments across subjects; these spanned the interval std.dev.≤3 to std.dev.≤11.5, in steps of 0.5, and then std.dev.≤15; std.dev.≤17.5; std.dev.≤23; std.dev.≤26; full sample.

Figure 2 – Treatment effects on KS3 percentiles; by different percentile cut-off points for top and bottom peers



Note: The figure plots regression coefficients and 95% confidence intervals (standard errors clustered at the school level) obtained regressing pupil KS3 achievements on the following treatments: percentage of top 5% new peers; percentage of top 5-to-10% new peers; percentage of top 10-to15% new peers; percentage of top 15-to-20% new peers; percentage of top 20-to25% new peers; percentage of bottom 5% new peers; percentage of bottom 5-to-10% new peers; percentage of bottom 10-to15% new peers; percentage of bottom 15-to-20% new peers; percentage of bottom 20-to25% new peers. The regression further includes: pupil fixed-effects; pupil KS2 achievement in same- and cross-subject interacted with subject dummies; average new peer quality; controls for old peer quality; subject and subject-by-gender dummies. Treated pupils include students with KS2 achievements between 25th and 75th percentile of the cohort-specific distribution of KS2 for every subjects. Number of observations: approx. 2,580,000 (860,000 pupils) in 2193 schools.

Appendix Tables

Appendix Table 1 – Within and between variation in pupil test scores and treatment measures

	Regular students				Sample including boys only				Sample including girls only			
Variable:	*Mean*	*Overall Std.dev.*	*Between Std.dev.*	*Within Std.dev.*	*Mean*	*Overall Std.dev.*	*Between Std.dev.*	*Within Std.dev.*	*Mean*	*Overall Std.dev.*	*Between Std.dev.*	*Within Std.dev.*
KS2 percentiles	49.19	24.31	21.15	11.98	48.71	24.38	21.10	12.21	49.66	24.22	21.19	11.73
KS3 percentiles	49.10	25.61	22.99	11.29	48.02	25.76	23.02	11.56	50.19	25.42	22.92	11.01
Average peer achievement at KS2	49.80	8.38	7.96	2.61	49.79	8.38	7.97	2.58	49.82	8.377	7.95	2.63
Percentage, bottom 5%	3.79	2.78	2.62	0.94	3.79	2.80	2.64	0.93	3.79	2.77	2.61	0.94
Percentage, top 5%	3.97	2.81	2.49	1.29	3.96	2.79	2.49	1.27	3.98	2.82	2.50	1.32

Note: Number of observations in the sample of regular students: approximately 3,600,000 corresponding to 1,200,000 pupils and 3 subjects. Number of observations in samples of boys and girls only: approximately 1,800,000 corresponding to 600,000 pupils and 3 subjects. Peer quality measures refer to new peers only.

Appendix Table 2 – Balancing of indiv. characteristics with respect to treatments

	Average achievement at KS2 (percentiles)	*Percentage of top 5% pupils*	*Percentage of bottom 5% pupils*
Dependent variable:	(1)	(3)	(5)
Panel A: OLS regression results			
First language is English (x100)	0.054 (0.024)*	-1.287 (0.150)**	-0.086 (0.061)
Eligible for free school meals (x100)	-0.161 (0.018)**	1.395 (0.074)**	-0.094 (0.040)*
Changed school between Year 7 and KS3 (x100)	-0.280 (0.056)**	-1.763 (0.118)**	-1.284 (0.134)**
Ethnicity: White British (x100)	0.056 (0.030)*	-1.493 (0.184)**	-0.138 (0.094)
Ethnicity: White other (x100)	0.007 (0.005)	0.031 (0.025)	0.036 (0.017)*
Ethnicity: Asian (x100)	-0.025 (0.019)	1.077 (0.131)**	0.083 (0.057)
Ethnicity: Black (x100)	-0.025 (0.008)**	0.321 (0.060)**	-0.022 (0.029)
Ethnicity: Chinese (x100)	-0.001 (0.001)	0.004 (0.003)	0.007 (0.003)**
Ethnicity: Other (x100)	-0.012 (0.012)	0.061 (0.042)	0.034 (0.034)
KS2 percentiles	0.059 (0.005)**	-0.103 (0.011)**	0.161 (0.011)**
Panel B: Within-pupil regression results			
KS2 percentiles	0.022 (0.006)**	-0.024 (0.013)	-0.043 (0.010)**

Note: The table reports regression coefficients and standard errors in round brackets from regressions of the dependent variable on treatments. Standard error clustered at the school level. **: at least 1% significant; *: at least 5% significant. All regressions include all three treatments together. The table displays the coefficients on treatments based on new peers only. All regressions control for quality of old peers and include subject and subject-by-gender dummies. Top panel (panel A) presents OLS estimates; the bottom panel (panel B) presents within-pupil estimates. Individual characteristics do not vary within-pupil except for KS2 percentiles. So within-pupil models can only be estimated for the latter variable. Further note that the coefficients and standard errors on the individual characteristics have been multiplied by 100 to improve legibility. Number of observations: approx. 3,600,000 (1,200,000 pupils), in 2193 schools.

Appendix Table 3 – OLS estimates of the impact of peer quality on KS3 educational attainments: by different subjects separately; full sample

	Average achievement at KS2	*Percentage of bottom 5% pupils*	*Percentage of top 5% pupils,*
Dependent variable is:	(1)	(2)	(3)
KS3 percentiles, controlling for KS2 (same- and cross-subject): English only	0.148 (0.013)**	-0.468 (0.037)**	0.165 (0.035)**
KS3 percentiles, controlling for KS2 (same- and cross-subject): Math only	0.138 (0.012)**	-0.513 (0.034)**	0.253 (0.034)**
KS3 percentiles, controlling for KS2 (same- and cross-subject): Science only	0.152 (0.014)**	-0.519 (0.038)**	0.257 (0.036)**

Note: The table reports regression coefficients and standard errors in round brackets from regressions of the dependent variable on treatments. Standard error clustered at the school level. **: at least 1% significant. All regressions include all three treatments together. The table displays the coefficients on treatments based on new peers only. All regressions control for quality of old peers. Pupil characteristics controlled for in columns (1), (2) and (3). Number of observations: approx. 3,600,000 (1,200,000 pupils), in 2193 schools.

CHAPTER V

EVIDENCE ON THE IMPORTANCE OF TIME ZONE ASSIGNMENT

1 Introduction

There are many reasons to believe that the timing of daylight matters for individual utility and welfare. Humans do not usually derive joy from sitting in the dark and we know that daylight has important impacts on health outcomes, i.e. van den Berg (2005). This study estimates the effect of the timing of daylight on electricity consumption. Electricity consumption presents an interesting case because US households spent 125 billion US-$ for electricity in 2005 alone (USdOT, 2010), which corresponds to about one per cent of total GDP. I argue that a better understanding of the effect of the timing of daylight on electricity consumption could potentially result in significant cost savings and welfare improvements.

Surprisingly, we know very little about the economic effects of local time and daylight on human activity. This is because for a given location, local times of sunrise and sunset only vary in a very smooth pattern over the year, which makes credible empirical estimation difficult. In terms of variations in local time, the exceptions are changes due to daylight-savings time (DST), and this variation has indeed been used to estimate effects on residential energy consumption (USdOT, 1975; Rock, 1997; CEC, 2001; Kandel, 2007; Kotchen and Grant, 2012; Kellogg and Wolff, 2008).[1] The latter two are empirical studies that focus on local changes in DST regimes in Australia and the state of Indiana and use a local difference-in-difference approach for estimation. Overall the results are inconclusive.

Contrary to Kotchen and Grant (2012) and Kellogg and Wolff (2008), I argue that rather than focussing on local changes in DST regimes, nationwide geo-temporal variation in local times of sunrise can be used for estimation. My approach has the advantage that I can fully examine the heterogeneity in the response of residential energy consumption across different latitudes and climate zones.

To obtain credible estimates for the elasticity of residential energy consumption with respect to local sunrise times, I use geo-temporal variation that has never, to my knowledge, been used before. To do this, I use Geographical Information Systems (GIS) to calculate the exact solar time for each county of the mainland US. This allows me to compute the length of the solar day (sunrise to sunset) and seasonal patterns

[1] Economists have developed an interest in time zones/DST to understand the costs of coordination (Hamermesh et al., 2008), effects on trade and FDI (Marjit, 2007), on financial markets (Kamstra et al., 2000), and car accidents (Sood and Ghosh, 2007).

in daylight for each county. Using additional information on time zones and daylight saving[2], I demonstrate that local standard times of sunrise and sunset depend on solar time, time zone, daylight savings regime, and the position within the time zone. Building on these stylised facts, I show in a simple model that these geographic and institutional patterns generate two different sources of variation that can be used to estimate the effect of the timing of daylight on residential electricity consumption. In principle both variation within time zone and across its boundaries can be used for estimation.

The mainland US presents a very good case study for the effects of the timing of daylight on residential electricity consumption because it is large enough to span four time zones, yet all counties share common institutional factors. Moreover, the US Department of Energy publishes panel data on residential electricity sales for each year between 2001 and 2009 for the entire country. This data contains information on annual retail revenue, sales, and customer counts, by state and by class of service[3], for each electric distribution utility, or energy service provider in all 50 states. In total, over 3,400 providers generate and sell electricity to residential customers in the US, which can be mapped into the counties of operation. The resulting data is a county-level panel of average annual residential electricity consumption, which can be directly used for estimation of the effects of the timing in daylight.

The coefficient for averaged annual sunrise time is insignificant for the US overall. However, this results masks stark heterogeneity across latitude and climate. In the North, later sunrise is associated with increases in residential electricity consumption, whereas in the hot South the effect goes the opposite direction. These patterns are remarkably robust. In the most demanding specification I include controls for geographical latitude, time-varying county-level industry structure and employment, and county-level census data on climate, land area, population, educational attainment, median age and poverty, and state fixed effects, and my general conclusions remain unaffected. I include this rich set of controls to hopefully capture all the unobserved geographically correlated factors that might otherwise invalidate the approach. I also test the robustness of these findings against a number of potential threats, including measurement error and specification of the functional form.

[2]A synonym for 'British Summer Time'.
[3]including the Transportation sector, new in 2003.

This chapter offers a first explanation of the channels that could give rise to differences in the effects of the timing of daylight in the North and the South. As I show, people living in the North and the South get the same overall amount of daylight over the year in principle. However, the North is much colder and has a larger seasonal variation in sunrise times. A simple analysis suggests that in the hot South later sunrise could lead to lower residential electricity consumption if this shifts the hours of human activity into the colder morning hours. Such a change could result in a reduced demand for cooling, which is one of the major sources of residential electricity consumption in hot areas. However, in the North, temperature-related arguments cannot explain why early sunrise would reduce electricity consumption since most heating uses fossil fuels. I argue that the extent of people's waking hours at home (versus at work) can generate a situation where early daylight is associated with lower residential electricity consumption through changes in the demand for lighting in the dark mornings.

The finding of this heterogeneity in the effect of the timing of daylight on residential electricity consumption is completely novel and has potentially important welfare consequences. This is because the timing of daylight is determined by institutional factors which policymakers can directly influence. However, further work is required to gain a better understanding of the economic channels that give rise to these effects. While this chapter offers a first attempt to explain potential behavioural channels, it is left to future research to examine these in detail.

The rest of the chapter is structured as follows. The next section reviews the literature and explains where the approach taken here differs from the existing literature on electricity savings and daylight savings time. Section 3 presents stylised facts about the geo-temporal variation in local daylight times that results from geography, time zones and daylight savings regimes. A short historical discussion highlights the roles that (exogenous) geography and (endogenous) institutions play in the generation of this variation. Next, section 4 presents a simple model to show how this geo-temporal variation can be used to estimate the effect of the timing of daylight on residential electrify consumption. Section 5 presents the data. In section 6, I discuss the results obtained from two different sources of variation. Section 7 presents a series of robustness checks, before I offer a first explanation of the behavioural channels that could explain the new set of stylised facts in section 8. Finally, section 9, concludes and outlines directions for future research on this topic.

2 Literature Review

To my knowledge, there exists no direct evidence for a link between electricity savings and time zones. However, the existing literature on the effect of daylight savings time regimes on electricity consumption can be seen as an indirect test of the general effect of time zone assignment on electricity consumption. This is because standard time varies across time zones by exactly one hour, equivalent to the variation around DST time changes, when clocks are adjusted one hour forward or backward. As a result, observing DST can be interpreted as changing time zones for the summer period.

While DST was originally established to reduce energy demand as first advocated by Benjamin Franklin in 1784, there is a lack of empirical evidence as to whether it achieves this aim. Aries and Newsham (2008) conclude in their literature on DST and electricity savings that we are far from an understanding. They write: "There is general consensus that DST does contribute to an evening reduction in peak demand for electricity, though this may be offset by an increase in the morning." (p. 1858). This is in line with the most recent study by Kotchen and Grant (2012), who present the only microeconometric study in the field using US data to study DST and electricity consumption in the state of Indiana. They use the fact that some counties in Indiana changed their DST policies in 2006 and track changes in electricity consumption using household level data. They find that for Indiana DST in fact increased residential energy consumption, as there is a trade-off between electricity consumption in the evening, and energy consumption for heating in the morning. The only other recent econometric study looking at DST and energy consumption is by Kellogg and Wolff (2008), who use a natural experiment in Australia, where some regions altered their DST patters for the Sydney Olympics. Their main finding is that morning and evening reductions and increases in electricity consumption offset each other.

However, studies comparing DST regimes across contiguous localities ignore the effects of synchronisation. By this I mean that the existing literature on daylight and electricity consumption neglects (or assumes away) the fact that we derive benefits from coordinating activities across space. If a neighbouring locality, but not my own, changes its DST policy and I happen to work there, for example, I will need to adjust my work patterns in accordance with this locality's time policy, regardless of my own. As a result, work schedules or national TV schedules do not necessarily change in line with DST policies for each locality (i.e. holding solar time constant). Indeed, Hamermesh et al. (2008) demonstrates that there are large benefits to synchronising economic activity over space. They show that national TV scheduling has large effects on the timing of economic activity. Hamermesh et al. (2008) shows that if your locality just changed to Summer time, for example, whether your neighbour also changes time affects when you get up. As a result, it is unclear to what extent any local differences in DST, as in Kotchen and Grant (2012) for example, result in changes of behaviour that in turn effect energy consumption. This is a general problem of difference-in-difference DST studies, as localities with different DST regimes must be otherwise as similar as possible for credible estimation.

The alternative is to compare electricity consumption before and after the actual DST change, in Spring and Autumn. This is, of course, also not a viable approach, as any results would be the local effects found around the dates of clock changes. Since DST is introduced to generate summer savings, we would in fact expect the local effect around the DST-changing dates to be close to zero, if DST was set optimally. Therefore, in order to answer more general questions regarding year-round timing of sunlight and electricity consumption, using the DST time-discontinuity would not be useful.

The existing literature holds daylight constant and examines changes in local times, thus exposing itself to the issue of synchronization. As I show later, the variation that I am using in this study is not affected by these issues. I can control for synchronisation by holding local current times constant but varying daylight. A further problem of local studies is that they cannot uncover potentially heterogeneous effects across different climates. For these reasons, it is unclear whether the local effects of DST can be generalised. To my best knowledge, this study is the first to use nationwide data on electricity consumption to fully examine heterogeneity across different climates and latitudes.

The next section describes the geo-temporal variation in sunrise times that is used for estimation.

3 A short primer in astronomy and time zones: variation in local time of sunrise

3.1 Sunrise and sunset of the solar day

Sunrise and sunset of the solar day and seasonal variations are geographically determined for each location depending on the exact position of that location on the surface of the earth and the position of the earth with respect to the sun. As a result it is possible to calculate these two variables for any location directly using a mathematical approximation for the shape of the globe and the path around the sun[4]. Figure 1 shows the resulting spatial variation in annual daylight for summer (June) and winter (December). Seasonal differences in daylight depend on latitude offset over the year. The North gets shorter days during December, but longer days in summer. Over the whole year, differences in total minutes of daylight are negligible[5]. Independent of the season, patterns in solar day-length only differ on the vertical axis, as can be seen by the horizontal layers in Figure 1. This means that any two locations on the same latitude experience exactly the same seasonal patterns of solar day-length. Ignoring cloud cover, for any given day of the year, all locations on the same latitude band have the same number of minutes of sunshine. Overall, each location in the US gets about 734 minutes of daylight per day on average, annually. These facts are exogenously determined by geography.

3.2 A short history of time zones and daylight savings time (DST)

In order to derive local standard times of sunrise and sunset, i.e. the time shown on local clocks, it is necessary to combine solar information with the respective time zone (off-set from GMT) and daylight savings regime. Even if we regard time zones as daily reality, they are only a relatively recent phenomenon. Historically, local timekeeping only emerged with the development of mechanical clocks, and the word 'punctuality' only emerged in the English language in the late 18th century (Levine, 1998). Dur-

[4]The U.S. Department of Commerce, National Oceanic & Atmospheric Administration, provides a solar calculator that is highly accurate for locations within the US at URL: `http://www.srrb.noaa.gov/highlights/sunrise/NOAA_Solar_Calculations_day.xls,whichIuseforcounty-levelcalculation`

[5]Note that I ignore differences due to local weather or cloud cover, which are negligible over the long run according to Hamermesh et al. (2008).

ing the 19th century, villages would each have their clock tower and set noon to the highest point of the sun. As a result, over 70 different time zones are recorded for the 1860s in the US alone (ibid.). The four time zones in the mainland US as we know them today were only introduced in 1883 and formally established in 1918, and only marginally changed thereafter (see Levine (1998) for a fully-fledged historical discussion).

Daylight saving time is defined as temporarily advancing the time by one hour during summertime, which is referred to as British Summer Time in the UK. This procedure was first advocated in the US by Benjamin Franklin in 1784 and in the UK by William Willett in 1907 (Aries and Newsham, 2008). The idea was to shift human activity one hour backwards to save energy used for lighting. DST was first introduced during WW1 by Germany and subsequently adopted by other European countries. The US first introduced DST in 1918. Contrary to time zones, daylight-savings time has continuously been modified. The US, for instance, was on 'year round DST time (YRDST)' in 1974-1975. The current British Prime Minister, David Cameron, wants to put the UK on double-DST, effectively putting the United Kingdom into the GMT+1 time zone, for a trial period. Regarding the US, the last change in DST policies was in 2007, when it was lengthened, and Indiana started observing DST in 2006 (all from Aries and Newsham (2008) who discusses the historical background of DST in more detail).

3.3 Variations in local standard time or sunrise

Combining local information on solar time, time zone and daylight saving, the local standard time for sunrise can be calculated for each geographical location in the mainland US. In order to do this, I augment the mathematical model that calculates solar times with county-level information on time zone and daylight savings policy, by year[6]. Figure 2 shows the local standard time for sunrise in summer (June), winter (December) and annually (lower panel). Contrary to the previous exercise, time zones and daylight-savings regimes matter here in the sense that they influence the spatial pattern. The four time zones are clearly visible. Further, the local standard time for sunrise changes discontinuously at the borders of time zones. Within each zone, local

[6]This program is written in visual basic, building on the solar times calculator used in section 3.1.

standard time for sunrise increases smoothly as we move from the east to the west. This is because the sun rises on the eastern horizon in the morning, and hence rises earlier in the east, depending on the time zone. On average the sun rises at 6:49am, and one standard deviation in the average annual sunrise time is about eighteen minutes. Strong differences in seasonality are displayed in the upper panel of Figure 2. These arise because the sun rises from the northeast in summer and from the southeast in winter. Again, these differences cancel each other out over the year, so that time-bands for local standard sunrise time run vertically through the time zones in the lower panel. The total 'width' of each time zone corresponds roughly to one hour. That is, at the eastern border of each time zone, the sun rises about one hour earlier than close to the western border, for any two locations within the same time zone[7]. Finally, Arizona and large parts of Indiana did not observe daylight savings time, which is clearly visible in both the annual figure and also the top left-hand panel, showing sunrise times for June. In December, on the other hand, Arizona and Indiana do not stand out, as everyone is on standard time now[8].

In a nutshell, local standard time for sunrise exhibits a non-standard variation across space, depending on solar time, geographical location, and on position within time zone and daylight saving. While the former is geographically determined, the latter are policy variables, and daylight savings regimes have frequently changed over recent decades for reasons not related to robust empirical evidence. This is important to note, since it shows that policy can indeed affect the timing of daylight, which makes the question of timing of daylight and electricity consumption relevant from a policy perspective.

[7]This is this norm, given that there are 24 time zones for 24 hours on the globe. However, in other parts of the world time zones follow actual solar time less closely. Europe has a single time zone at GMT+1 that is spanning a region from eastern Poland to western Spain (about two and a half hours differences in sunrise-times), and China is on a single time zone.

[8]This map is drawn for the year 2003.

4 A simple model of annual residential electricity consumption

4.1 Intuition for the model

Figure 3 illustrates how the geo-temporal patterns described in the previous section can be used for the estimation of the effect of the timing of daylight on residential electricity consumption in a stylised way. The two boxes represent two time zones which have a one-hour difference in local time. For example, the right (eastern) box could represent the Central time zone, and the left box Mountain time. The smaller white boxes inside show the sunrise times for people living close to the western or eastern border within each time zone. Focussing on the left box, someone who lives close to the western border observes local sunrise at 7am. Another person living in the same time zone but close to the eastern border observes sunrise (and sunset) one hour earlier in local time: here, the sun rises at 6am. This pattern is the same in the other box (Central time zone). As a result, two sources of variation in the local time of sunrise emerge. First, moving within time zone, it is possible to compare the electricity consumption of people living close to the western versus the eastern border. Generally, moving horizontally within each time zone, daylight occurs later in local time. For now, assuming that everyone gets up at 7am local time, this would generate a variation of one hour in the timing of daylight.

The second source of variation comes from moving across time zone boundaries. A person who lives close to the eastern border of the Mountain time zone in Figure 3 observes sunrise at 6am local time, whereas a person living close to the western border of the Central time zone observes sunrise one hour later in local time, at 7am. In principle, both sources could be used for estimation.

There are some important factors to consider. First, within time zones actual solar time is changing but local time is constant. Everyone in the same time zone has the same local time. In contrast, at the boundary actual solar time does not change (the sun rises 'at the same time') but local times differ by one hour. As we will see, this has important consequences for the interpretation of the estimates. So far we assumed that everyone always gets up at 7am local time. Indeed, different sunrise times can only have real economic effects on electricity consumption if they are not mirrored by an exact behavioural response of getting up in the morning and going to bed in the evening. For example, if people living in a county towards the western border of a time zone get up about one hour later than people living in counties that are close to a time zone boundary to the east, we would not expect to find any impact on electricity consumption because their work-sleep patterns would not be different with respect to solar time.

Indeed, the assumption that has always been implicit - but never tested - in the existing literature is that people perfectly adjust their behaviour according to their local current time. This assumption implies that people always get up at the same local time regardless of the position of the sun, i.e. solar time. Similarly, issues such as coordination costs across space have been ignored. At the boundary, or when comparing counties that observe DST with neighbouring counties that do not, it is usually assumed that there are no coordination costs across space (Kotchen and Grant, 2012). However, these arise if people commute across a county or time zone boundary to get to work, or simply because people watch live events on television at the same time. Hamermesh et al. (2008) show that coordination costs across boundaries are non-trivial, and we should therefore not assume them away without knowing the consequences.

4.2 Setup

In order to understand how these different behavioural responses to changes in the timing of daylight could affect reduced form effects, I present a simple model of local times of waking and sunrise in the following[9].

Sunrise in current time $SRCT$

$$SRCT = f(longitude, TZ) \tag{1}$$

From Figure 2 we learn that sunrise in current time is a function of geographical longitude and time zone. Sunrise in current time defines the actual time that is shown

[9]I continue to refer to sunrise times and waking times in the morning rather than the evening. While sunrise in local time in the morning and sunset in the evening change symmetrically as we move across longitude, there is the possibility that people who, say, get up earlier do not go to bed earlier by the same time difference. For simplicity, we shall assume that the total hours people are awake does not vary depending on geographical location within a time zone. However, this is an important assumption that should be tested using the American Time Use Survey in future research.

on the clocks in each location. Accordingly, $SRCT$ is a function of longitude and time zone, denoted by the term TZ. We ignore differences in daylight savings regimes for simplicity.

Waking time in current time WCT

$$WCT = g(longitude, X) \tag{2}$$

As discussed, the effect of the current time of sunrise on electricity consumption also depends on how waking behaviour changes with sunlight. This is why waking in current time, WCT, is a function of longitude to capture the potential response to sunlight. The term X captures other influences over space that might affect the time people get up in the morning. A potential candidate for X is coordination costs across space, which will be discussed in more detail in section 6.2. Finally, WCT is not a function of time zone. This implies the assumption that people in different time zones in principle get up at the same time (i.e. at 7am). Taken together, residential electricity consumption then depends on how daylight changes, controlling for changes in waking times.

Specifically, differencing equations 1 and 2 we get:

Residential electricity consumption REC

$$REC = f(SRCT) - g(WCT) \tag{3}$$

Assuming linearity:

$$REC = [(\alpha_1 longitude + \alpha_2 TZ) - (\beta_1 longitude + \beta_2 X)] \tag{4}$$

We can now partially differentiate equation 4 at the time zone boundary and within time zones to shed some light on potential behavioural responses to changes in current sunrise times. Imposing linearity is a potentially strong assumption, which will be tested later on. Here, we keep the linear notation for simplicity.

4.2.1 Partial derivative: Variation within time zone

Using equation 4 and taking the derivative with respect to longitude within time zone, we get:

$$\frac{\partial REC}{\partial longitude} = [\alpha_1 - \beta_1] \tag{5}$$

where α_1 is how much SRCT changes when we move within TZ and β_1 how much WCT changes with changes in sunlight.

We can clearly see that if β_1 is positive the reduced form effect will be a combination of the effect of position within time zone on sunrise times and waking behaviour.

4.2.2 Partial derivative: Variation at the time zone boundary

Using equation 4 and taking the derivative with respect to longitude at the time zone boundary, we get

$$\frac{\partial REC}{\partial boundary} = [\alpha_2 - \beta_2] \tag{6}$$

where α_2 is equal to one since current time changes by one hour for each TZ in the US. α_1 is close to zero since latitudes of counties close to the time zone boundary are similar. What this highlights, however, is that some measure that captures longitude should be included as running variable in regression analysis that exploits the boundary discontinuity.

More importantly, β_2 depends on coordination costs/changes in conventions in WCT at the boundary. If people commute across the boundary to get to work, they effectively need to live on the neighbouring time schedule. There might be other reasons why people on both sides of the boundary would get up at different local times, and hence simultaneously. As shown by Hamermesh et al. (2008) air times of popular television programs have a significant effect on the time people get up in the morning and go to bed in the evenings. This is important because all major television channels air their programs simultaneously in the Eastern and Central time zone, for example. This pattern is less clear at the other time zone boundaries and depends on actual channels, but naturally all live events are aired simultaneously throughout the US. Therefore, the assumption that β_2 equals zero is a strong one and we have reasons to believe that β_2 is positive. If this is the case, again the reduced form estimate would be a combination of the time change at the boundary and the behavioural response and would go towards zero. In the extreme, if coordination costs at the boundary were prohibitive, and people got up simultaneously, there would be no effect on electricity consumption.

4.3 Summary

In this section we have seen that, in principle, two different sources of variation emerge from the spatial patterns described in section 3. Reduced form estimates of electricity consumption on sunrise times can be estimated using either within-time-zone variation across longitude, or the boundary discontinuity. However, effects of daylight timing on electricity consumption also depend on how people adjust their sleep patterns with respect to changes in the timing of daylight. I have shown that the effects of the two sources of variation will thus differ depending on the behavioural responses of waking time. Since the behavioural response is due to different reasons at the boundary versus within the boundary, it is not clear a priori why reduced form estimates should be comparable in magnitude and significance.

For example, if coordination costs across space are high, the behavioural adjustment in waking times within time zones should be minimal. This is because when coordination matters, people who live close to the eastern border of a time zone will need to get up simultaneously with other people living further west in the same time zone. As a result, local clock time will determine when people get up, and not the position of the sun. In contrast, behavioural adjustment with respect to local time at the boundary would be large. This is because if two people living on opposite boundaries of a time zone need to get up simultaneously, they will in fact get up with a one-hour difference in their local times. As shown by the model, the estimated effects would vary accordingly. I will return to these important considerations when comparing the estimates obtained from the two distinct sources of variation later in section 6.[10]

[10]The sum of β_1 and β_2 define the total time budget available, i.e. the total difference in waking times that we can find across the whole of mainland US. If, for example, people in the Pacific time zone get up exactly three hours later on average than people in the Eastern time zone, β_1 and β_2 needed to sum to one. Whether this equality needs to hold is something that future research should examine using geo-coded data from time use surveys.

5 Data

5.1 Residential electricity consumption

The U.S. Energy Information Administration (EIA) requires all energy utilities in the USA to report their annual residential electricity sales. Specifically, the form called 'EIA-861' contains information on annual retail revenue, sales, and customer counts, by state and class of service (including the Transportation sector, new in 2003), for each electric distribution utility, or energy service provider in all 50 states. Each utility or service provider also lists all counties of operation. Therefore, combining this information, it is possible to extract annual per-consumer (which is per electricity-meter) residential energy sales. [11]

In total, 3,420 different energy utilities sold electricity to residential customers in the US between 2001 and 2009. Over ninety per cent of energy utilities both produce and sell electricity. However, about six per cent of utilities do only produce and not sell electricity to residential customers themselves. This electricity is sold through the other providers or a small number of sales-only providers, on which information is available, as well. Since we are interested in the location of residential electricity consumption (and not production), we need to drop the six per cent of utilities that do not directly sell to end consumers themselves [12].

The EIA also collects seasonal information on residential electricity consumption[13], however this information is only collected for a subsample of energy providers. Therefore seasonal electricity consumption data is available for a sample of counties as well.

Table 1 shows descriptive statistics for annual county level per-customer electricity consumption in MWh. The first two columns show descriptive statistics for all counties in mainland US, and the remaining columns split the US into time zones. The first row gives averages for all latitudes, whereas the remaining rows split the US into quintiles based on latitude of county centroid. The first quintile includes the twenty per cent of counties furthest North, for example. All data is averaged over the period from 2001 to 2009 and weighted by county population according to the 2001 census.

[11] All power-utilities are required to provide this information through an online portal knows as 'Single sign-on'.

[12] Many thanks to Paul Hesse from the EIA for helpful explanations.

[13] This is done through the form EIA-862.

Turning to the statistics, the first entry in column (1) is average annual per-customer residential electricity consumption, which is 12.1 MWh. The remaining rows in column (1) show how this consumption varies over five latitude quintiles. We can clearly see that electricity consumption is higher in the South than in the North. In the most southerly quintile, per-customer electricity consumption averages 13.82MWh, compared to 10.63MWh in the most northern quintile. However, as the standard deviations in column (2) show, there is substantial variation within these geographical bands. As a result, these differences do not turn out to be statistically significant at conventional levels.

Looking at time zones individually, it is interesting to see that the North-South pattern documented in column (1) is not present in all four time zones. In fact, in the Pacific and the Mountain time zones, overall consumption is higher in the North than in the South. The overall pattern is thus driven by the Central and Eastern time zones. In the Eastern zone in particular, there is a clear pattern of higher electricity consumption in the North than in the South. The most northerly counties in the Eastern time zone have an average electricity consumption of about 7.85MWh per customer, compared to 14.12MWh in the South. These differences turn out to be significant, as we can see from the standard deviations reported in column (10).

To conclude this section, overall there is a North-South pattern in residential electricity sales. Splitting the US into latitude quintiles, we can see that electricity use is somewhat higher in the South. As I will argue in section 8.1 this is probably because the use of air conditioning for cooling is very electricity-intensive in the hot South. Patterns are somewhat different in the Mountain and Pacific time zones. These differences could be partly driven by climatic patterns. The next section describes data on climatic variables in detail.

5.2 Climate: Cooling Degree Days and Heating Degree Days

Different climates might affect energy consumption, and the effects of daylight might be different depending on climate. Table 2 shows indexes for Heating Degree Days (HDD) and Cooling Degree Days (CDD) for the four time zones of the mainland US and by latitude percentile. HDD and CDD are common measures in the energy sector. As we can see from column (1) in table 2, the South has significantly more CDD than the North, which has significantly more heating degree days. This is unsurprising, of course. Indeed, the correlation between geographical latitude and these measures is very high. The correlation coefficient for latitude and HDD is 0.9357 and for latitude and CDD -0.8737.

Turning to columns (3) to (10), CDD and HDD measures are shown for each time zone individually. The general North-South pattern of increases in CDD and decreases in HDD as we more further South is present in all four time zones. Interestingly, the South in the Central and Eastern time zones have a much higher index for CDD compared to the other time zones. For the most southern quintile, for example, the CDD index is 7.21 and 7.18 for the Central and Eastern time zones respectively, and significantly lower at 5.88 and 5.40 in the Pacific and Mountain time zones. The way the CDD index is constructed, this difference translates into a factor of about two, meaning that the Central and Eastern time zones have a much higher potential absolute demand for cooling. Since cooling is very demanding in electricity, these patterns might explain the overall higher electricity consumption in the south in these time zones, which we detected in section 5.1.

5.3 Further control variables

Appendix table A.1 shows descriptive statistics of additional county-level variables. The table shows data in the first five panels on population (measured in 2001), land area (square miles), median age, educational attainment (high school graduate or higher in 1990 and the number of persons below poverty level. This data is taken from the ICPSR 2896 Historical, Demographic, Economic and Social Data DS81:2000 County Data Book, and I will include these variables as additional control variables in some of the specifications that I discuss in the next section. The lower part of the table further shows statistics on county-level industry specialisation and overall employment and the information on these two variables is extracted from the County Business Pattern dataset for every year between 2001 and 2009. Since this is county-level data, again all entries are weighted by county population as recorded by the 2001 census.

Overall, the table shows some regional variation across both time zones and latitude, but these patterns do not seem significant. For example, there is a clear North-

South pattern for both educational and poverty levels. Column (1) shows that in the North people are on average better educated and less poor. However, there is also substantial variation within the latitude bands and these differences are not significant. Similarly, differences across time zones are not remarkable.

6 Regression Analysis, main results

This section presents regression results from using two different sources of geographical variation in the timing of daylight. Section 6.1 presents results from using geographical variation across longitude and within latitude bands for estimation. Section 6.2, on the other hand, uses variation in the timing of daylight that arises because of different current times on opposing sides of inland time zone boundaries in the mainland US. Before turning to the specific analysis, let us first discuss two technical notes that apply to all regression specifications presented below.

First, in all specifications I cluster the error term at the county level to account for the fact that each county is observed in nine consecutive years and that the residual is likely to be correlated within a county over time. Alternatively, I can use robust standard errors to control for heteroscedasticity only, which results in similar estimates. Since my treatment varies across geographical latitude I also clustered at the state*year-level result, which results in even smaller standard errors. I also estimated most specifications using two-way clustering to simultaneously control for potential autocorrelation in the residual over time and across geographical latitude[14]. In particular, I clustered the error at the county level and additionally along twenty-four latitude bands, which I constructed based on the integer values of the geographical county-centroid latitude coordinates. This two-way clustering also only marginally changed the estimated standard errors and never changed the interpretation of my coefficients. I therefore concluded that autocorrelation in the error term across latitude is not a major concern and cluster all standard errors only at the county level in all of the analysis below.

Secondly, I will use county-level averages in residential electricity consumption as dependent variable throughout. In principle, however, we want to make claims

[14]Two-way clustering was implemented in STATA using the cluster-command of the Ôivreg2’-routine, which allows for multiple level clustering of the error term.

about population electricity consumption. In order to do this we would ideally use individual-level data. However, as explained in section 5 my data is only available on the aggregated county level. Since counties differ in population, treating them all as equal would not make it possible to make statements about overall electricity consumption. Stated intuitively, this is because a change in the average electricity consumption in a county with a very large population would result in a larger change in national electricity consumption than a similar change in average electricity consumption in a county with smaller population.[15]

We can solve this problem by using weighted least squares and assigning analytic weights to the county-level regression. This can be done with the command 'aweight' in STATA, which I use to scale the assumed variance of the county-level data by the inverse of the county population. The data on county population is taken from the 2001 Census as described in section 5. Notice that using WLS is justified solely because I have grouped data. This is different to issues of heteroscedasticity or frequency weighting because of non-random sampling. For notational simplicity I will ignore the weighting matrix in the specifications spelled out below, but all results presented are based on WLS using analytic weighting as described here.

6.1 Analysis using within time zone spatial variation in the timing of daylight

6.1.1 Specification

The most basic specification that I estimate is the following:

$$\begin{aligned} lnY_{c,t} =& \gamma_0 + \gamma_1(avsunrise)_{c,t} + \gamma_2(timezone)_c \\ &+ \gamma_3(year)_t + \gamma_4(timezone)(year)_{c,t} + \varepsilon_{c,t} \end{aligned} \tag{7}$$

In this specification, the term $Y_{c,t}$ represents annual per-household electricity consumption for county c in year t. Sunrise times are in local current times. Further, time zone and year dummies are included to capture any potentially unobserved time zone year specific shocks. The coefficient γ_1 is the main coefficient of interest.

[15]This is similar to Angrist (1998) who estimates the labour market impact of military service using averaged data on earnings, see Angrist and Pischke (2008), p. 40 for a discussion.

If the within-time-zone variation in local sunrise times were truly exogenous to other factors that determine electricity consumption, then this simple specification should already reveal an unbiased estimate of the reduced form relationship of timing of daylight on electricity consumption as discussed in the model in section 4. However, there is the potential that historical time zone assignment has not been truly random, or that firms or people sorted themselves into specific geographical locations in ways that would confound causal interpretation. In order to alleviate these concerns I also estimate specifications including additional controls. Specifically, I include variables on the geographical latitude of the county centroid, Cooling and Heating Degree Days (CDD, HDD), industry specialisation and employment numbers, land area, population, a poverty measure and education[16]. These controls are included to capture factors that potentially correlate with local time of sunrise and residential electricity consumption over space. If unobserved, these factors could induce omitted variables bias. Since it is not clear a priori which variables are likely candidates to capture such geographical patterns, I follow a 'kitchen sink' approach and include this wide range of controls. The hope is that conditional on general control variables on education, production and climate, there are no relevant unobserved factors correlated with local times of sunrise and electricity consumption.[17]

To further alleviate potential concerns of omitted variable bias I also estimate regressions that include state-by-year fixed effects. This is to control for any institutional differences of states that could affect electricity consumption and also correlate to within time zone geography. Estimating the effect of the timing of daylight on electricity consumption within states is very demanding because state fixed effects alone explain about 44 per cent of the variation in annual electricity consumption conditional on year and time zone.[18].

6.1.2 Estimation results

Table 3 shows the estimates for the main coefficients of interest for specifications that try to explain residential electricity consumption as a function of local times of sunrise

[16]For descriptive statistics on controls see Appendix Table A.1

[17]Since it is hard to see these general control variables as outcomes themselves, the hope is that including these variables does not cause 'bad control' issues, i.e. bias in the main coefficient of interest, as explained by Angrist and Pischke (2008).

[18]Obtained by keeping the residual of the specification in column (2) with explanatory variable *avsunrise* excluded. Over 43 per cent of the remaining variation in the residual is between states.

using the spatial variation of local sunrise times within time zones. The first column presents the γ_1 estimate of specification 7 above. In the second column controls are added to the specification, and the third column further includes state and state-times-year fixed effects. Columns (4) to (6) and (7) to (9) repeat these regressions on a subset of counties, splitting the US into two equal halves based on geographic latitude of county centroids.

What we can see from the estimates in the first column of table 3 is that there is a positive association between sunrise times and residential electricity consumption across the whole US. The estimated effect is significant at the one-per cent level, and very large: A one-hour-later sunrise is associated with about twenty per cent higher annual residential electricity consumption in column (1). However, this estimate is almost halved once we include a rich set of control variables in column (2). The fact that the estimate is sensitive to the inclusion of controls shows that the regression in column (1) suffered from omitted variable bias. Jointly, the additional control variables included in column (2) correlate with within-time-zone geography and electricity consumption. Indeed, the adjusted R2 rises to 0.45 in this specification, a dramatic increase compared to the previous specification with an adjusted R2 of 0.16. Further, adding state fixed effects in column (3) completely removes any association between average annual sunrise times and residential electricity consumption across the US. Controlling for state-by-year averages in residential electricity consumption makes it possible to predict almost 80 per cent of the variation in residential electricity consumption. At the same time the estimated standard error in column (3) remains unaffected. This suggests that the insignificance of the estimate for the effect of the timing of daylight on residential electricity consumption is not driven by lack of within-state variation in the outcome or explanatory variables. Taken at face value, this estimate means that shifting existing time zone boundaries towards the East or West would not result in any overall residential electricity savings.

Moving to columns (3) to (9), where the US is split into North and South, it becomes evident that there is substantial heterogeneity across geographical latitude. We should note that the US is split into two halves crudely based on geographical latitude of county centroid and ignoring any other boundaries or location-specific features.

The estimate in column (4) shows that in the North an one-hour-later annual average sunrise is associated to about a thirty-five per cent increase in electricity consumption in the unconditional specification. This is an unrealistically large effect and once the rich set of control variables is included the estimated coefficient reduces from 0.341 to 0.250. Again, this suggests that the control variables are not randomly distributed over the within-time-zone geography. Adding state fixed effects in column (6) further reduces the effect to a sixteen percentage-point change in residential electricity consumption. Overall, this set of results demonstrates that while the estimates are sensitive to the inclusion of controls, the estimates remain large in size and highly significant even in the most demanding specification.

Columns (7) to (9) show the estimates for the southern half of the United States. Here, the results are opposite to the North. In the South, a one-hour-later average annual sunrise is associated with a reduction in residential energy sales of about sixteen per cent in the unconditional specification. Notice that this effect in the South is very robust to the inclusion of a wide range of control variables. Indeed the point estimate remains virtually identical in column (8). Here, even including state fixed effects does not significantly alter the estimated coefficient. The unconditional estimate in column (7) is estimated at -161, including state fixed effects reduces the coefficient only to -0.131. This difference in estimates between the unconditional and most demanding specification is not significant at the five-per cent level.

To recap, using the spatial variation within time zones to estimate the relationship between residential electricity consumption and the timing of daylight we found two results: first, there is no robust evidence for an overall association between average annual sunrise times and electricity consumption. However, this overall result masks significant heterogeneous effects across latitude. Splitting the US into two halves along county latitude, in the North a delay in sunrise is associated with an increase in residential electricity sales, whereas later sunrise with lower electricity consumption in the South. While the estimated effect is sensitive to the inclusion of controls for the North, the estimates for the South are remarkably robust to the inclusion of a wide range of controls variables. We can even include additional state fixed effects, which take out over 40 per cent of the variation used in the estimation, and the estimates remain unchanged compared to the unconditional specification.

6.2 Analysis using time zone boundary spatial variation in the timing of daylight

An alternative approach to estimating the effect of daylight on residential electricity consumption is to focus on time zone boundaries.

In order to implement regression analysis of boundary counties it is necessary to first identify all counties that are close to the boundary. Initially, I focussed on counties that are contingent to a time zone boundary only. However, it turned out that using counties that share a border with the time zone boundary resulted in large estimates of the standard errors due to the small sample size. Also, since counties in the eastern US tend to be smaller than counties in the West, the overall area included was not balanced across space. Therefore I now focus on 612 counties that lie within a 100km buffer around a time zone boundary. Figure 4 shows these counties divided into 'treatment' and 'control' groups. First, note that a few counties, mainly around Arizona, are not grouped into control or treatment group, because they followed different daylight-savings regimes for at least one year of the study period. While this is taken care of in the construction of the average sunrise variable, it is less clear what would happen to the discontinuity. In particular, it is not clear which would be a control county and which ones would be treated. To be on the safe side, I exclude these counties from the control and treatment groups for the boundary analysis. The remaining counties are grouped into a treatment and control group, where the treatment group consists of counties that lie east of the respective time zone boundary. These counties have a local time one hour later than the control group, hence the estimated coefficient can be interpreted as adding one hour, or as sunlight beginning one hour later under the following two conditions.

Firstly, when using this variation it is only possible to estimate the coefficients of interest on a subset of counties, namely counties that are close to a time zone boundary. One concern is that these counties might not be representative and it might not be possible to examine heterogeneity because of small numbers. Appendix Tables A.2, A.3, and A.4 replicate Table 1, 2 and Table A.1 previously described in section 3, but for the sample of boundary counties only. Notice that since counties around the state of Arizona could not be included, there are some missing entries for the southern quintiles in the Pacific time zone in these tables.

First turning to Appendix table A.2, which shows the average residential electricity consumption the counties that lie within 100 km of an inland time zone boundary, what we see is promising: the boundary counties are quite similar to the rest of the US. Again, there is the overall North-South pattern with higher electricity consumption in the South, as shown by column (1): the average customer in the most northern boundary country uses about 9.57MWh, whereas this figure is 13.37MWh for the most southern quintile. Notice that I again split the US into five latitude quintiles based on the latitude of the county centroid. The climate variables on Cooling and Heating Degree Days also follow a similar pattern, and they are tabulated in Appendix Table A.3.

However, Appendix Table A.4 shows the descriptive statistics for the boundary counties. Focussing on column (1), comparing numbers across to table A.1 it becomes clear that the boundary counties are indeed not representative. The first column, for example, shows that the average boundary county has a population of about 100,000, which compares to 150,000 in the full sample. This is a potential caveat when trying to generalise results from the boundary estimation.

A second concern, which I already pointed out in section 4, is that at a time zone boundary daylight does not change much. Indeed, very close to the boundary the real change in solar time is negligible. Instead, local current time changes by one hour. We know from Hamermesh et al. (2008), who study time use in adjacent counties in Arizona that followed different daylight savings regimes, that there is extremely little impact on behaviour in terms of waking time when clocks are changed but neighbours remain in a different time zone. The key problem is that it is not daylight that varies across the boundary, but local current time. While we can probably assume that people get up at the same local time within a time zone, it is harder to assume that they get up with a one-hour time difference at the boundary. This would imply that β_2 in Equation 6 is likely to be greater than zero. In fact, in order to compare estimates to the previous exercise, one would need to assume an elasticity of waking up with respect to local time of one at the boundary. If this is not met, the reduced form estimate will be lower depending on β_1 and β_2 of equation 6, as shown in section 4.

6.2.1 Specification

The simplest specification that I estimate is the following:

$$\begin{aligned} Y_{c,t} = \delta_0 + \delta_1(treatment)_c + \delta_2(tzboundary)_{c,t} \\ +\delta_3(year)_t + \delta_4(longitude)_c + v_{c,t} \end{aligned} \tag{8}$$

Here, δ_1 is the main coefficient of interest and should capture the effect of being in the treatment group, i.e. a one hour later local time, on electricity consumption. δ_2 is an estimate for the difference between the group of boundary counties shown in figure 4 overall, compared to all other counties which are still included in the regression to reduce the Residual Sum of Squares. A significant coefficient here would indicate that boundary counties are on average significantly different to the average other county in the US. Note that it is now not possible to include time zone fixed effects but time fixed effects are still included to capture any overall differences in annual electricity consumption. As highlighted by the theoretical discussion in section 4 a measure of longitude is included as running variable, here the longitudinal coordinate of the county centroids. I continue to cluster the residual at the county level and weight each county by its overall population using weighted regressions.

6.2.2 Results

Column (1) of table 4 shows the estimates for specification 8. Here, all latitudes and time zones are bunched together. First, note that the estimate for β_2, reported in the second row, is negative and significant. This raises important concerns from an external validity perspective as this shows that boundary counties have lower residential electricity sales compared to the rest of the mainland US. The main coefficient of interest reported in the first row is also significant (and positive), but these are only the unconditional results.

However, it turns out that the inclusion of the usual set of control variables in column (2) does not change much, and even state fixed effects (column (3)) do not change the message: using the boundary variation there seems to be an overall positive association between sunrise times and electricity consumption. The finding of a positive effect in the most robust specification in column (3) especially seems to contradict the earlier finding that there is no significant overall effect. However, recall that we had to drop a significant number of counties in the South due to the changing DST regimes around Arizona from the control and treatment groups. In the within-time-zone analysis, splitting the US into two halves by county centroid ensured an equal number of counties in the North and South in the previous analysis. Here, more counties in the North are treated than in the South. Therefore, these results are less informative. This becomes clearer when looking at the effects for northern and southern counties separately.

Columns (4) to (6) show estimates for the same regressions but using counties in the northern half only, and (7) to (9) are for the South. Again, the estimates in the second row all turn out significant and negative. Counties close to the boundary are non-representative as they have lower per-customer annual electricity consumption.

Turning to the estimates for the treatment, later sunrise is significantly associated with higher electricity consumption. The estimated effect is always significant at the one-per cent level. The unconditional estimate reported in column (4) is 0.094, which is only reduced to 0.089 by the inclusion of the usual set of control variables. Further including state fixed effects reduces the coefficient to 0.060. To summarise, the estimated effect in the North is robust to the inclusion of controls, though it does decrease by about three percentage points. However, this reduction is not significant at the five-per cent level.

Columns (7) to (9) show the respective estimates for southern counties. In the South, there seems to be less of a problem in terms of representativeness, which documents itself in the fact that the 'tzboundary'-estimate is insignificant in two of the three specifications, and also smaller in absolute terms. In contrast, the treatment coefficients are consistently estimated at negative values. The unconditional estimate in column (7) is -0.024, which is significant at the five-per cent level. Including control variables in column (8) reduces the coefficient to -0.021, which makes it just non-significant. However, including state fixed effects increases the estimated effect to -0.030, which is precisely estimated due to the reduction in the residual and therefore significant at the one per cent level. Again, we conclude that there is a negative association between average sunrise time and electricity consumption that is robust to the inclusion of a rich set of control variables, and even state fixed effects.

Summarising the boundary estimates, overall the results point in a similar direction to the findings using the within-time-zone variation. The estimates again suggest that the North could benefit from earlier sunrise, while the South would benefit from later sunrise. As before, the estimated effects are remarkably robust against the inclusion of controls, especially in the South. This is exactly what we previously found using the totally different variation in average sunrise times within time zones.

However, the results of the boundary counties should be taken with a pinch of salt. First, as the significant estimates for the dummy variable indicating boundary status indicates, these boundary counties are significantly different to the rest of the US in terms of electricity consumption. Therefore it is not clear if these results can be generalised across the US.

In addition, in order to interpret the estimate as the effect of sunlight on electricity consumption at the boundary, we have to make the unrealistic assumption that the elasticity of getting up with respect to local time equals one. As argued before, this is not very likely to be the case due to coordination costs across the boundary. For both of these reasons, the magnitudes of these results are not directly comparable to those obtained from the within-time-zone variation.

Keeping these caveats in mind, we do find the same overall pattern using both completely orthogonal sets of variation: the North benefits from early light, whereas the South suffers.

7 Robustness checks

7.1 Functional form

All findings so far come from specifications assuming linearity. As already mentioned in section 4 this is a potentially strong assumption, which is relaxed in table 5. Here, I estimate nine different regressions and the first three columns present results from regressions for the mainland US, columns (4) to (6) for the North and (7) to (9) for the South. Note that these results estimate the effect using within-time-zone variation in the timing of daylight, as in section 6 only. This is because the 'treatment' close to the boundary is not continuous, which makes it impossible to consider alternative functional forms.[19]

[19]Technically, this is not possible simply because the boundary treatment is captured by a dummy variable, and a dummy is equal to its own square.

In contrast to the results presented in table 3, the explanatory variable for average local sunrise time is now also included as a quadratic and first two rows show the respective estimates. Since it is difficult to compare these results to the previous linear specifications directly, table 5 also reports computed marginal effects at variable means in the third row.

The first thing to note is that most estimates of both the linear and quadratic term are significant at the one-percent level. In principle, this would suggest that the quadratic term should be included. It is only in columns (7) and (8) that the estimates are non-significant. However, when turning to the marginal effects, the results are very close to the linear specification results presented in table 3, both in terms of magnitudes and significance. Indeed, the estimated marginal effects are almost identical and never different from the linear model in any of the specifications at any conventional significance level. Comparing the most robust specifications that include state fixed effects, the estimated marginal effects are -0.008, 0.174** and -0.158**, which compares with -0.002, 0.163** and -0.131** in table 3. Therefore I conclude that the linearity assumption is defensible on grounds of simplicity.

7.2 A closer look at heterogeneity by latitude

The findings so far suggest that there is heterogeneity across latitude in the effect of average local sunrise times on residential electricity sales. This section examines this finding in more detail, splitting the US not only into North and South but into five latitude bands based on quintile of county centroid. This is again only possible when looking at variation within time zones. Around the boundary, there are not enough observations for each quintile to obtain precise estimates.

Columns (1) to (3) of table 6 mimic the regressions of the first three columns of table 3, but coefficients are estimated separately for each latitude *quintile*. All effects are estimated from running separate regressions for each quintile, thus table 3 reports results obtained from fifteen different regressions. As before, in columns (2) and (3) we subsequently add control variables and state fixed effects.

Turning to the results, column (1) shows estimates for specification (7) broken down by latitude quintile. Moving from the North to the South, there is a strong and similarly-sized positive estimated effect in the northern two quintiles, which then turns negative in the third quintile to -0.011, but not significantly different to zero at conventional levels. Moving further south, the fourth and fifth quintile both have large negative associations between average sunrise and residential electricity sales. In fact, the coefficient for the most southern quintile is somewhat smaller than for the fourth quintile. However, these are only the unconditional results.

In column (2) the usual set of control variables is included. Here, there are some marginal changes in the coefficients resulting in a smooth pattern as we move from the North to the South. The inclusion of additional state-times-year fixed effects again does not change much. As we can see in column (3), here the magnitudes of the effects are reduced in the North, but amplified in the South, resulting in a very similar overall pattern.

To summarise the findings so far, breaking up the US into latitude quintiles confirms the previous finding that the effect of average sunrise times on residential electricity sales is heterogeneous by latitude. The results in table 3 for the North and South of the US are not driven by some outliers or few counties, but there is evidence for an overall North-South pattern across latitude quintiles. For the middle quintile of the US there is no evidence for a significant association between sunrise times and electricity consumption in any of the specifications. Therefore, we can conclude that this pattern is robust and the later sunrise is indeed associated with higher electricity consumption in the North, and lower consumption in the South.

7.3 Measurement error

There is no measurement error in the timing of daylight variable but as explained in section 5, some power utilities serve more than one county, and whenever this has been the case, per-customer sales have been averaged over the entire service area. Theoretically, it is unclear why measurement error in the dependent variable should bias my results. Nevertheless, table 7 reports the main results relying only on county level electricity consumption data that was derived using utilities that serve at most 10 counties $Panel A$, or exactly one county $Panel B$. While these restrictions result in a loss of up to 65 per cent of the counties for which electricity data is available, none of these changes significantly affects the main results.

8 Interpretation

Due to the lack of empirical evidence it has not been clear a priori what to expect in terms of findings. Equally there are no clear theoretical predictions of the effect of the timing of daylight on electricity consumption. This is because it is difficult to generate clear theoretical predictions without any empirical guidance, and the relationship is further complicated by the fact that people do not maximize their daily schedules with respect to sunlight and electricity consumption only. Indeed, the timing of daylight certainly matters for individual utility and welfare in many other dimensions.

This study presents the first nationwide empirical assessment of the timing of daylight and residential electricity consumption. Guided by these new empirical results I present a first attempt to rationalism the findings in the following. In particular, I am proposing two different mechanisms to explain the documented associations between daylight and electricity consumption in the North and the South. In any case, I acknowledge that more research is needed to examine these, and potentially other, channels in more detail.

8.1 The demand for cooling in the hot South

According to the US Annual Energy Review (USdOT, 2010), American households used electricity equivalent to 0.88 quadrillion Btu[20] for cooling in 2005, which constituted 20 per cent of overall household electricity consumption. Unfortunately, regional data is not available, but given that the South has a much higher demand for cooling, it follows that electricity use for cooling is responsible for a high share of overall electricity consumption in the hot South.

One possible explanations for the finding that later sunrise could reduce electricity consumption in the South is illustrated by figure 5: the functions show a typical relationship between air temperature and daytime. In particular, the coldest point of the day just after sunrise, whereas the hottest time of the day is in the afternoon.[21] Further assuming that demand cooling is higher when people are awake during daytime, shifting daylight later can result in electricity savings. For instance, if people get up at 7am and go to bed at 11pm and only demand cooling during this time, the area

[20]One kilowatt-hour=3.412 Btu

[21]Source: http://www.wisegeek.com/what-is-the-coldest-time-of-the-day.htm

between 7am and 11pm that lies under curve (A) represents potential total demand for cooling. Shifting daylight later, the temperature schedule also shifts as shown by function (B). Since it is colder in the morning than in the evening, the area between 7am and 11pm under function (B) is strictly smaller compared to function (A). Put simply, a relatively later sunrise shifts the hours of human activity into the cooler times of the day, which can potentially result in savings for cooling demand. These effects are exacerbated if people are at work from 9am to 6pm and only demand cooling when at home as the savings from the early morning hours would be a larger proportion of overall consumption of cooling.

This diagram can be used to generate a number of predictions that should be brought back to the data in future research: in particular, a critical assumption is that demand for cooling is lower when people are asleep. This could imply that the association between timing of daylight and electricity consumption should not hold in the South in months when it is so hot that people leave the air conditioning running 24 hours a day. Future research should address this prediction using seasonal data on electricity consumption in hot areas.

8.2 The demand for lighting in dark mornings in the North

The demand for cooling cannot possibly explain the findings for the North since the overall demand for cooling is low in cold places. Instead, heating mainly relies on fossil fuels rather than electricity, which only makes up about 6.5 per cent of total energy use for residential space heating. Again, no regional information is available, but it seems plausible to assume that colder areas are less likely to use electricity for space heating since fossil fuel is more efficient. As a result it seems unlikely that a temperature-related story gives rise to the patterns documented for the North of the mainland US.

Figure 2 shows that the sun rises at the same local time in the North and South in the annual average. However, what the lower panel of figure 2 does not show is that there is much larger variation in sunrise and sunset times in the North. In the summer, the sun raises extremely early and days are very long, as we can see in the upper panel. During winter days are very short and the sun raises after people would normally get up.

Figure 6 shows that in situations when people get up (and switch on the light) before the sun raises, early sunrise can result in savings. Making a few additional assumptions, this is because the equivalent 'loss' of daylight in the evenings occurs at a point in time when people are still at work. To see this, notice that the top part of figure 6 shows total hours of daylight over the time of the day. Two situations are compared, when the sun rises after people get up in scenario (A) and when the sun rises exactly when people get up, scenario (B). Assuming that people do not consume electricity for lighting when at work, the lower part of the figure backs out the hours when people would need to switch on the light under both regimes. If the sun rises late (A), people consume lighting before going to work and after coming back. In contrast, if the sun is already up when people get up in the morning, they only consume lighting after work (B). Again, future research should examine these channels and test whether, for example, effects only emerge where and when people get up before sunrise.

9 Discussion of results and concluding remarks

In this chapter I have shown that the variation in local standard times of sunrise are non-standard across space, and depend on geographical position, time of the year, time zone, daylight savings regime, and position within the time zone. Building on these stylised facts I have demonstrated that two different sources of geographical variation in the timing of daylight can be used in order to estimate the effects on residential electricity consumption. First, variation in the timing of daylight that arises along latitude bands within time zones can be used. Alternatively, we can use differences in local current times that arise in counties in proximity to either side of inland time zone boundaries.

Using the within-time-zone variation in the timing of daylight along geographical latitude bands for estimation, I find no evidence for an overall effect of average sunrise times on residential electricity sales in the most robust specification. However, this finding masks substantial heterogeneity along geographical latitude. In particular, I show that a one-hour-later sunrise in the annual average is associated to an about 16-per cent increase in residential electricity sales in the North. Contrary, in the South a one-hour-later average annual sunrise is associated with a reduction in residential electricity sales of about 13 per cent in the most demanding specification. Especially for the South these estimates are insensitive to the inclusion of a rich set of county level control variables, including industry specialisation, industry employment, population, area, educational levels, median age, climate variables, latitude, a poverty index, and state-times-year fixed effects. Further, the heterogeneity across latitude is shown to be a general pattern that is present throughout latitude quintiles.

Next, the variations in local times across time zone boundaries are used for estimation. Using this totally different source of variation, I can confirm the general pattern of the previous findings. In the most robust specification a one-hour-later average annual sunrise is associated to a six-per cent higher electricity consumption, whereas the effect in the South is estimated to be negative at three per cent in the most demanding specification.

I have also shown in a theoretical discussion of these two sources of variation that the reduced form estimates from within-time-zone variation and at the boundary capture different behavioural responses towards the solar position of the sun and local time. In particular, coordination costs at the boundary could explain why the estimates coming from the boundary variation are lower than those from the analysis that uses the within-time-zone variation in daylight for estimation. Therefore I am not overly concerned by differences in the point estimates, but future research should examine the proposed behavioural responses to explain the differences in findings. This could be done using geographically localised time use data, for example.

In this paper, I also present a first attempt to highlight potential channels that could give rise to this new set of stylised facts, namely that early daylight is associated with increased electricity consumption in the South and lower consumption in the North. I argue that later sunrise in the hot South could shift human activity into the cooler hours of the day, which would then result in electricity savings. In the North, additional assumptions about work times are necessary to generate a situation where earlier sunrise can reduce electricity demand if people get up before sunrise otherwise. I believe that testing these theoretical channels or finding better explanations is a fruitful path for future research.

Finally, another potentially important channel that should be examined by future research are supply side reactions to changes in electricity demand. If we believe my results that early sunrise creates long-term higher demand in the South, for example, then in principle this can be used to estimate the slope of the long-run supply curve of electricity production in the South. This is because a change in demand induced by the timing of daylight is unlikely to enter the production function of electricity directly. As a result, both the within-time zone and the time zone boundary variation in the timing of daylight should be valid instruments for estimating long-run electricity supply.

As a final note of caution, the supply side also matters for the interpretation of the results presented so far. If electricity production, for example, exhibits increasing returns to scale, this would affect the interpretation of the reduced form results that I estimated here. This is because people who live in counties with lower electricity demand would potentially pay higher prices for their electricity, thus further reducing their demand depending on the exact slope of the demand curve. Once the slopes of the supply and demand curves are known, the reduced form effects of daylight on electricity consumption could be decomposed into a price and pure quantity effect.

With all these precautionary notes in mind, interpreting my reduced form findings at face value my results would imply that introducing a new time zone boundary which splits the US horizontally along the median latitude could result in substantial residential electricity savings. In 2005 annual residential electricity sales totalled 124.74 billion US dollars (USdOT, 2010) (Table 2.5). Taking my estimates, this means that introducing a horizontal time zone boundary would result in residential electricity savings of about 13 billion US dollars annually, which is equivalent to over 0.1 per cent of GDP in 2005. However, changing the timing of daylight is likely also to affect other outcomes, in particular expenditure for fossil fuel heating, and these should be examined. Future work should also validate the behavioural channels that give rise to the large effects documented here, either using seasonal, or even better micro-data, as I outlined above.

References

Angrist, J. (1998). Estimating the labor market impact on voluntary military service using social security data on mility applicants. *Econometrica*, 66(2):249–288.

Aries, M. B. and Newsham, G. R. (2008). Effect of daylight saving time on lighting energy use: A literature review. *Energy Policy*, 36(6):1858 – 1866.

CEC (2001). Effects of daylight saving time on california electricity use. california energy commission (cec). Technical report.

Hamermesh, D. S., Myers, C. K., and Pocock, M. L. (2008). Cues for timing and coordination: Latitude, letterman, and longitude. *Journal of Labor Economics*, 26(2):223.

Kamstra, M. J., Kramer, L. A., and Levi, M. D. (2000). Losing sleep at the market: The daylight saving anomaly. *The American Economic Review*, 90(4):1005–1011.

Kandel, A. (2007). The effect of early daylight saving time on california electricity consumption: A statistical analysis. Technical report.

Kellogg, R. and Wolff, H. (2008). Daylight time and energy: Evidence from an australian experiment. *Journal of Environmental Economics and Management*, 56(3):207–220.

Kotchen, M. J. and Grant, L. E. (2012). Does daylight saving time save energy? evidence from a natural experiment in indiana. *The Review of Economics and Statistics*.

Levine, R. (1998). *A geography of time: The temporal misadventures of a social psychologist, or how every culture keeps time just a little bit differently*. Basic Books.

Marjit, S. (2007). Trade theory and the role of time zones. *International Review of Economics & Finance*, 16(2):153–160.

Rock, B. (1997). Impact of daylight saving time on residential energy consumption and cost. *Energy and Buildings*, 25(1):63–68.

Sood, N. and Ghosh, A. (2007). The short and long run effects of daylight saving time on fatal automobile crashes. *The BE Journal of Economic Analysis & Policy*, 7(1):11.

USdOT (1975). The daylight saving time study: A report to congress from the secretary of transportation. u.s. department of transport (usdot). Technical report.

USdOT (2010). *Annual Energy Review 2009, U.S. Department of Energy (USdOT)*. Energy Information Administration.

van den Berg, A. (2005). *Health impacts of healing environments: a review of evidence for benefits of nature, daylight, fresh air, and quiet in healthcare settings*. Foundation 200 years University Hospital Groningen.

Table 1: Residential Electricity Consumption in MWh

	All		Pacific		Mountain		Central		Eastern	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Electricity Cons.	12.10	2.70	10.86	3.46	9.70	2.13	12.85	2.20	11.96	2.75
North (1)	10.63	2.86	13.84	1.82	11.68	1.60	11.04	2.32	7.85	1.55
(2)	10.42	2.05	9.90	2.23	9.45	1.31	11.31	1.65	10.18	2.13
(3)	12.29	2.51	7.44	1.48	8.54	0.92	12.39	1.47	13.48	1.71
(4)	13.28	2.31	7.94	2.48	7.85	2.10	13.85	1.76	13.75	1.25
South (5)	13.82	1.86	8.74	2.32	9.26	1.92	14.09	1.65	14.12	0.83
N	26891		1527		2744		12389		10231	

Notes Average per-customer residential energy sales in MWh. Data from the U.S. Department of Energy, forms EIA-f826 and EIA-f861 and utility-level residential energy sales from 2001 to 2009 are matched to US counties based on area of operation of respective utility. Number of counties matched: over 2600, about 1300 in Central TZ, 900 in Eastern, 250 in Mountain and 100 in the Pacific TZ. The first row gives averages for all mainland counties. NORTH (1) to SOUTH (5) divide the USA into five latitude-bands based on the percentile of county centriod. Counties weighed by 2001 census population.

Table 2: Cooling and Heating Degree Days

	All		Pacific		Mountain		Central		Eastern	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Cooling Degree Days	4.56	1.81	2.77	1.47	2.98	1.37	5.38	1.64	4.33	1.65
NORTH (1)	2.46	0.69	1.83	0.68	2.41	0.80	2.89	0.50	2.30	0.49
(2)	3.42	0.85	2.45	0.75	2.43	0.83	4.21	0.59	3.25	0.67
(3)	4.39	0.96	3.28	1.02	2.54	0.99	5.11	0.26	4.37	0.68
(4)	5.20	1.01	4.39	1.33	3.42	1.17	5.75	0.57	4.96	0.92
SOUTH (5)	7.10	0.87	5.88	1.93	5.40	1.13	7.21	0.66	7.18	0.84
Heating Degree Days	5.39	2.02	5.92	1.65	7.10	1.49	5.00	2.11	5.40	1.89
NORTH (1)	7.86	0.74	6.92	0.94	8.08	0.67	8.18	0.40	7.90	0.48
(2)	6.87	0.66	6.62	0.89	7.75	0.73	6.96	0.55	6.76	0.60
(3)	5.54	0.86	5.09	1.51	7.49	0.78	5.42	0.39	5.46	0.58
(4)	4.35	0.82	4.32	1.15	6.13	1.07	4.21	0.45	4.28	0.82
SOUTH (5)	2.55	0.77	2.64	0.70	4.25	0.76	2.64	0.54	2.19	0.82
N	2617		108		250		1317		942	

Notes: Index for heating and cooling degree days in the USA by latitude and time zone. HDD and CDD are indexed from 1 to 7. (HDD: 1 under 1001, 2 1001-2000, 3 2001-3000, 4 3001-4000, 5 4001-5000, 6 5001-6 > 000, 7 6001+. CDD: 1 under 101, 2 100-400, 3 701-1000, 4 701-1000, 5 1001-1500, 6 1501-2000, 7 2001-2500). Source: U.S. Department of Commerce, National Climatic Data Centre, NOAA Satellite and Information Service, NNDC Climate Data, GIS map in ESRI shapefile format, county-level data calculated using the Şzonal statisticsŤ tool. The first row for each panel gives averages for all mainland counties. NORTH (1) to SOUTH (5) divide the USA into five latitude-bands based on the percentile of county centriod. Counties weighed by 2001 census population..

Table 3: Within-TZ analysis: the timing of daylight and residential electricity use

	All latitudes			North			South		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Av. Sunrise Time	0.198**	0.114**	-0.002	0.341**	0.250**	0.163**	-0.161**	-0.162**	-0.131**
	(0.020)	(0.017)	(0.017)	(0.022)	(0.023)	(0.026)	(0.016)	(0.015)	(0.023)
N	26891	26891	26891	13511	13511	13511	13380	13380	13380
Adj. R2	0.16	0.45	0.79	0.24	0.38	0.74	0.64	0.70	0.80
Controls		✓	✓		✓	✓		✓	✓
State FX			✓			✓			✓

Sunrise-time is annual average in local time. All regressions include dummy variables for the year of observation interacted with time zone. Data is for mainland USA countries from 2001 to 2009. Residential energy consumption measure as before. Controls are: year-specific ln(employment), [ln(emp)]2, and index of industry specialisation (which is the number of employees in dominant two-digit industry divided by overall employment), also entered squared, and average Heating Degree Days (HDD), average Cooling Degree Days (CDD), latitude for county centroid, land area, population on 1st July 2001, median age of population in 2000, educational attainment, high school graduate or higher (rate from 1990), persons below poverty level, persons under 18 years of age (percent). North-South split is by latitude-median of county centroids. Standard errors in parentheses and clustered at county level.

** $p < 0.01$

Table 4: TZ boundary analysis: the timing of daylight and residential electricity use

	All latitudes			North			South		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	0.055**	0.042**	0.033**	0.094**	0.089**	0.060**	-0.024*	-0.021	-0.030**
	(0.023)	(0.018)	(0.012)	(0.027)	(0.023)	(0.019)	(0.014)	(0.013)	(0.012)
Boundary Dummy	-0.033*	-0.030**	-0.052**	-0.095**	-0.122**	-0.096**	0.008	0.029**	0.017
	(0.019)	(0.014)	(0.010)	(0.017)	(0.016)	(0.015)	(0.013)	(0.011)	(0.011)
Controls		✓	✓		✓	✓		✓	✓
State FX			✓			✓			✓
N	26891	26891	26891	13511	13511	13511	13380	13380	13380
Adj. R2	0.01	0.45	0.79	0.11	0.36	0.74	0.45	0.67	0.81

Sunrise-time is annual average in local time. For definition of *tzboudnary* and *treatment* look at Figure 4. Data is for mainland USA countries from 2001 to 2009. Residential energy consumption measure as before. Controls are: year-specific ln(employment), [ln(emp)]2, and index of industry specialisation (which is the number of employees in dominant two-digit industry divided by overall employment), also entered squared, and average Heating Degree Days (HDD), average Cooling Degree Days (CDD), latitude for county centroid, land area, population on 1st July 2001, median age of population in 2000, educational attainment, high school graduate or higher (rate from 1990), persons below poverty level, persons under 18 years of age (percent). North-South split is by latitude-median of county centroids. Standard errors in parentheses and clustered at county level.

* $p < 0.05$, ** $p < 0.01$

Table 5: Main results robustness: Including a quadratic term for average sunrise time

	All latitudes			North			South		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Av. Sunrise Time	3.206**	3.725**	3.927**	5.470**	7.177**	3.710**	-0.251	-0.057	3.985**
	(0.857)	(0.657)	(0.753)	(0.924)	(0.806)	(1.183)	(0.680)	(0.569)	(0.891)
Squared: Av. Sunrise Time	-0.223**	-0.267**	-0.291**	-0.379**	-0.513**	-0.262**	0.007	-0.008	-0.305**
	(0.063)	(0.049)	(0.056)	(0.069)	(0.060)	(0.087)	(0.051)	(0.042)	(0.067)
Marginal Ef.	0.190**	0.104**	-0.008	0.354**	0.257**	0.174**	-0.161**	-0.163**	-0.158**
	(0.020)	(0.017)	(0.017)	(0.020)	(0.022)	(0.026)	(0.017)	(0.017)	(0.024)
N	26891	26891	26891	13511	13511	13511	13380	13380	13380
Adj. R2	0.16	0.46	0.79	0.26	0.41	0.74	0.64	0.71	0.81
Controls		✓	✓		✓	✓		✓	✓
State FX			✓			✓			✓

Regressions as in Table 3 but with quadratic term. Marginal effects estimated at means of respective sample.

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$

Table 6: Main results by five latitude bands

	(1)	(2)	(3)
NORTH (1)	0.217**	0.179**	0.178**
	(0.026)	(0.027)	(0.038)
N	5418	5418	5418
Adj. R2	0.62	0.67	0.79
(2)	0.373**	0.282**	0.194**
	(0.027)	(0.033)	(0.039)
N	5438	5438	5438
Adj. R2	0.38	0.49	0.65
(3)	-0.011	-0.039	0.010
	(0.029)	(0.023)	(0.019)
N	5263	5263	5263
Adj. R2	0.68	0.76	0.87
(4)	-0.242**	-0.195**	-0.101**
	(0.023)	(0.024)	(0.031)
N	5394	5394	5394
Adj. R2	0.70	0.76	0.86
SOUTH (5)	-0.154**	-0.167**	-0.222**
	(0.023)	(0.021)	(0.042)
N	5378	5378	5378
Adj. R2	0.51	0.64	0.69
Controls		✓	✓
State FX			✓

Standard errors in parentheses & clustered at county level.

* $p < 0.05$, ** $p < 0.01$

Table 7: Main results robustness: Utilities serving at most 10 counties or at most 1 county

	All latitudes			North			South		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A:									
Av. Sunrise Time	0.177**	0.088**	0.003	0.327**	0.220**	0.191**	-0.172**	-0.184**	-0.134**
	(0.024)	(0.021)	(0.023)	(0.029)	(0.030)	(0.037)	(0.020)	(0.021)	(0.032)
N	23179	23179	23179	11688	11688	11688	11491	11491	11491
Adj. R2	0.13	0.38	0.65	0.21	0.33	0.60	0.49	0.59	0.67
Panel B:									
Av. Sunrise Time	0.069*	0.014	0.087*	0.152**	0.091	0.190**	-0.201**	-0.204**	-0.094
	(0.034)	(0.035)	(0.043)	(0.039)	(0.047)	(0.059)	(0.033)	(0.034)	(0.063)
N	9498	9498	9498	5618	5618	5618	3880	3880	3880
Adj. R2	0.05	0.29	0.59	0.11	0.20	0.49	0.39	0.51	0.64
Controls		✓	✓		✓	✓		✓	✓
State FX			✓			✓			✓

Regressions as in Table 3. Panel A uses only utilities that serve at most ten counties for the calculation of county level residential electricity consumption to minimise measurement error. Panel B uses only utilities that serve exactly one county. Here county level residential electricity consumption can be exactly calculated, at the cost of loosing 65% of counties.

* $p < 0.05$, ** $p < 0.01$

Figures

Figure 1: Length of the solar day, sunrise to sunset

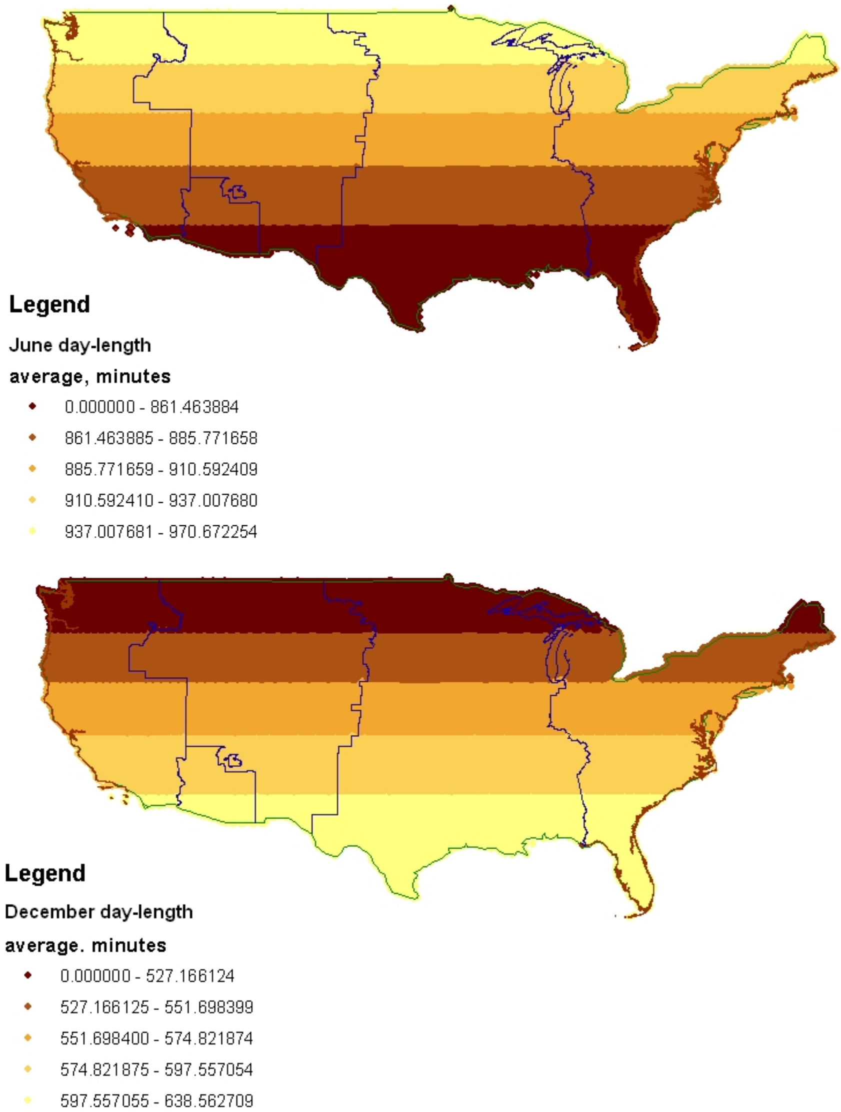


Figure 2: Local standard time of sunrise, June, December and Annual Average

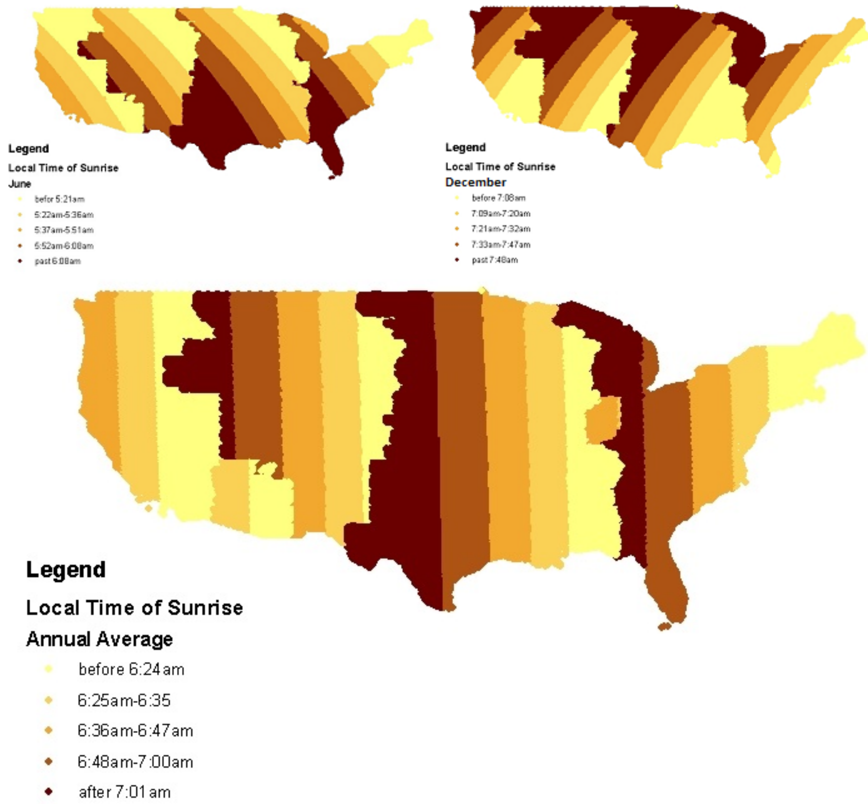


Figure 3: Model: some intuition first

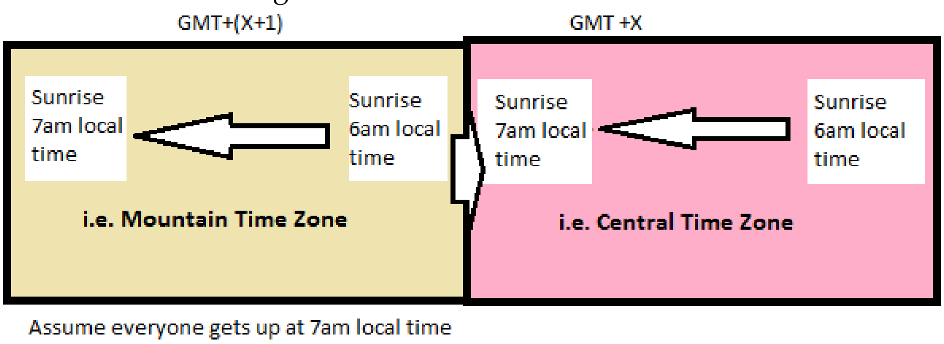


Figure 4: Boundary counties to inland time-zones, excluding Arizona

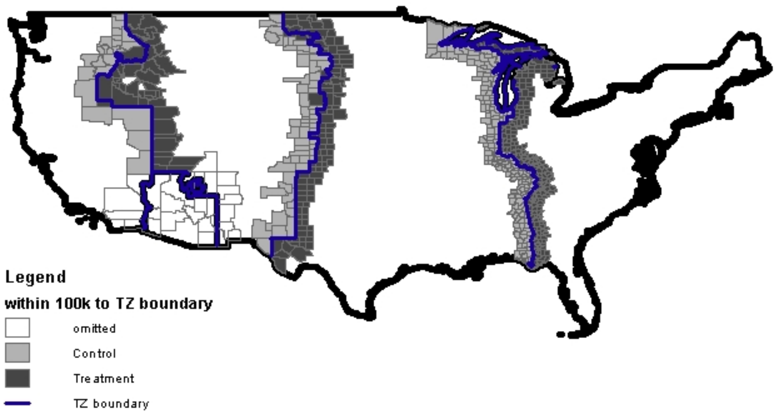


Figure 5: South: later daylight reduced demand for cooling

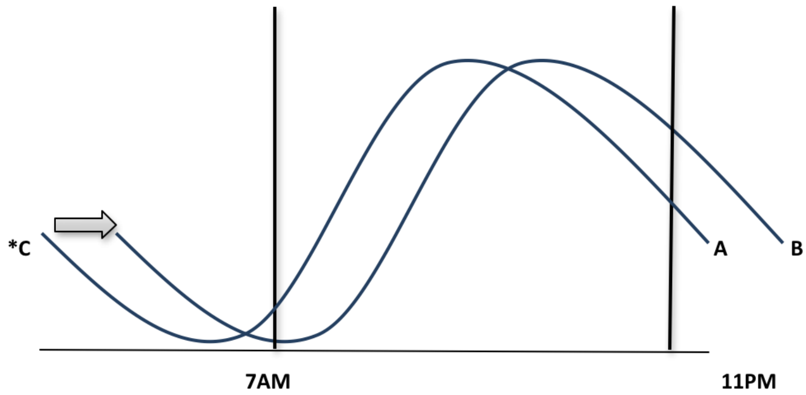
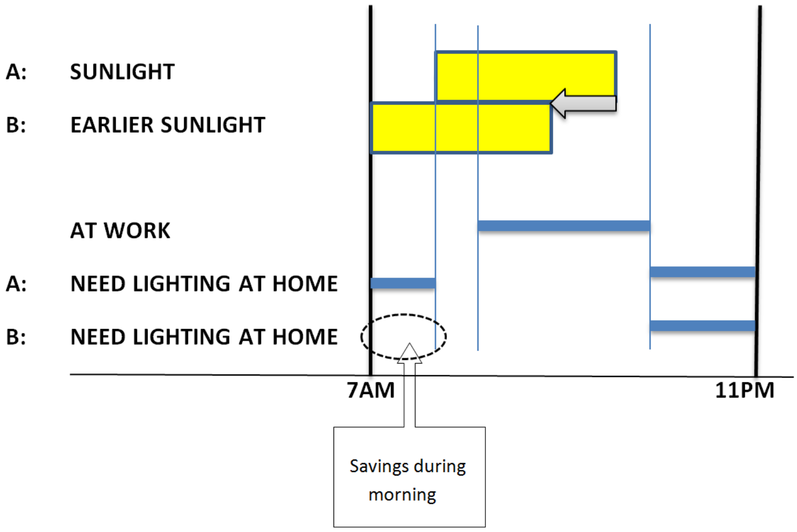


Figure 6: North: earlier daylight reduced demand for lighting



Appendix

Table A.1: County-level control variables

	All		Pacific		Mountain		Central		Eastern	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Population	154321.86	404043.91	398943.26	1052533.16	140465.16	357415.93	104236.65	308626.45	163492.12	275505.79
NORTH (1)	113523.52	202639.97	175691.37	287549.72	44446.81	59877.23	83667.49	155654.64	141252.39	217335.84
(2)	220006.11	430467.83	96042.12	95941.16	168278.36	226491.84	153891.30	572117.97	260886.44	383647.83
(3)	112297.43	211054.63	310562.46	436385.75	100505.19	164698.27	80403.91	147613.91	98185.43	167459.81
(4)	108584.05	205455.41	682318.49	598020.16	95939.09	129144.21	64438.49	108424.12	102943.84	140589.65
SOUTH (5)	201019.32	678989.10	3003335.89	3787174.65	437193.40	877075.18	132460.25	342584.65	157678.16	263005.51
Land Area	955.81	1416.78	2907.60	3259.36	3020.42	2766.38	746.50	431.65	555.86	397.72
NORTH (1)	1265.98	1198.67	1952.46	1369.84	2520.74	1776.43	870.95	677.37	921.92	823.13
(2)	816.15	1241.60	4536.92	3625.07	2051.05	2010.02	611.88	236.87	536.22	262.82
(3)	710.28	1112.66	1935.24	3004.44	1964.69	1296.16	622.13	221.83	400.46	170.10
(4)	1028.91	2058.50	6364.90	5504.57	5074.34	4384.97	676.57	263.97	482.48	198.11
SOUTH (5)	977.34	1108.59	3766.93	1679.38	4783.30	2630.24	846.86	451.52	624.50	362.13
Median Age	36.50	3.68	36.14	4.60	34.70	4.84	36.23	3.55	37.05	3.33
NORTH (1)	37.16	3.77	36.78	4.16	35.54	5.14	37.18	3.56	37.94	2.94
(2)	36.72	3.26	36.56	5.36	32.70	4.80	36.87	3.43	37.09	2.45
(3)	36.75	3.44	36.54	4.88	36.10	4.04	36.64	3.64	36.93	2.91
(4)	36.36	3.38	33.37	4.08	34.70	5.21	36.49	3.10	36.67	3.18
SOUTH (5)	35.62	4.33	33.86	4.38	34.35	3.56	34.99	3.54	36.99	5.29
Education	71.12	10.02	77.72	7.11	77.93	9.14	69.96	9.92	70.18	9.84
NORTH (1)	77.63	5.71	79.61	5.77	79.14	6.56	76.98	5.79	76.78	4.82
(2)	77.11	5.73	77.08	4.15	83.80	6.51	78.27	4.75	75.92	5.54
(3)	71.09	10.48	78.36	8.02	80.59	8.69	72.60	8.53	67.81	10.60
(4)	64.25	9.32	73.07	6.92	71.35	9.12	64.18	8.89	62.84	9.20
SOUTH (5)	65.99	9.39	70.10	11.37	69.18	8.26	65.10	9.24	67.01	9.53
Poverty Rate	20.00	7.99	20.63	7.67	20.89	9.47	20.32	8.17	19.50	7.62
NORTH (1)	16.23	5.76	17.96	4.79	19.53	7.16	13.72	5.25	17.28	5.04
(2)	15.36	5.92	21.97	7.98	13.77	5.31	14.13	4.40	15.62	6.08
(3)	19.15	7.79	20.37	8.46	18.68	9.04	18.19	6.02	19.57	8.40
(4)	23.02	6.88	25.73	7.88	29.04	9.27	23.34	6.71	21.87	6.26
SOUTH (5)	25.98	7.78	30.80	10.74	29.63	7.24	25.75	7.78	25.72	7.55
Industry Spec.	0.26	0.10	0.21	0.07	0.24	0.10	0.26	0.10	0.26	0.10
NORTH (1)	0.24	0.08	0.21	0.06	0.25	0.11	0.27	0.09	0.23	0.06
(2)	0.25	0.09	0.24	0.08	0.22	0.09	0.25	0.09	0.25	0.09
(3)	0.26	0.09	0.21	0.09	0.25	0.10	0.25	0.09	0.27	0.10
(4)	0.28	0.11	0.19	0.05	0.26	0.08	0.29	0.11	0.29	0.11
SOUTH (5)	0.25	0.10	0.18	0.07	0.22	0.09	0.25	0.10	0.26	0.10
Employment	35392.16	134444.38	99248.83	348241.04	19937.03	96567.02	22758.57	103083.08	45243.56	115045.41
NORTH (1)	24541.91	77751.64	42996.32	123072.34	6252.29	17752.56	19842.29	65956.20	39816.05	90817.44
(2)	56504.21	164213.71	21294.05	38019.53	29758.29	83415.78	29359.00	163178.63	88056.00	179921.12
(3)	25684.37	77950.56	86175.28	186269.90	11763.48	36436.79	18359.23	58189.35	25482.55	67429.96
(4)	26550.03	74033.12	205743.09	224925.70	16219.75	42429.73	15380.98	46266.32	31553.63	74851.73
SOUTH (5)	43670.77	212213.57	1078392.30	1370528.48	91445.76	299237.52	28996.59	125634.13	34969.90	92032.99

Notes: Number of counties: 2979, 170 in Pacific, 305 in Mountain, 1375 in Central and 1129 in Eastern time zone. Variable descriptions: From ICPSR 2896 Historical, Demographic, Economic, and Social Data, DS81: 2000 County Data Book (County and State): Population in 2001 (July 1), Land area in square miles, Median age in 2000, Education: Educational attainment, high school graduate or higher (1990), Poverty Rate: Persons below poverty level, persons under 18 years of age (percent). From County Business Pattern, averages for 2001 to 2009: Industry specialisation: main two-digit employer divided over total industry employment. Employment: total employment. The Census 2001 county-level population information is used as analytic weight.

Table A.2: Residential electricity consumption in MWh for time zone boundary counties

	All		Pacific		Mountain		Central		Eastern	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Electricity Cons.	12.06	2.70	12.16	1.71	10.90	2.10	11.41	2.80	12.67	2.65
NORTH (1)	9.57	2.40	12.85	1.10	12.29	1.40	8.82	1.26	7.32	0.81
(2)	10.97	1.77	11.41	2.15	10.72	1.67	10.03	1.42	11.52	1.66
(3)	13.36	1.86	10.52	1.23	8.39	1.03	12.84	1.70	14.11	0.99
(4)	13.90	2.04	.	.	9.38	1.68	14.17	2.34	14.16	1.27
SOUTH (5)	13.37	1.88	.	.	9.40	1.53	13.22	1.90	14.04	0.89
N	5508		369		1062		1746		2331	

As in Table 1 but only for the 612 countries at time zone boundary, see Figure 4

Table A.3: Cooling and Heating Degree Days for time zone boundary counties

	All		Pacific		Mountain		Central		Eastern	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Cooling Degree Days	4.25	1.46	2.32	0.79	3.68	1.16	4.24	1.41	4.59	1.40
NORTH (1)	2.45	0.61	1.96	0.64	2.92	0.77	2.60	0.46	2.19	0.35
(2)	3.63	0.59	2.64	0.47	3.58	0.33	3.77	0.57	3.71	0.48
(3)	4.80	0.47	3.32	0.90	4.22	0.78	5.00	0.00	4.86	0.27
(4)	4.87	0.50	.	.	4.44	0.77	5.03	0.26	4.81	0.54
SOUTH (5)	6.39	0.79	.	.	5.10	1.08	6.38	0.59	6.59	0.63
Heating Degree Days	5.79	1.86	7.47	0.95	6.63	1.59	6.02	1.95	5.31	1.74
NORTH (1)	8.00	0.66	7.63	1.05	7.75	0.81	8.21	0.45	8.03	0.37
(2)	6.95	0.48	7.44	0.49	7.22	0.44	7.04	0.54	6.79	0.37
(3)	5.42	0.55	6.78	0.91	6.33	0.52	5.30	0.43	5.31	0.37
(4)	4.58	0.56	.	.	4.98	0.82	4.46	0.46	4.62	0.57
SOUTH (5)	2.99	0.73	.	.	4.34	0.94	3.07	0.50	2.74	0.55
N	612		41		118		194		259	

Notes: As in Table 2 only for countries at time zone boundary, see Figure 4

Table A.4: Control variables for TZ-boundary counties

	All		Pacific		Mountain		Central		Eastern	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Population	103227.90	348927.73	42341.31	31397.32	42286.01	63554.73	169484.72	576696.51	77594.61	121466.15
NORTH (1)	80259.78	137996.17	41859.62	34066.57	57447.82	84393.50	118281.03	175999.61	58290.39	123415.66
(2)	239094.06	769162.94	39202.68	18969.32	16441.96	12936.75	577024.92	1277085.36	78473.83	66122.23
(3)	55963.05	99783.66	50465.81	41154.64	10548.25	6270.95	39435.36	41977.82	66127.96	119289.80
(4)	87868.25	142556.07	.	.	22826.96	14752.99	65444.29	63399.40	106581.05	175357.77
SOUTH (5)	63721.16	60508.07	.	.	44515.26	22005.27	60253.90	56677.97	68454.02	66122.81
Land Area	933.96	1415.71	3729.93	3652.10	2357.25	1693.68	751.61	727.03	482.21	230.76
NORTH (1)	1246.06	1158.29	2187.18	1394.31	1862.25	1335.86	986.97	1060.17	763.64	350.84
(2)	1162.20	2395.18	7131.23	5405.09	1830.25	743.13	607.20	348.95	494.87	161.47
(3)	650.69	1037.53	4145.51	2695.31	2165.86	1499.24	505.54	234.39	369.79	113.58
(4)	585.56	557.63	.	.	2270.68	1367.36	577.33	303.72	438.19	165.90
SOUTH (5)	960.31	1336.96	.	.	4314.80	2077.28	921.99	795.69	461.56	196.25
Median Age	36.19	3.51	34.50	5.78	35.44	4.51	36.42	3.17	36.36	3.12
NORTH (1)	37.69	4.26	37.18	5.35	35.39	5.23	37.97	3.16	38.97	3.86
(2)	35.17	3.21	30.72	3.14	39.31	2.71	34.63	3.00	35.87	2.54
(3)	35.97	3.12	29.79	4.77	37.13	3.04	36.58	3.51	36.05	2.36
(4)	36.36	2.76	.	.	33.16	3.84	36.17	2.30	36.76	2.71
SOUTH (5)	34.94	2.93	.	.	34.19	2.08	35.51	2.68	34.77	3.15
Education	69.73	10.58	79.17	5.44	73.53	8.52	70.83	10.34	67.26	10.60
NORTH (1)	76.38	6.00	79.29	5.33	76.24	7.78	76.86	5.56	74.28	4.99
(2)	77.05	5.46	76.60	5.22	79.21	2.68	79.48	5.35	75.42	5.24
(3)	68.37	9.97	83.54	4.04	73.41	5.43	67.93	8.03	67.19	10.33
(4)	60.68	10.12	.	.	67.42	7.98	61.85	10.35	59.41	9.94
SOUTH (5)	64.15	9.25	.	.	66.91	8.79	63.30	7.92	64.16	9.97
Poverty Rate	19.64	7.87	18.57	5.09	23.39	8.45	17.62	7.69	20.44	7.81
NORTH (1)	17.45	7.07	20.53	3.93	21.42	8.78	13.29	6.35	19.52	4.68
(2)	13.73	4.96	14.08	6.03	16.22	2.64	12.31	5.58	14.36	4.41
(3)	19.06	6.75	18.41	2.24	25.22	7.84	19.29	5.09	18.57	7.26
(4)	22.87	7.23	.	.	30.59	7.67	20.93	5.53	23.29	7.58
SOUTH (5)	26.28	7.20	.	.	26.77	4.56	26.87	5.72	25.90	8.19
Industry Spec.	0.28	0.11	0.23	0.08	0.26	0.13	0.27	0.11	0.29	0.11
NORTH (1)	0.26	0.09	0.21	0.06	0.26	0.14	0.28	0.09	0.26	0.06
(2)	0.28	0.12	0.27	0.10	0.28	0.13	0.21	0.09	0.33	0.12
(3)	0.28	0.10	0.24	0.09	0.29	0.12	0.27	0.10	0.28	0.10
(4)	0.31	0.13	.	.	0.29	0.10	0.32	0.12	0.32	0.13
SOUTH (5)	0.27	0.12	.	.	0.22	0.10	0.28	0.13	0.28	0.11
Employment	44083.09	163082.07	14156.20	13908.27	15702.72	33562.79	73552.36	258944.71	33209.86	81970.77
NORTH (1)	35094.93	72830.12	14011.99	15067.77	24463.73	44997.95	54703.08	90349.13	23666.13	69109.54
(2)	102329.37	343172.89	12484.94	8593.60	4703.72	4331.23	254755.36	562018.94	29559.33	33253.86
(3)	23084.82	56961.35	17985.76	16075.42	2314.91	1592.94	16326.52	26548.00	27588.37	67913.69
(4)	41783.34	108000.56	.	.	4771.37	3866.34	24217.27	31510.23	55187.80	135544.08
SOUTH (5)	21113.23	25777.76	.	.	11615.34	6395.06	20562.06	28371.24	22865.18	25961.55

Notes: Number of counties within 100km of inland boundary: 612, 41 in Pacific, 118 in Mountain, 194 in Central and 259 in Eastern time zone. Variable descriptions as in Table A.1.

CHAPTER VI

CONCLUSION

Conclusion

I have argued in the introduction of this thesis that econometric methods can be applied to geographical questions in order to gain a better understanding of the human-environment relationship. In this thesis, I have presented four main chapters that look at specific settings where geography is arguably related to socio-economic outcomes.

In Chapter II I estimated the effect of moving into deprived social housing neighbourhoods on teenage test scores in England. Numerous sociological and economic theories hypothesis a negative effect from living in deprived neighborhoods on individual outcomes such as cognitive outcomes at school. Overall, my analysis shows that there is little evidence to support such claims at least in the short run. I do not find evidence of negative effects of moving into high density, deprived social housing neighbourhoods during the early teenage years once I credibly control for otherwise unobserved geographical sorting. I believe that this finding is informative for social housing and area based policies.

Chapter III presents work undertaken jointly with Steve Gibbons and Olmo Silva where we examine the importance of peers in the neighbourhood. Using a very rich panel of pupils in England, we can identify similar-aged students at the neighbourhood of residence and use variation coming from residential mobility for estimation. Notice that contrary to Chapter II, this study isolates the effect of social interaction on student outcomes. Overall, there is little support for the notion that peers in the neighbourhood matter for own cognitive outcomes.

Chapter IV is based on joint work with Victor Lavy and Olmo Silva and a slightly extended version of this work has recently been accepted for publication in the Journal of Labor Economics. Here, we turn the attention towards peers at schools and show that there are negative effects on cognitive outcomes from having a high proportion of poorly performing peers at school. Contrary, there is no overall evidence for positive effects stemming from a high proportion of very bright peers, but this finding masks some heterogeneity along the gender dimension with girls marginally benefitting and boys loosing out from the presence of very academically able peers. Again, our results provide important evidence for policy intervention.

Finally, in Chapter V I ask the question how human behaviour is affected by the timing of daylight, specifically studying household electricity consumption in the

United States. To do this I first calculate the average solar and current times of sunrise for all locations in mainland US to derive stylised facts about spatial variations in the annual timing of daylight. Building on these findings I show in a model that two different sources of geographical variation in local sunrise times can be used to estimate the effect on electricity consumption. Using both I find that earlier daylight is associated with higher electricity consumption in the South, whereas the opposite holds for the North. I provide a first attempt to explain these findings in terms of behavioural channels and outline directions for future research. If my results are confirmed, additionally splitting the U.S. into time zones horizontally could result in substantial cost savings.

Overall, I believe that my work presented here makes a compelling case for using applied econometrics in geographical settings. As I have demonstrated particular attention has to be paid towards the geographical sorting problem as any unobserved geographical heterogeneity could potentially confound the analysis. I believe that in addition to advances in the availability and handling of geographically localised data, close attention to research design will continue to play an important role in the analysis of socioeconomic phenomena.