CESifo WORKING PAPERS

8846
2021

January 2021

Home Broadband and Human Capital Formation

Rosa Sanchis-Guarner, José Montalbán, Felix Weinhardt

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8846*

Home Broadband and Human Capital Formation

Abstract

This paper estimates the effect of home high-speed internet on national test scores of students at age 14. We combine comprehensive information on the telecom network, administrative student records, house prices and local amenities in England in a fuzzy spatial regression discontinuity design across invisible telephone exchange catchment areas. Using this strategy, we find that increasing broadband speed by 1 Mbit/s increases test scores by 1.37 percentile ranks in the years 2005-2008. This effect is sizeable, equivalent to 5% of a standard deviation in the national score distribution, and not driven by other technological mediating factors or school characteristics.

JEL-Codes: J240, I210, I280, D830.

Keywords: broadband, education, student performance, spatial regression discontinuity.

*Rosa Sanchis-Guarner**
Queen Mary University of London
School of Economics and Finance
London / United Kingdom
r.sanchis-guarner@qmul.ac.uk

José Montalbán
SOFI at Stockholm University
Stockholm / Sweden
jose.montalban@sofi.su.se

Felix Weinhardt
European University Viadrina
Frankfurt (Oder) / Germany
weinhardt@europa-uni.de

*corresponding author

January 14, 2021
For comments, we thank Antonio Cabrales, Martin Fernández-Sánchez, Alexandra de Gendre, Clément Malgouyres, Stephan Maurer, Marion Monnet, Diego Puga, and participants at seminars and conference presentations. We are especially grateful to Ben Faber, without whom this project would not have materialized. Felix Weinhardt gratefully acknowledges ESRC seed funding (ES/J003867/1) as well as support from the German Science Foundation through project WE5750/1-1 and the CRC TRR 190 (project 280092119). Rosa Sanchis-Guarner acknowledges funding from LSE/CEP and from the British Academy. This is a revised version of ”ICT and education: Evidence from student home addresses” National Bureau of Economic Research (No. w21306).

1 Introduction

Over the past two decades, technological advances in information and communication technology (ICT) have dramatically changed the education landscape. Currently, the educational technology (*edutech or edtech*) industry is booming, and most researchers, policymakers and educators agree on the importance of incorporating these technologies into the learning environment. However, as Escueta et al. (2017) note, "researchers and educators are far from a consensus on what types of EdTech are most worth investing in and in which contexts" (p. 3), and Deming et al. (2015) call for more research to investigate the impact of online technology on education.[1] One crucial ICT is the internet, which is used for educational purposes not only in schools but also at home, complementing classroom education with additional online learning. While home internet can increase learning productivity and widen access to educational opportunity, it may also lead to unproductive distraction, making its net effect on student human capital formation ambiguous (Bulman & Fairlie, 2016).[2] Because online learning is likely to remain a key input in the education system, identifying whether high-speed (broadband) internet impacts student educational outcomes has important economic and policy implications. Notably, this is crucial in the advent of temporary shocks, such as the COVID-19 pandemic, during which many countries closed schools for several months, and as a consequence, home online learning became pivotal to children's education.

Even though understanding the relationship between high-speed internet and education is a first-order empirical question, major empirical challenges have limited the scope of previous research. This is because the estimation of this effect entails several identification issues. First, *observed* home internet subscription choices (i.e., connection speed) are nonrandom and likely related to learning outcomes through unobserved household-level confounding factors. We refer to this as *active* selection. An analysis of the relationship between package internet speeds and educational outcomes would probably suffer from selection bias. An alternative is to use a measure of local *available* internet speed, which is determined by the distance between the household and the telephone local exchange (LE) station providing them with telephone and internet services (for DSL connections). While it is unlikely that households actively sort in locations on the basis of potential DSL-speeds, locations at different distances of the LE station may also have different local neighborhood characteristics that do matter for household sorting. For this reason, simply comparing households in locations connected to the same LE station can still lead to biased esti-

[1] Escueta et al. (2017) define edtech as "any ICT application that aims to improve education". Some examples are e-learning platforms, distance learning tools, and massive online open courses (MOOCs). Edtech can also include provision of software or hardware such as computers.

[2] The accumulation of human capital, i.e., the stock of skills, traits and knowledge that an individual possesses (Burgess, 2016), is key for growth, employment and earnings (Schultz, 1961; Becker, 1962; Mincer, 1974; Barro, 2001). Educational outcomes, such as the cognitive skills that people have learned, have been found to be a reliable proxy of human capital (Hanushek & Woessmann, 2011, 2012, 2015).

mates. We call the correlation between local geography and local available internet speed *passive* selection. Hence, estimating causal effects requires an identification strategy that overcomes both types of selection issues.

In this article, we apply a careful identification strategy able to overcome these challenges. We estimate the causal effect of home high-speed internet on teenagers' test scores. For this purpose, we combine rich comprehensive geolocated administrative data from England to exploit a (fuzzy) spatial regression discontinuity (SRD) design.

In Section 2, we describe the multiple sets of administrative microdata that allow us to meet the extensive data requirements for implementing our empirical strategy. First, we use administrative standardized and externally marked test score records for the population of 14-year-old English students in national Key Stage 3 (KS3) tests over the period 2005–2008, along with rich information on their background characteristics. This includes a key variable, the student pre-internet score, which allows us to estimate value-added regressions. A key feature of our data is that we are able to georeference student residence and school at the most disaggregated spatial scale, i.e., the postcode level, which roughly corresponds to blocks of approximately 15 households. Second, we use telecommunication network data including the position of the universe of English LE stations (approximately 3,900) and their assignments to each of roughly 1.45 million full postcodes. We complement this with postcode-level internet speed measures for 2012–2014. Third, we employ a rich vector of georeferenced control variables that allows us to compute residential proximity to a comprehensive list of local amenities. These data include the universe of property transaction values (from which we can construct local average house prices).

To overcome *active* and *passive* selection, we use a well-known feature of digital subscriber line (DSL) broadband technology in the design of our estimation strategy: the length of the copper wire that connects residences to the telephone LE station, which is a key determinant of *available* local internet connection speeds. In this context, to deal with *active* selection, a potential strategy would be to compare outcomes of students whose residences are located at different distances to LE stations and who hence enjoy different potential home internet speed quality. However, residential distances to the connected LE stations are not randomly assigned across space because stations are located in places with particular location characteristics, potentially leading to *passive* selection. To address these concerns, we explain in Section 3 that we focus on the invisible boundaries generated across LE stations. We note that each LE station has an invisible catchment area of residential addresses that it serves in its surroundings. The extent and shape of this catchment area is a byproduct of history: rapid growth in fixed-line telephony during and after World War II, in combination with capacity constraints at the exchange switchboards, led to invisible and essentially randomly placed station-level catchment area boundaries. In our strategy, we focus

on households whose residence is located in the vicinity of these invisible boundaries, exploiting variation in distances to the connected station across small segments, each side connected to a different LE station. The causal effect of broadband speed on student test scores is identified by comparing "lucky" households that are supplied with faster broadband access (the side with shorter distances on average) to otherwise similar counterparts that were "unlucky", supplied with slower broadband access (the side with longer distances on average).[3] Due to the irregular geographic shape of the boundaries, some households with short cables (long cables) might live on the slower side (faster side). Hence, our SRD design is *fuzzy*, with the *sharp* SRD design affected by substantial attenuation bias pushing the estimates towards zero.

Our main finding is that broadband quality has positive effects on national externally marked test scores. We find that moving 100 meters closer to the LE station increases student test scores by 0.122 percentile ranks. In Section 4, we present a battery of robustness checks to validate our identification strategy and support our conclusions. In particular, the results are robust to controlling for school-specific broadband availability features and to including school fixed effects, which suggest that the findings are not driven by school characteristics.

Although our main estimate may seem relatively small, in Section 5, we show that its effect size is economically meaningful. To assess the magnitude of the baseline estimates, we employ out-of-sample-period data to reverse engineer the distance-speed relation present during our study period. We combine linked postcode-to-exchange station telecom network data with data on local internet speed measures experienced by households. We find that for each additional 100 meters closer to the connected LE station, the local average speed increases by 0.089 Mbit/s. This means that for each increase in the average broadband speed of 1 Mbit/s, test scores increase on average by 1.37 percentile points. This average effect of one additional Mbit/s is equivalent to approximately 5% of a standard deviation in the national test score distribution.

The empirical setting of this study is England over the time period 2005–2008, which offers several advantages. First, home broadband was predominantly delivered via a stable technology in our study setting: that is, via asymmetric DSL (ADSL) through telephone copper wires.[4] In addition, other non-distance sensitive technologies (cable and fiber) became more widespread in the UK only after 2008, which also coincides with the last years of availability of the detailed KS3 records. Second, at this time, the broadband market was already developed in England. In 2008,

[3] We use the words broadband availability and access interchangeably to refer to the possibility of subscribing to broadband services from one or more providers in a postcode.

[4] In the empirical context that we study, the transition from dialup connections to ADSL occurred in the early 2000s. Initially, the usage of broadband subscriptions was low due to prices being high and consumers not considering higher internet speeds attractive enough to pay a premium. By 2005, the broadband market in England had matured in the sense that coverage was close to universal (see Nardotto et al., 2015, for a detailed discussion of the English broadband market). Similar to Amaral Garcia et al. (2019), we therefore use years following 2005 to estimate the effects of faster broadband. In this period, the ADSL, ADSLmax and ADSL2+ technologies were available, but all are affected by signal decay with increasing copper cable length.

84% of students used the internet, and among those, 90% used it for their homework, with an average connection speed of 4.1 Mbit/s (Livingstone & Bober, 2005; OfCom, 2009a) . Currently, 59% of the world population is online (Clement, 2020), and for those connected, the average worldwide connection speed in 2015 was 5.6 Mbit/s (Inc., 2015). We therefore believe that our estimates based on English student population data in the mid- to late-2000s have high external validity for other countries today and can inform current policy.

Our paper is related to the large and growing literature on the relationship between ICT and education outcomes.[5] An important strand of the literature analyzes the effects of ICT in school settings. Most of this previous work uses experimental and quasiexperimental methods, finding mixed results but typically no consistent impacts on math or reading educational achievement (Angrist & Lavy, 2002; Rouse & Krueger, 2004; Goolsbee & Guryan, 2006; Machin et al., 2007; Belo et al., 2014, 2016; Falck et al., 2018). Our study is different because we study the effects of broadband access at home. A different set of studies has analyzed the impact of providing access to academic software specially designed for students, with many papers showing positive effects on math and reading (Banerjee et al., 2007; Barrow et al., 2009; Barrera-Osorio & Linden, 2009; Muralidharan et al., 2019). In contrast, we study the effects of a policy variable (home broadband) that is not directly targeted toward education. Another group of papers focuses on the relationship between home computer access and education outcomes, in which some of the earliest papers identify empirical associations rather than causal estimates (Battle, 1999; Fairlie et al., 2010; Fiorini, 2010). Recently, this literature has used quasiexperimental methods as well as randomized interventions to identify the causal effect of home computer access on student outcomes (Malamud & Pop-Eleches, 2011; Fairlie & Robinson, 2013; Vigdor et al., 2014; Beuermann et al., 2015; Cristia et al., 2017). These articles often report positive effects on outcomes directly related to computer access but no impact – or only a modest one – on student academic outcomes.

Our paper is most closely related to the narrower and relatively recent literature linking home broadband technology to student test scores. Malamud et al. (2019) find no significant effects of home internet access on student achievement. This result is based on a credible randomized controlled trial implemented in several low-achieving primary schools in Peru. Our paper expands on Malamud et al. (2019) by identifying effects based on a broader population that covers all socioeconomic levels of school-age teenagers. In another key study, Dettling et al. (2018) show that students with broadband access in their postal codes perform better on the SAT and apply to a larger set of colleges in the US. We complement Dettling et al. (2018) by analyzing the impact on

[5] A growing literature on the impact of broadband in socioeconomic outcomes includes papers on its positive effects on labor productivity and wages (Akerman et al., 2015), economic growth (Czernich et al., 2011), capitalization of the property market (Ahlfeldt et al., 2017), health choices (C-sections) (Amaral Garcia et al., 2019) and marriage rates (Bellou, 2015), and negative effects on political participation (Falck et al., 2014; Campante et al., 2018; Gavazza et al., 2019) and sex crime (Bhuller et al., 2013).

tests that are low stakes from the student perspective and are specifically designed to test cognitive ability with no explicit online-training resources.

The contribution of our paper is fourfold. First, we link several sources of administrative microdata to trace the broadband available to the universe of English students over four years at the finest geographical level. The richness of the data allows us to exploit discontinuous changes in broadband quality across neighboring residences to implement an estimation strategy that causally estimates the impact of broadband quality on student test scores, addressing *active* and *passive* sorting and attenuation bias. Second, while our main estimates are for the impact of broadband availability on test scores, our methodological approach allows us to underpin the direct relationship between broadband speed and student test scores. This second parameter is relevant for assessing the impact of policy interventions aimed at boosting local speeds or subsiding the takeup of higher-speed packages. Third, a main advantage of our outcome of interest is that it is a low-stakes exam from the students' perspective and designed to test student progress in the English education curriculum with no explicit online-training resources. This means that our estimates inform us about the impact of the home environment on the learning and knowledge accumulation that determines human capital formation in a general sense, in contrast to specific ICT skills or targeted preparation for specific test-taking. Finally, our paper shows the importance of human capital accumulation in the home environment for outcomes measured at the school level. We are able to identify the isolated impact of home broadband on student test scores, abstracting from school mediating (technological) factors that may affect student performance. Therefore, our findings imply that broadband technology affects the learning nexus of home and school education, complementing school learning.

2 Background, Data and Descriptive Statistics

2.1 Broadband Expansion in the UK

The rollout of DSL broadband technology in the UK started in the major urban centers at the beginning of the 2000s and proceeded rapidly. This process involved technological upgrades of the infrastructure of telephone LE stations – the same ones that provide telephony services to a number of connected premises around them – to allow them to offer broadband internet services through copper cable. By the end of 2004, 80% of the LE stations had been equipped to provide broadband services, covering 97% of local residences, which could subscribe to receive broadband services at home. That year, 54% of households had an internet connection, of which 6.2 million (approximately 25%) were broadband. By the start of our estimation period in 2005, 99% of English addresses were connected to broadband-enabled telephone LE stations.

Even if most of the technological upgrades took place between 2000 and 2005, penetration rates were low in the first years (approximately 10% in 2003) and only started growing in 2004. By then, infrastructure was completely rolled out across space, and the takeup rate increased steadily. The broadband internet takeup rate rose from approximately 30% in 2005 to over 60% by 2008 (Eurostat). This increase in takeup was related to decreases in prices and changes in attitudes and internet content. Due to this, and similarly to existing work (Nardotto et al., 2015), in our analysis, we focus on the post-2005 period. In this context, we can focus on the impact of broadband speed for a given state of technology and exploit very local variations in quality.

In 2007, more than half of UK homes had broadband access, with an average connection speed of 4.6 Mbit/s (OfCom, 2009a). While today that average speed is faster, the coverage and speed available to households in the UK between 2005 and 2008 is comparable to the infrastructure currently available in large parts of the world. For comparison, in 2017, only approximately 14% of the world population had broadband access, with average connection speeds of 7.2 Mbit/s (McKeay, 2017). Note that our period of analysis ends in 2008 for two reasons: (i) mobile broadband and cable/fiber internet technologies became more widespread in the UK after 2009, reducing the efficacy of our empirical approach, and (ii) the standardized exam that we use to measure the educational achievement of teenagers in this paper (the KS3 for students at age 14) was discontinued. The combination of the testing regime and the state of development of the broadband infrastructure in the UK in the period 2005-2008 offers a unique opportunity to study the effects of home broadband quality on student performance.

2.2 Data

2.2.1 Administrative Student Records

In the English educational system, student academic performance is assessed in national exams that are administered through externally marked tests. The English education curriculum is organized into four key stages (KSs). Compulsory education starts at age 6 and ends at age 16 with the fourth and final KS4 (the General Certificate of Secondary Education [GCSE] examinations). There are several reasons why the KS3 exam at age 14 is the most suitable for our analysis. First, the KS3 test is externally marked and thus comparable across students and schools. Second, the test is low stakes, so there are no incentives for teachers or students that would drive a wedge between test scores and real achievement. Third, the test is finely graded (mostly zero to 100); therefore, in combination with our sample size, it is possible to detect even small effect sizes. Finally, all students are tested in the three main compulsory subjects: English, mathematics and science. Students have very limited options in choosing subjects or specializing according to interest or ability before the KS3, in stark contrast to the educational period before the KS4 test two years later. This feature of

the KS3 exam makes it particularly suitable to test for heterogeneity across groups that might later on (endogenously) specialize in different fields.

We employ administrative data containing information on the universe of students enrolled in English state schools (approximately 95% of pupils) who took the KS3 test from 2005–2008 in England. These data are supplied by the Department for Education (DfE). To match the student information with the telecom network data that we describe below, we first use the restricted-access version of the National Pupil Database (NPD), from which we extract the full residential postcode for each registered student in a given year. British postcodes are associated with a small number of addresses (15 on average) and in denser areas usually correspond to housing blocks. In the second step, we use the unique student identifiers to link their residential information to individual test score results, which are also provided as part of the NPD.[6]

Following the education literature, we transform these scores into percentile ranks for each test and cohort, i.e., separately by year-subject. These subject percentiles are then added into a total score, which we percentilize to obtain an average total score ranging from 1 to 100. We conduct this transformation to make our results comparable to other countries' national exams as well as across cohorts/subjects. Transforming raw scores into percentile ranks has the goal of keeping the ordinal information in the outcome variable and removing the cardinal differences between units of interest, which might be driven by the setup of any particular exam paper, for instance.

We use additional data from the DfE NPD for each pupil in our KS3 2005–2008 sample and collect information on their KS1 test scores (taken at age 7). For this sample, this corresponds to tests taken during 1998–2001, when most of the rollout of broadband internet had not yet taken place and the level of broadband takeup was essentially zero. In contrast with the KS3, this test is marked by the schools, is only available for the subjects of mathematics and English, and is graded on a coarse scale. However, adding this information to our empirical models allows us to estimate individual-level value-added results, controlling for pupil-specific time-invariant ability and background characteristics, which in turn improve the precision of the estimates and the explanatory power of the models. We also obtain information on the location, size and type of the school that the pupils attend, which we use to construct school-level controls and, for some specifications, school fixed effects.

In addition to test scores, the administrative data give us access to a series of observable student characteristics, such as gender, ethnicity, and student eligibility for a free school meal (FSME), which is a common proxy for family income. We exploit these data at two scales: to construct individual-level controls and to calculate postcode-year-specific demographics based on the pop-

[6]The DfE formerly distinguished between the NPD and the Pupil Level Annual School Census (PLASC), which is now treated as part of the NPD. Note that no information is available on private schools, which enroll approximately 6–7% of the English student population (Ryan & Sibieta, 2010).

ulation of pupils of all ages, which we also use as local area control variables in the regressions.

2.2.2 Average House Prices and Area Socioeconomic Characteristics

We use a number of additional datasets to improve precision and to validate our approach. First and foremost, we use transaction-level data on property sales in England over the estimation period. The data are administrative records from the England and Wales Land Registry, covering all property transactions over this period. We use the reported property address information to link these property transaction values to individual residential postcodes. The postcode-year averages are based on several million individual property transactions that occurred in England over the period 2005–2008. Local house prices capitalize many desirable (and undesirable) local attributes and are likely to capture a large number of unobserved spatial characteristics of the areas.

Even though our empirical analysis is based on a spatial discontinuity design that compares only very proximate households, it could still be the case that catchment area boundaries coincide with physical barriers such as roads or rivers and that either the slower or the faster side of the boundaries has a higher likelihood of hosting a given type of local amenity, the combination of which could lead to bias in the boundary effect. Using a GIS with detailed attribute data from the UK Ordnance Survey, the commercial real estate consultancy CBRE and the DfE, we compute euclidean distances between each English postcode and the following features: nearest school (primary or secondary), nearest road (class A, class B and motorways), nearest rail station (which captures centrality), nearest water body (river, stream, marsh or lake) and nearest supermarket. One of the major concerns is that *passive* endogeneity arises because local geography correlates with the location of the LE stations and of households. Taking these variables into account allows us to properly test whether observable geographic features are an endogeneity concern in our setting and ultimately control for these variables in our empirical specification to increase statistical precision.

Finally, we combine data from the Office for National Statistics Postcode Directory (ONSPD) and the DfE to control for local density by calculating the number of premises in each postcode (which is fairly stable over time) and the number of students (of all ages) per premise.

2.2.3 Postcode Broadband Speed Data

Note that in Section 5, we use additional data on postcode-level realized internet speed from Ofcom, the British telecom regulator in the UK, to estimate the distance-to-LE speed relationship.[7] The major fixed-line broadband suppliers (ISPs) provide data on individual speed tests to Ofcom,

[7] These data are available from the Ofcom Infrastructure reports – now called Connected Nations – accessible via the Ofcom webpage and the National Archives webpages.

Table 1: Summary Statistics.

	All Sample	Baseline Sample within 300 Meters
	(1)	(2)
A. Outcome Variables		
Average Percentile Rank Score (Mean)	50.21	49.76
	(28.67)	(28.66)
Average Percentile Rank Score in English	50.26	50.21
	(28.59)	(28.51)
Average Percentile Rank Score in Math	51.60	51.09
	(28.03)	(28.05)
Average Percentile Rank Score in Science	51.50	50.84
	(27.97)	(27.98)
B. Discontinuity Variables		
Distance to the Segment (Meters)	679.13	156.29
	(547.1)	(79.1)
Distance to the LE Station (Meters)	1,511.96	1,866.97
	(862.7)	(878.3)
Share on the "Fast" Side	0.54	0.51
	(0.50)	(0.50)
Average "jump" (Meters)	763.36	930.29
	(659.7)	(598.5)
C. Pupils & School Characteristics		
Distance to School (Meters)	2,602.5	2,413.35
	(3,433.14)	(3,008.3)
White	0.838	0.785
	(0.37)	(0.41)
Male	0.499	0.498
	(0.50)	(0.50)
Free School Meal	0.141	0.155
	(0.35)	(0.36)
Pre-KS3 Score	44.45	43.98
	(24.61)	(24.65)
Number of Schools	2,864	2,610
D. Area Socioeconomic Characteristics		
Share of White Pupils	0.825	0.769
	(0.29)	(0.32)
Share of Free School Meal Pupils	0.151	0.165
	(0.24)	(0.24)
Share of Community Schools	0.638	0.620
	(0.48)	(0.48)
Average House Prices (Pounds)	193,092.6	190,596.8
	(118,208.4)	(110,831.9)
Observations	1,115,594	183,892

Notes: This table shows descriptive statistics for the outcome variables (Panel A), treatment variables (Panel B), pupils and school characteristics (Panel C), and density and area socioeconomics (Panel D). The first column reports statistics for the whole sample of pupils and postcodes. The second column shows statistics for our baseline sample, which are pupils and postcodes located within 300 meters of the invisible LE station boundary segment. Standard deviations are reported in parentheses.

which aggregates the information by area in different years. The data from these suppliers cover over 80% of the market.[8] Data at the finest geographical level, the postcode, have been available since 2012 and are published yearly. Some quality measures are put in place, and only postcodes with a sufficient number of tests have usable data points. These are large datasets with close to one million postcode-level observations per year.

Postcode-level measures (average and median speed) are calculated from information on millions and millions of active broadband connections provided to the regulator and are based on *modem sync speed*, which captures the highest possible speed at which data can be transferred across the line with the use of a particular DSL technology (OfCom, 2012). The indicator captures the speed at which the modem in a customer's home connects to the equipment in the telephone exchange, and it is directly related to the subscription package headline speed. This way of measuring the line speed contrasts with speed tests obtained using modems at home and performed by users, who usually report slower speeds, which are affected by the time of the day at which the data transfer is done, the number of devices connected simultaneously and the quality of home software and internet equipment. This second type of measure is influenced by household socioeconomic variables, which are correlated with our outcome of interest and, in the context of our study, are less preferable than the measure based on line subscriptions. However, local modem sync speeds are still a reflection of resident demand for different broadband packages, which in the raw connection data used by Ofcom include a mix of technologies: primarily ADSL but also cable or fiber internet. In the results in Section 5, we include the same large set of local varying and time-invariant characteristics as that used for the main results; we expect these to comprehensively control for local characteristics correlated with speed demand.

A second concern is the deployment of non-distance-affected technologies in Britain from 2008 (cable) and 2010 (fiber). If reliable, rich data on local speed were available for our estimation sample period (2005–2008), we could use them directly in the estimations. However, small geography data were only made available from 2012, when superfast technologies were already available in some areas.[9] First, we note that even if in 2012 68% of England already had access to superfast broadband (yielding speeds over 30 Mbit/s), in this period, approximately 75% of the subscribed broadband connections used ADSL technology. In this sense, local averages for 2012 are the result for a majority of ADSL connections and are sensitive to distance to the LE. Furthermore, we use information on the potential available speeds that are realistic for the period 2005–2008, when most packages offered 8–10 Mbit/s headline speeds (OfCom, 2009a). With this in mind, when using data from after 2012, we restrict our sample to postcodes with average (download) speeds that

[8]The suppliers include BT, Virgin Media, Everything Everywhere, O2, KCom, TalkTalk and Sky.

[9]In previous years, the data were aggregated at a higher geographic scale (local districts), or studies on average speeds by region were based on smaller samples of tests.

are realistic for our sample period, e.g., up to 10 Mbit/s.[10]

2.3 Summary Statistics

Table 1 provides the descriptive statistics of key variables in the whole and our baseline estimation sample, with mean values and standard deviations. In the full sample, our data cover slightly more than 1.1 million students living in over 400,000 postcodes and attending more than 2,860 schools in England over the period 2005–2008.[11] Our estimation sample is constructed by focusing on households within 300 meters of an LE catchment area boundary segment. This procedure is explained in detail in the coming section.

Panel A provides the descriptive statistics of our outcome variables for the subject-specific tests and the mean of the three. Panel C reports the pupil-level characteristics and shows that the vast majority of the pupils are white, approximately 14% are entitled to free school meals, and students live on average 2.5 kilometers from their schools. Panel D displays the postcode-level characteristics, which show similar values in the proportion of white and FSME students at the local level, a majority of community-type schools and an average house price of approximately £190,000. The table shows that the composition of pupils and area characteristics for the whole sample are highly similar to those of the estimation sample.

3 Empirical Strategy

3.1 Sorting Issues and Identifying Variation

To estimate the effects of home broadband on education outcomes, a major identification challenge has to be addressed: household-level observed broadband speed – i.e., package choices – is likely related to learning outcomes through confounding factors that are difficult to directly control for. We refer to this as *active* sorting; e.g., better-off households invest in better connections to boost outcomes. As a result of this type of sorting, using data on observed broadband speeds is problematic. The approach used by the existing literature is therefore to focus on variation in *available* broadband speeds, which depends on location choices but not on broadband subscription choices.

In this context, we exploit a feature of the DSL-broadband technology: a salient feature of the technology is that once a home is connected to a broadband-enabled LE station, the *available* con-

[10] The rollout of ADSL2+, which allows speeds of up to 24 Mbit/s, did not start until 2008.

[11] The raw data include approximately 500,000 pupil observations per year. To prepare the sample for our empirical strategy, we exclude observations that (i) have implausible or inconsistent values, (ii) are assigned to segments that only have observations on one side, thus making it impossible to perform within-segment-year comparisons, (iii) cannot be linked to school or local area characteristics, or (iv) are located in postcodes that had broadband services enabled for less than six months.

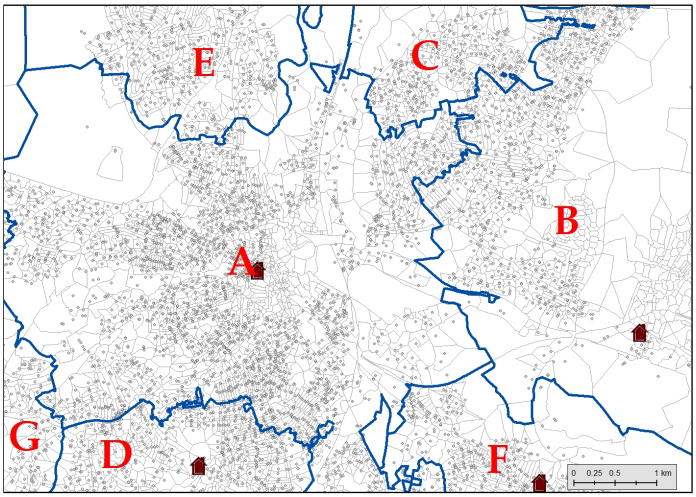
nection speed depends on the length of the copper wire connection between the residence and the LE station.[12] Falck et al. (2014) are the first to exploit this feature, estimating the effects of information disseminated over the internet on voting behavior in Germany, where during their sample period entire (small) locations happened to be located too far away from an exchange station to access broadband. In particular, towns farther than five kilometers from an exchange could not obtain any broadband internet without costly further technological upgrades. This characteristic allows the authors to exploit differences in outcomes between places that were connected to the DSL network and those that were not. However, this approach is not directly applicable to the British context because of a much denser network of LE stations, related to the smaller size of the country, with all places connected to the network, and the relatively quick rollout of the broadband infrastructure.

We use a modification of Falck et al. (2014)'s approach by exploiting "jumps" in distance to the LE across catchment area boundaries.

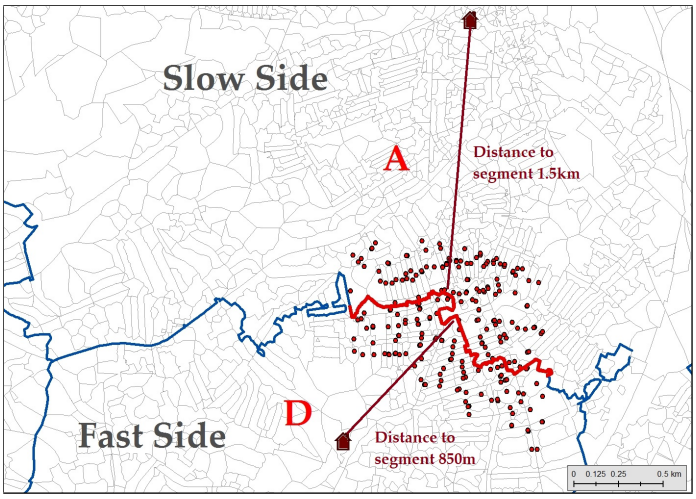
In the UK, the location of LE stations was determined during the deployment of the English landline telephony network, which mainly occurred before and during World War II. Importantly, distance to the LE station did not affect the quality of traditional telephone services, so the setup of the network was designed to maximize the number of connections from the minimum number of stations (with the goal of achieving cost-effectiveness). However, while we believe that it is unlikely that households in the past or present actively sort on the basis of distance to the connected exchange station, there are several reasons to believe that LE station location is far from random and is potentially correlated with other local neighborhood characteristics that do matter for household sorting. One can argue that most households are not likely to be aware that the quality of their broadband connection is related to their location choices, and even if they are, they might not precisely know where LE stations are located. Households might not be located at different distances from the LE because they know and care about the speed-distance relationship (abstracting from *active* sorting). However, they might sort with respect to other geographical features also correlated with station location. For example, LE stations appear to have been placed at central locations (local town centers) that were also close to major road junctions for hosting the exchange switchboard infrastructure. We refer to this as *passive* sorting. As local geography correlates with both the location of LE stations and the location of households across space, comparing households located within an LE would lead to biased estimates.

[12]Distance to the LE is not the only driver of the variation across DSL subscriptions, as other factors can also affect observed speed, such as the quality of the hardware and software used, the number of simultaneous users in the household, the day of the week and time of day, the size and upkeep of the LE station and other technical factors such as varying quality of in-house wiring, unconnected microfilters, or varying performance by the ISP. See OfCom (2009a) and OfCom (2009b) for more details. See Ahlfeldt et al. (2017) for estimates of the determinants of household observed broadband speed.

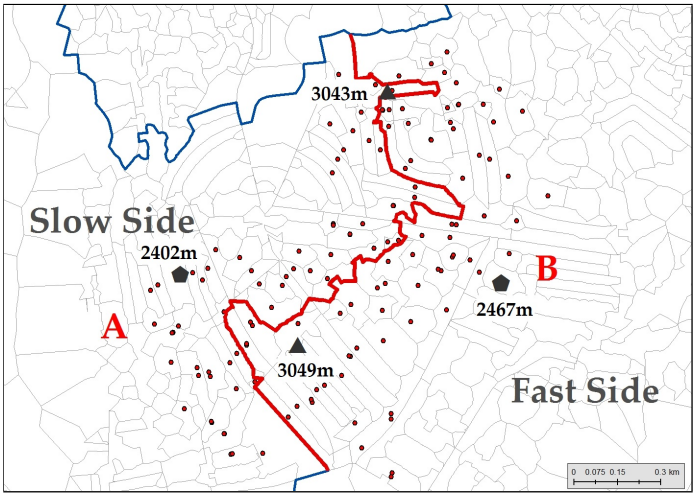
Figure 1: Graphical Illustration of Empirical Strategy.



(a) Overview of LE catchment areas and boundary segments



(b) Postcode sample boundary A–D segment – Sharp



(c) Postcode sample boundary A–B segment – Fuzzy

Notes: The light gray dots represent the precise location of the postcode centroids for a selection of LE catchment areas at the 1 meter-resolution precision. The blue lines represent the invisible LE station boundaries. The building symbols display the exact locations of LE stations. Subfigure (a) shows an overview of LE station areas with underlying postcode area polygons and centroids. One letter is allocated to each LE. Note the irregular shapes and the fact that LE stations are not always located in the center of the LE area. Subfigures (b) and (c) zoom in on the boundary segments of two particular LE catchment areas, AD (b) and AB (c). The red dots mark the postcodes located within 300 meters of the red boundary segments.

To overcome identification issues related to *passive* sorting, instead of comparing different locations within LEs (e.g., households connected to the same LE but at different distances from the station), we use a strategy that compares variation in DSL-cable length across neighboring locations. We compare households located very close to each other and thus with similar geographical features but with connections to different LEs and hence with different broadband *availabilities*. These boundaries give rise to substantial cross-sectional variation in the quality of the available DSL-broadband speed due to discontinuous jumps in the length of the copper wire that connects residences on either side of the invisible boundary to their assigned LE stations. The different shapes and sizes of the LE catchment areas give rise to discontinuous changes in average distances to the LE on each side of small boundary segments (which we call "jumps").

Note that in this paper, we do not compare outcomes between DSL-connected and unconnected places but exploit differences in broadband quality across locations.[13] The period for which we could compare locations with internet dialup connections to those already connected to a broadband-enabled LE is that encompassing the rollout of the DSL infrastructure, e.g., mostly 2000–2004. The technological upgrade between these two technologies would have allowed us to exploit a 10x increase in expected internet speeds; however, the takeup rate of broadband services before 2005 was negligible, reducing the probabilities of finding an identifiable treatment. Instead, we use data for the years 2005–2008, when broadband infrastructure was almost universally enabled and when takeup rates were already substantial and growing. We argue and provide evidence below that using the variation in available broadband speed across local boundaries addresses the discussed endogeneity concerns. This variation has to be exploited at very small scales to avoid spatial confounders correlated with location and outcomes. To leverage the richness of the data, it is essential to use very disaggregated information, in terms of both geographical scale and sample size; thus, the available geolocated administrative data are key for the application of a robust empirical strategy. Next, we explain how we take this setting to our data and our estimation approach.

3.2 Construction of the Discontinuity and Treatment Variables

The core of our empirical strategy is the construction of the boundaries of LE catchment areas. The first thing to note is that these boundaries do not coincide with any other administrative boundaries, in particular school district boundaries, and are in practice difficult for households to know.

To set up the empirical strategy, we use information on the precise geolocation of the universe

[13] Other technological and regulatory changes also took place during our period of analysis; however, by comparing within-year cross-boundary segments, we focus solely on cross-sectional variation to obtain our estimates.

of English LE stations (approximately 3,900). In particular, we use the assignment of each of the English postcodes to the LE that provides telephone and internet services to its premises. There are approximately 1.45 million full postcodes in England. Each postcode contains approximately 15 households on average, and the postcode areas are often as small as a single building, especially in denser areas. This georeferenced dataset allows us to construct precise LE station-level catchment areas, which are usually unobserved, and from them infer the exact boundaries between different LE areas. We construct the catchment areas by aggregating the polygons of all the postcodes connected to the same LE station.[14] It is important to use the correct LE-postcode pairing, as due to constrained capacities and natural accidents, not all the postcodes are served by the closest station. Figure A.1 depicts all LE catchment areas in England. We construct detailed polygons for each catchment area, which are then transformed to create boundaries (lines) identified by a pair of LEs, one on each side. Then, the boundaries are divided into smaller segments (henceforth called boundary segments), which are on average 3.2 kilometers long (S.D. of 1 kilometer). The details of the underlying data can be appreciated in Figure 1, where we can observe how some postcodes correspond to portions of streets. Next, we assign all postcodes in England to particular boundary segments based on their proximity, conditional on which LE they are connected to.[15] This determines which side of the boundary the postcodes belong to. We exclude boundaries on the outline of the country to ensure that we can pair coastal postcodes with segments with neighbors on the other side.

The following step is the construction of the treatment and SRD variables. For each postcode in England, we calculate the euclidean distance to the connected LE station; this approximates the connecting copper cable length, which is a measure of internet quality (speed).[16] Our goal is to compare households that live close to each other but on different sides of the invisible LE catchment area boundary segment. We therefore also calculate the distance between postcode centroids and the closest boundary segment. This distinction is important and worth reiterating: there are two different types of distances: The first is the distance to the connected LE station, which is an important determinant of the available broadband speed. This distance increases as we approach the boundary segment and changes discontinuously when we cross an LE boundary segment. This is the distance measure that gives rise to the variation in broadband quality across boundary segments. The second measure is the distance to the LE boundary segment. This distance is used

[14] Instead of approximating postcodes with centroids, we use Ordnance Survey CodePoint with Polygons data, which provide very detailed polygons for each postcode in the UK.

[15] We know the precise geolocation of the postcode centroids using the British National Grid Eastings and Northings to the 1 meter precision from the National Statistics Postcode Directory.

[16] A shortcoming of our approach (common to other, similar papers) is that we can only calculate crow-fly distances between the centroids of the postcodes of the location of pupils' homes and the LE station to which they are connected; in urban areas, there can be a substantial gap between this and the actual length of the connection cable (OfCom, 2009a). Nevertheless, given the very small scale of our geographical units of observation, we can approximate this in a more precise way than other studies that use data for larger geographical units.

to identify close neighbors, i.e., to select which locations we use as comparisons within a short segment, and it is our SRD running variable. Panel B of Table 1 provides summary statistics on these two distances for both the full and estimation samples.

To make this geographical setting operational, one important step is necessary: to define who is on the *fast* and who is on the *slow* sides of each boundary segment. Initially, there are over 40,000 boundary segments (some of them very small) and over 1.45 million (active) postcodes in England. Figure 1(a) provides an illustration of the geographical details of the data. As we show below, households located inside of the LE are different from households located at the edge of it, but they are similar to households on the other side of the boundary. We therefore first restrict the postcode sample that we use to postcodes located close to the nearest boundary segment. For our main analysis, we use the sample of postcodes within 1 kilometer of the boundaries. For this sample, we then construct the segment-specific variables to implement the (fuzzy) SRD (approximately 65% of the total sample). Using all the postcodes assigned to a particular boundary-segment side, we calculate the average distance to the connected LE of the postcodes on that boundary-segment side. The side that has shorter average distances is defined as the *fast* side, and the other side is defined as the *slow* side.[17] The difference in the average distances between the two sides measures the jump in average cable length when we cross the boundary segment; from this jump, we identify the impact of quality broadband. Sometimes the jump between the two sides is small, so the variation in average distances is relatively low. For this reason, we exclude segments in which the jump is below 100 meters, and for our main results, we focus on segments with jumps of at least 300 meters. We extensively discuss the robustness of this choice in Section 4.4.

Finally, we match the postcode-segment-side and distance (to LE and to segment) data to the KS3 pupil information based on the home address postcode. For each boundary segment and year, we observe the universe of 14-year-old pupils who live in households located at different distances from the segment and at different distances from the LE station. In essence, we group all pupils (postcodes) closest to the same boundary segment into a local-segment neighborhood. The invisible boundary cuts through each of these neighborhoods, splitting them into the *fast* and *slow* sides. Within each neighborhood, the invisible boundary line thus produces variation in distance to the connected LE station. To compare households with similar geographical surroundings, we use postcodes within 300 meters of the boundary segment in our preferred estimation sample. Given the large size of the underlying dataset, even this narrow definition still provides a sample size over 180,000 pupil observations (living in over 60,000 postcodes).[18] As becomes apparent from the nonparametric estimation results of the boundary discontinuity effect, none of the presented

[17] Using a 500 meter sample around the segments to construct these variables provides very similar results.

[18] For the 1 kilometer sample, this corresponds to 580,000 observations in almost 300,000 postcodes.

findings are sensitive to increasing or decreasing this sample threshold.[19] Robustness checks on the sensitivity of our findings to the baseline sample selection are discussed extensively in Section 4.4.

Figure 1 provides a graphical illustration of our empirical setup and our strategy. Figure 1(a) shows the population of postcodes (gray dots) in several telephone exchange areas (e.g., A, B or C), where the location of the LE station is indicated with a building symbol. All the postcodes inside a given area are connected to the LE station, and telephone and broadband service are provided from the copper cable connecting the LE station and the premises. The boundaries of the catchment areas of each LE are shown in blue. Each boundary has one LE station on each side, allowing the identification of boundaries from a combination of two LE stations (e.g., AD, AB, EC or CB). Boundaries are split in smaller segments. In Figures 1(b) and 1(c), specific boundary segments are shown as thicker red lines. We expect – and show in Section 5 – that households in postcodes closer to the station have faster internet connection speeds.

Because of the topology and different sizes of the LE catchment areas, there might be higher or lower differences in the average distance to the LE station between both sides of the segment, which gives rise to a higher or lower "treatment" change when we compare pupils across the segments. This is illustrated in Figure 1(b): side A of the segment is on average 1.5 kilometer from the LE station, while side D is 850 meters from the station.[20] Postcodes located within 300 meters of the two highlighted boundary segments are marked in red. In this particular case, it is clear that all households within 300 meters of segment AD on the *fast* side have shorter individual distances to the LE than all households on the slower side, as the difference in the average distances is 650 meters. In this case, when we compare pupils from both sides of the segment in a given year, the SRD is sharp and the treatment (distance to the LE station) changes discretely for all households when we cross from the *slow* to the *fast* sides.

However, it could be the case that two addresses located on different sides of the segment are not necessarily different in the way that we would expect. Due to the irregular geographic shape of several invisible boundaries, some households with short cables (long cables) might live on the slower side (faster side). Hence, our SRD design is fuzzy, and sharp RD estimates would suffer from attenuation bias.[21] This is illustrated in Figure 1(c). For segment AB, the average distance to the connected LE is quite similar on both sides, approximately 1.6–1.7 kilometers. The shape of the

[19] For completeness, we also report estimation results for a wider distance band around the boundaries covering more than 97% of the student population in England.

[20] To calculate the discontinuity variables, we use the population of postcodes within 1 kilometer of the boundaries. The aim is to obtain more representative boundary-segment variables that are independent of the choice of pupil sample. We drop extreme outliers (located further than 3.25 kilometers from a segment or 5.5 kilometers from an LE station (less than 10,000 postcodes)). After excluding postcodes assigned to the segments with observations on one side only, we are left with 824,000 postcodes in England and 17,000 operational segments.

[21] For the sharp SRD estimates, see Table A.2.

segment is irregular and slightly diagonal, tilted to the west. Two sets of postcodes are selected to explain this situation. The triangular-shaped ones are both around 3 kilometers away from the LE station, but the shorter of the two (3,043 meters) is located on the *slow* side, and the longer segment in the pair (3,049 meters) on the *fast* side.[22] If the jump in distances between sides is attenuated because some households are "assigned to the wrong side", we would not have enough variation to estimate the coefficients with precision. We resolve this situation by using an IV strategy, which we explain in detail in the next subsection.

3.3 Specification and IV Strategy

Our goal is to estimate the causal effect of broadband internet speed on test scores for 14-year-old students. The basic framework of our analysis capturing the relationship of interest is the following:

$$TestScore_{ipnlst} = \beta DistLE_{pl} + g(D_{pn}) + X_i'\Lambda + Z_{is}'\Theta + A_{pt}'\Phi + L_p'\Psi + \delta_{nt} + \epsilon_{ipnlst} \tag{1}$$

where $TestScore_{ipnlst}$ is the percentile rank in the KS3 test of pupil i living in postcode p in boundary-segment neighborhood n associated with LE station l and attending school s at time t; $DistLE_{pl}$ is the distance to connected LE station l from postcode p; X_i' is a vector of student background characteristics, such as preinternet student performance on the KS1 test, gender and free school-meal eligibility status; Z_{is}' is a vector of characteristics of the school attended by pupil i, such as school type and distance between home and school; A_{pt}' is a vector of postcode-year-specific characteristics, such as local average housing prices, share of students eligible for free school meals, and white (population) pupils; L_p' is a vector of time-invariant postcode attributes such as density (e.g., number of delivery points) and distance to different amenities (e.g., nearest rail station or road); δ_{nt} is boundary segment-by-year fixed effects, which guarantee that we are comparing students within the same segment-year of the LE boundary; and ε_{ipnlst} is the error term. To ease interpretation of the estimates, we measure $DistLE_{pl}$ as "negative distance", e.g., proximity to the LE station. Thus, *beta* captures the changes in test scores when we come closer to the LE by one meter and broadband quality improves. We cluster the standard errors at the segment-by-year level.[23]

To apply the empirical model to the data, we need to specify two additional pieces of information. First, the definition of $g(D_{pn})$ captures the relationship of the postcode distance to the boundary segment. This deterministic function has the spirit of the running variable in a nonspa-

[22] The pentagon-shaped postcodes display the inverse situation: the one with longer distance, 2,467 meters, is located on the *fast* side, while that with the shorter distance (2,402 meters) is on the side with longer distances on average.

[23] Table A.5 shows that different clustering choices for our preferred specification do not change our conclusions.

tial regression discontinuity design. In our setting, we are interested in controlling for distance to the boundary segment because of *passive* sorting, as explained in Section 3.1 above. As we have a large sample of pupils, even when we restrict observations to postcodes within 300 meters of the border, we can carefully control for distance to the boundary segment, where each side is connected to a different LE station. This means that we can effectively compare the test scores of students with differences in home broadband access living very close to each other.

Our preferred control function for the relation with the running variable is the following:

$$\begin{aligned} g(D_{pn}) &= \sum_{b=1}^{B_s} \left(\gamma_b^{slow} \, Slow_{pn} * I\left(DistSegment_{pn} = b\right) * DistSegment \right) \\ &+ \sum_{b=1}^{B_f} \left(\gamma_b^{fast} \, Fast_{pn} * I\left(DistSegment_{pn} = b\right) * DistSegment_{pn} \right) \end{aligned} \tag{2}$$

where we control for the distance to the LE invisible boundary segment by using distance bin dummies of 100 meters (b) interacted with the distance to the boundary segment on both sides of the cutoff. In terms of the notation, $Slow_{pn} * I\left(DistSegment_{pn} = b\right)$ is an indicator variable for postcode p on the *slow* side of boundary segment n that equals one if the distance of postcode p to segment n is within bin b. $Fast_{pn} * I\left(DistSegment_p = b\right)$ is defined analogously for all postcodes on the *fast* side of the boundary segment n. This semiparametric approach allows for more flexibility in controlling for distance to the segment. Instead of imposing a certain functional form on the polynomial, we estimate the coefficients γ_b^{slow} and γ_b^{fast} for each small distance bin for the segments (i.e., 0–100, 100–200, and 200–300), thereby capturing the shape of the polynomial in a flexible way on each side of the boundary segment. This approach is flexible and avoids either oversimplifying the underlying relationship (as would be the case if, for example, we used a linear polynomial) or overfitting by using high-order polynomials (Gelman & Imbens, 2014). The results are robust to alternative definitions of distance bins, to the use of distance as a continuous variable, and for a more flexible approach of higher-order polynomials.[24]

One final step is required for estimation. Using the information on which side of each boundary segment has a lower average distance to the LE, we construct segment-side-specific *Fast* dummies that we use as instruments for the actual postcode-level distance to the LE. The first-stage equation is:

$$DistLE_{pl} = \pi Fast_{pn} + g(D_{pn}) + X_i'\Lambda + Z_{is}'\Theta + A_{pt}'\Phi + L_p'\Psi + \delta_{nt} + \varepsilon_{ipnlst} \tag{3}$$

where $Fast_{pn}$ is a dummy variable equal to one if student i living in postcode p is located on the *fast* side of invisible LE station boundary segment n and zero otherwise. As above, $DistLE_{pl}$ is

[24]These results are available upon request.

measured as proximity to the connected LE station. The coefficient π captures how much closer postcodes are to their connected LE stations on the *fast* side relative to the average distance on the *slow* side.

Since the invisible telephone LE station boundaries are historically given and under the assumption that households do not sort on each side of the boundary within the boundary segment, we can focus on households whose residences are located in the vicinity of the invisible LE boundary segment, considering those on the *fast* side very similar to those on the *Slow* side. The broadband speed assigned to these households can be considered "locally" randomly assigned within segment-by-year. Assuming that in the absence of the treatment, the outcome variable is a smooth function of distance to the LE boundary, the causal effect of broadband internet speed is identified by comparing outcomes for pupils who live close on the *fast* side of an LE station boundary (treatment group) with those who are near but live on the *slow* side (control group). This effect is captured by the IV estimate of β in specification 1. This strategy estimates a local average treatment effect (LATE) of quality to high-speed internet on student performance by comparing "lucky" households that are supplied with faster broadband access to otherwise similar counterparts that were "unlucky" in terms of being supplied with slower broadband access. A specific feature of this strategy is that it generates variation over multiple thousand telephone LE station boundary segments, which vary in distance to the LE station and in the jump in distance across the invisible boundaries.

3.4 Estimating the Impact of Broadband Speed on Test Scores

In a simple framework, faster broadband internet could affect teenagers' school performance in different ways. High internet speed allows students to access more online content per unit of time. If test scores are determined in a learning production function, we can think of speed impacting learning productivity. For each hour of study, students can access more information, shifting the learning production function upwards if, for example, they can access more learning resources such as Wikipedia or online interactive materials. However, broadband could affect learning investment by reducing study hours if students divert time to nonlearning online activities, such as gaming or using social media. This second channel could be more or less relevant depending on whether online distractions replace offline distractions. The first channel would have a positive effect on human capital formation, while the second would have a negative effect. Coefficient β captures the net impact of these channels, and our aim is to identify an unbiased estimate of this coefficient.[25]

[25]This model can become more complex if, for example, we take into account the changing nature of web content over time or the interaction of school and home ICT use, but discussing this is beyond the scope of this paper.

Henceforth, the causal estimation of coefficient β in equation 1 informs us about the impact of *available* broadband speed on test scores. It captures how much test scores change as we approach an LE station via the relationship between copper cable length and potential speeds. We estimate a LATE impact by using a proxy of broadband speed that changes over space. While the estimation of this parameter is desirable from an identification point of view, it also provides a measure of the intention-to-treat (ITT) impact, which is a relevant policy variable. By investing in network improvements and expansions, public policy can influence the availability of fast broadband in different locations, while afterwards, particular households might sort into different broadband packages for reasons that might correlate with the outcomes of interest.

However, we are also interested in the impact that broadband speed changes have on student performance. For the reasons explained above and due to the lack of data appropriate to the time period and geographical scale of our analysis, we cannot directly estimate this impact in the data. Our approach is to employ out-of-sample-period data to reverse engineer the distance-speed relation present during our study period, combining linked postcode-to-exchange station telecom network data with data on the local average internet speed experienced by households. We compute the Wald estimate dividing our LATE coefficient (effect of distance to LE station on test scores) by this parameter (distance to LE station on average speed) to obtain a measure of the impact of changes in speed on test scores and calculate the standard error using the Delta method.

A last step is to scale our Wald estimate by the relevant population of interest to approximate the treatment-on-treated (TOT) effect. We do this using additional survey data to obtain information on the appropriate broadband takeup rates during the period of analysis (those of families with teenagers during 2005–2008 in England) and on the proportion of pupils using the internet for school work. This allows us to discuss the economic relevance of our results and to contextualize the size of our estimates. We extensively discuss our quantification estimates in Section 5.

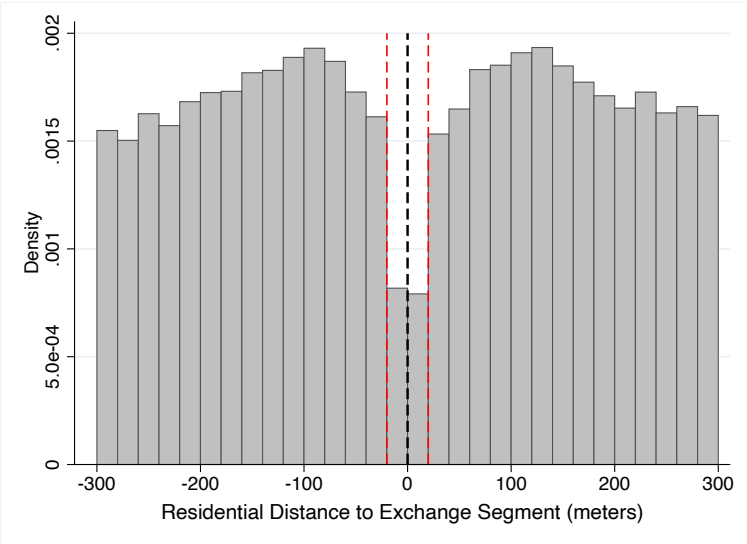
4 Results

4.1 Internal Validity of the Empirical Strategy

The internal validity of the fuzzy SRD requires that there is no endogenous sorting on either side of the LE boundary segment. In the case of this study, there are two key features that make manipulation at the LE station boundary segment highly unlikely. First, LE catchment areas are either invisible or unlikely to be known by households. Second, LE boundaries do not coincide

with administrative boundaries of any kind, such as school district boundaries.[26]

Figure 2: Distribution of Pupil Distance from Residential Distance to Exchange Segment.



Notes: The figure shows the density of pupils in our preferred estimation sample at the invisible LE boundary segment. The black dashed vertical line presents the invisible LE station boundary segment. The red dashed vertical lines show the boundaries of the donut strategy. Each bar contains bins of 20 meters' distance.

Figure 2 presents a histogram representation of pupil density at the LE station boundary segment for our preferred estimation sample. It allows for visual inspection of whether bunching takes place. It reveals no evidence of systematic manipulation of residential distance to the LE boundary segment around the thresholds. However, we observe that density drops within 20 meters of the LE boundary segment (indicated by the red vertical dashed lines). This drop is a byproduct of the data construction: postcode distances to the boundary segment are measured from the centroids, and there are very few oddly shaped postcodes where the centroid falls within 20 meters of the boundary segment. The fact that the drop is of the same magnitude on both sides of the discontinuity is a positive indicator that it is unlikely that this feature has a differential effect on one particular side of the LE boundary segment. We also perform more in-depth analyses and formal tests of bunching, as reported in Section A.2. We formally test for bunching following McCrary (2008) and Cattaneo et al. (2018). The McCrary (2008) test fails to reject the null of no significant jump at the LE boundaries.[27] Cattaneo et al. (2018) propose a test that is robust to bandwidth selection issues. We fail to reject the null hypothesis of no discontinuous jump at the LE boundary segment.

Since the density drop within 20 meters is a byproduct of the geographical resolution of the data, we use the so-called donut strategy, which excludes observations within 20 meters of the

[26] In the data construction, we removed boundary segments that intersect with natural boundaries (rivers). While it is now possible to determine (rough) boundary locations online, this was not possible during the period studied.

[27] We also test for various alternative residential distances to the LE boundaries. The results are robust to different specifications. See Table A.1.

discontinuity in our main specifications, following Angrist et al. (2019) and Leuven & Løkken (2020), among others. Discarding these observations improves the precision of our estimates since we eliminate potentially rare postcodes and spillovers across LE boundaries.[28]

Another direct check of a violation of instrument exogeneity is to test whether pupil, school and area/postcode baseline characteristics are "locally" balanced on either side of the LE boundary segment. If these variables are unbalanced on either side of the boundary segment, it would indicate selection problems around the discontinuities. Table 2 tests the balance of a battery of background characteristics. The first column of Table 2 regresses the main outcome on the predetermined background characteristics. The regressions further control for distance bins and include segment-by-year fixed effects. The results of this column show that the background characteristics are economically and statistically important in explaining the variation in the outcome variable. We strongly reject the null hypothesis that these variables are jointly equal to zero. The third column of Table 2 displays the results of local linear regressions for each of the predetermined pupil, school, density, area socioeconomic, housing price and amenity characteristics. We show that most of the regressors are not significantly different from zero, and the coefficients are very small. In particular, important determinants of our outcome variable, such as individual preinternet scores, free school meal eligibility and housing prices, are statistically indistinguishable from zero. However, we find that postcodes on the *fast* side of the LE boundary segment are more likely to be closer to a school and a rail station and less likely to be white. However, the magnitude of the differences in these characteristics with those on the *slow* side of the LE boundary segment is very small (0.6%, 0.3% and 1.8% with respect to the baseline mean, respectively). Given the number of variables tested, it is not surprising to find some small imbalances in individual variables. In line with this, we fail to reject the joint test in which all the coefficients are equal to zero at conventional levels of statistical significance. This means that we do not find evidence that third factors, some of which are extremely important in explaining later test scores, change discontinuously across our invisible boundary segment.

[28] At the boundary segment, very local spillovers might exist where households on the *slow* side could connect to the WiFi routers of neighbors on the *fast* side. The fraction of pupils who are discarded represents less than 1% of our estimation sample.

Table 2: Balance of Baseline Student, School and Area/Postcode Characteristics.

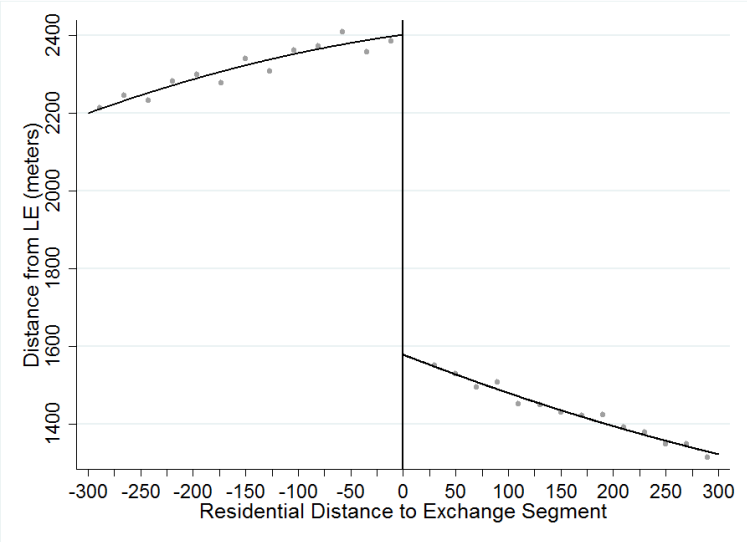
	Outcome	Instrument	
	Average Percentile Rank Score	*Slow* Side Baseline Mean	*Fast* Side Nonparametric Estimate
	(1)	(2)	(3)
A. Student Characteristics			
White	-0.592**	0.78	-0.014*
	(0.244)		(0.007)
Male	-2.430***	0.5	0.005
	(0.133)		(0.010)
Free School Meal	-13.749***	0.16	0.003
	(0.206)		(0.008)
Pre-KS3 Score	0.819***	43.89	-0.153
	(0.002)		(0.508)
B. School Characteristics			
Log Distance Home to School	3.699***	7.39	0.012
	(0.112)		(0.017)
Log School Size	1.305***	6.88	0.007
	(0.140)		(0.013)
Community School	-9.951***	0.62	0.003
	(0.214)		(0.009)
C. Density & Area Socioeconomics			
Log Number of Delivery Points	-0.944***	3.34	0.013
	(0.132)		(0.015)
Log Number of Pupils per Premise	-1.754***	-1.76	0.016
	(0.121)		(0.016)
Share of White Pupils	1.416***	0.77	-0.008
	(0.389)		(0.005)
Share of Free School Meal Pupils	-22.724***	0.17	-0.001
	(0.354)		(0.006)
D. Housing Prices			
Log Average Housing Price	13.310***	12.04	0.014
	(0.270)		(0.009)
E. Amenities			
Log Distance to Closest Road	0.045	5.44	-0.013
	(0.103)		(0.022)
Log Distance to Closest School	1.416***	5.83	0.033**
	(0.134)		(0.016)
Log Distance to Closest Supermarket	1.156***	6.56	0.013
	(0.191)		(0.012)
Log Distance to Closest Rail Station	0.118	7.22	0.020**
	(0.259)		(0.009)
Log Distance to Closest Water Body	-0.144	6.5	-0.002
	(0.136)		(0.017)
Joint F-test	10,381		1.23
Joint P-value	0		0.23
No Segm-years	7,096		7,096
No Observations	183,892		183,892

Notes: Column (1) regresses the main outcome on the predetermined background characteristics. Column (2) shows the baseline mean on the *slow* side for the predetermined background characteristics. Column (3) performs local linear regressions for each of the background characteristics. Each regression controls for distance bins and segment-by-year fixed effects. F-tests (and corresponding p-values) are for the joint significance of the variables reported in each column. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.2 Discontinuities in Distance to the LE Station

We now investigate the discontinuities across the boundary segment in the distances to the connected LE stations. Figure 3 plots the distance to the connected LE station as a function of household distance to the LE boundary segment. The left side shows the average "stacked" distances to connected exchange stations of all boundary segments on the *slow* side. As we move closer to the boundary segment, the distance to the exchange increases. Then, there is a clear discontinuity. On the *fast* side, the average distance to the (different) connected exchange in turn decreases as we move farther away from the boundary segment.

Figure 3: Distance from the Connected LE Station Jumps across the Boundary Segment.



Notes: The dots represent the average distance to the LE station per 25-meter interval of residential distance to the invisible LE boundary-segment boundaries. The solid lines are fitted values from a third-order polynomial approximation, which is estimated separately on both sides of the cutoffs. "Residential Distance to Exchange Segment (Meters)" refers to the residential distance to the invisible LE boundary-segment boundaries. The black vertical line is the stacked invisible LE station boundaries.

In Panel A of Table 3, we estimate the magnitude of this jump, which is the first stage in our fuzzy SRD setup. The results confirm that being on the *fast* side of the invisible threshold is a strong and statistically highly significant predictor of the distance to the connected LE station. Our preferred specification coefficient, shown in the fifth column of Table 3, estimates that households just on the *fast* side are 816.6 meters closer to their LE stations (relative to those on the *slow* side). This coefficient and the corresponding estimates from the various other specifications, which are discussed in more detail below, are statistically significant at the 1% level.

4.3 Impact on Student Test Scores

Panel B of Table 3 shows the fuzzy SRD nonparametric estimates on the average student percentile rank score. Each column presents estimates from increasingly saturated specifications. Column 1 includes student test scores from a prebroadband period in England (KS1 test scores from age 6/7)

as well as the distance bins and segment-by-year fixed effects. Adding this predetermined control variable effectively changes the empirical strategy to a fuzzy SRD *value-added* design. Using this specification, moving one meter closer to the LE station increases the national KS3 exam performance at age 14 by 0.00123 percentile ranks. The results are statistically significant at the 5% level. In column 2, we add further individual controls (e.g., ethnicity, free school meal eligibility, and gender), which on their own have explanatory power, as documented in Table 2. To increase precision, we include interactions between the different combinations of student individual controls. The main effect increases slightly further to 0.00131 per meter, which is statistically significant at the 1% level. Columns 3 and 4 introduce the additional time-variant and time-invariant area and amenity controls from Table 2. This hardly affects the estimates. Finally, the last column includes additional school-level controls. In column 5, we control for school type, size and distance from home. This does little to the estimates.

Because of the stability of the estimates across columns, we choose the estimate from column 5 as our preferred specification, for which we discuss more results below.[29] The baseline estimate is statistically significant at the 5% level (p-value=0.01). As explained above, due to the irregular geographic shape of several invisible boundaries, it is critical to use a fuzzy SRD design since sharp RD estimates would suffer from attenuation bias. Table A.2 illustrates the attenuation bias, showing its substantial magnitude. Although many of the estimates are statistically significant at conventional levels, the sharp SRD estimates are approximately five times lower than the preferred fuzzy SRD estimates.

At face value, this estimated effect of about 0.0012 percentile ranks per meter seems very small and in fact is only detectable due to the combination of our empirical approach with formidable student census data that includes national test scores (and previous test scores) at a very high geographical resolution. In Section 5, we provide direct estimates that allow us to quantify this positive effect in terms of broadband speed and usage to show that the effect size is economically meaningful.

Our headline estimate, an effect of 5% of a standard deviation, is not small in the context of the education literature. For instance, teachers have been identified as one of the most important factors in test performance, with a one standard deviation increase in teaching quality improving test scores by 0.1–0.2 standard deviations (Rivkin et al., 2005; Slater et al., 2012; Chetty et al., 2014a). Peer effects are smaller at between 0.01 and 0.08 standard deviations (e.g., in the English context Lavy et al., 2012; Gibbons & Telhaj, 2016), and various dimensions of neighborhood quality have even smaller effects on test scores (e.g., Jacob, 2004; Sanbonmatsu et al., 2006; Gibbons et al., 2013,

[29] The results are robust to the use of a triangular kernel. The point estimates are larger in magnitude (0.00147), but we cannot reject the null hypothesis that the estimates are different from the baseline estimates. These results are available upon request.

2017). Therefore, in the context of the education literature, our estimates represent medium-sized effects.

Table 3: Fuzzy SRD Estimates: Impact of Exchange Distance on Student Performance.

	(1)	(2)	(3)	(4)	(5)
A. Avg. jump in distance to connected exchange when crossing boundary to the *fast* side (in m)					
Nonparametric	-814.3***	-813.8***	-813.3***	-816.4*** *	-816.6***
Estimates	(16.00)	(15.99)	(15.99)	(15.93)	(15.93)
B. Effect of exchange distance on national KS3 exam scores at age 14 (in percentile ranks)					
Nonparametric	0.00123**	0.00131***	0.00120**	0.00120**	0.00122**
Estimates	(0.00049)	(0.00048)	(0.00048)	(0.00048)	(0.00047)
Student Preinternet Score KS1	Yes	Yes	Yes	Yes	Yes
Student Controls	No	Yes	Yes	Yes	Yes
Time-Variant Area & Amenities	No	No	Yes	Yes	Yes
Time-Invariant Area & Amenities	No	No	No	Yes	Yes
School Controls	No	No	No	No	Yes
Observations	183,892	183,892	183,892	183,892	183,892

Notes: The table shows the fuzzy SRD nonparametric estimates for the average distance to the LE station (Panel A) and test scores on the KS3 (Panel B). Columns report different specifications. Each coefficient comes from a separate regression, where the running variable is the residential distance to the LE station invisible boundary. The window size for the residential distance to the LE station invisible boundary is $\pm$ 300 meters. We only include boundaries where the distance jumps by at least 300 meters. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.4 Robustness Checks

4.4.1 Sensitivity to Sampling Choices

All results discussed so far are estimated using the sample of postcodes that fall within 300 meters of the invisible boundary segment and where the distance jump to the connected exchange across the boundary segment is at least 300 meters. To clarify that these choices do not affect our conclusions, in Table A.3, we estimate our preferred specification with different samples, using only observations falling within 100 meters of the boundary segment up to those falling within 1 kilometer on either side. The effect sizes are slightly larger in samples closer to the boundary segment (0.00157 in the 100 meter sample) but not significantly different from each other at conventional levels.[30]

In Table A.4, we use our main sample of postcodes within 300 meters of the invisible boundary segment but restrict our attention to boundaries based on the magnitude of the distance jump in crossing to the *fast* side. In the sample that includes all boundaries except those where the distance jumps by less than 100 meters, the average change in distance from crossing to the *fast*

[30]Note that the results are unchanged when we remove the area and amenities controls in comparing observations very close to the boundary segment, such as those at 100 or 150 meters.

side is 664 meters (see column 1, Panel A). In contrast, when we exclude all boundaries with jumps of less than 500 meters, the average treatment effect intensifies up to almost one kilometer (column 5). The estimates of the per-meter effect of distance are smaller in the samples that include more boundaries that contribute little to the variation, i.e., in columns 1 and 2. In contrast, the per-meter effect is reasonably constant in all samples that exclude (almost irrelevant) boundaries with jumps of 300 meters or higher. The resulting tradeoff between sample size and the exclusion of boundaries that offer little variation motivates the choice to use the 300 meter threshold as the baseline. Moreover, including many boundaries with small variation "on average" introduces more misclassification of local postcodes into the *fast* and *slow* categories.

4.4.2 Sensitivity to School ICT

Table 4: Fuzzy SRD Estimates: Impact of Exchange Distance on Student Performance with School ICT Controls.

Dependent variable: National KS3 exam scores at age 14 (in percentile ranks)					
	(1)	(2)	(3)	(4)	(5)
Nonparametric	0.00122**	0.00124***	0.00124***	0.00103**	0.00092**
Estimates	(0.00047)	(0.00047)	(0.00047)	(0.00047)	(0.00046)
School Proximity to LE Station		0.00047***	0.00047***	0.00058***	
		(0.00010)	(0.00010)	(0.00013)	
Months since ADSL School Upgrade		0.00666	0.00728		
		(0.00652)	(0.03590)		
School ADSL Upgrade Year Dummies	No	No	Yes	No	No
School LE Station Fixed Effect	No	No	No	Yes	No
School Fixed Effect	No	No	No	No	Yes
Observations	183,892	183,712	183,712	183,638	183,598

Notes: The table shows the fuzzy SRD nonparametric estimates for average test scores on the KS3. Columns report different specifications. Each coefficient comes from a separate regression, where the running variable is the residential distance to the LE station invisible boundary. The window size for the residential distance to the LE station invisible boundary is $\pm$ 300 meters. We only include boundaries where the distance jumps by at least 300 meters. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4 estimates the effects on test scores, including several important school technological controls. The positive effect of home broadband on student performance may be partially driven by other technological factors that might be correlated with home broadband. For instance, if the effects are mediated by the fact that the school has high-speed internet, not controlling for this variable would bias our baseline estimate upward. Table 4 compares our baseline estimates (column 1) with four different specifications. We find that student test scores are higher if the school is closer to an LE station and as the number of months since the ADSL upgrade increases (column 2). Although this is not a causal estimate based on an SRD, it is still prone to *passive* sorting. However, our main result is robust to the inclusion of this school-level control. In column

3, we add information on the exact year when the LE that the school is located in upgraded to ADSL technology. Again, the main coefficient does not move. In column 4, we include secondary school LE station fixed effects to absorb any variation that is common to the school LE station. In this specification, the point estimate is lower than the previous ones, but it is significant at the 5% level, and moreover, we cannot reject the hypothesis that it is identical to the baseline estimates at conventional levels of statistical significance. Finally, we add secondary school fixed effects (column 5). Note that this is possible in the English context only because secondary schools do not operate on the basis of residential catchment areas. As a result, there exist a sufficient number of students on both sides of each boundary segment attending different secondary schools. This highly saturated specification nevertheless places high demands on computing power. From this specification, we estimate an effect of crossing the invisible boundary of 0.00092 per meter. The point estimate is slightly lower than the previous ones, but we cannot reject the hypothesis that they are identical at conventional levels of statistical significance. The coefficient also remains significantly different from zero at the 5% level. Moreover, the school fixed effects might absorb possible interactions between home- and school-level technology. In conclusion, we cannot reject the null hypothesis that our baseline estimates are different from any of those obtained with the inclusion of these control variables. Hence, the results indicate that our baseline estimates are not biased by other mediating school technological factors but represent estimates of the causal effect of home broadband.

4.5 Heterogeneous Effects

4.5.1 By Subject

We next explore heterogeneity in the impact of broadband on student test scores. First, we analyze separate regressions for each subject (i.e., English, mathematics and science). This is motivated by the fact that subject differences are often found in the literature on education interventions.[31] Table 5 shows the fuzzy SRD nonparametric estimates on the average student percentile rank score by subject. Column 1 of Table 5 is equivalent to the estimate shown in column 5 of Panel B in Table 3, based on our preferred specification. Splitting the results up by subject, it becomes evident that the effect is the strongest for English, at 0.00141 percentiles per meter. In contrast, the effect is 0.00075 per meter for mathematics and 0.117 per meter for science. However, the confidence intervals of these coefficients overlap, and we cannot reject the hypothesis of equality at conventional levels.

[31] See, for instance, Vigdor et al. (2014) or Falck et al. (2018), who find differential effects between math and reading and math and science, respectively. Other empirical evidence, such as that in Malamud et al. (2019) or Cristia et al. (2017), cannot reject the hypothesis that effects on math and reading are significantly different at conventional levels, but the estimates in both papers have different statistical precision. Machin et al. (2007) find a similar pattern across subjects in their study of the payoff of ICT technology in English primary schools.

Table 5: Fuzzy SRD Estimates: Impact of Exchange Distance on Student Performance by Subject.

Dependent variable: National KS3 exam scores at age 14 (in percentile ranks)				
	Average Score (1)	**English** (2)	**Mathematics** (3)	**Science** (4)
Nonparametric Estimates	0.00122**	0.00141***	0.00075	0.00117**
	(0.00047)	(0.00051)	(0.00048)	(0.00050)
Observations	183,892	183,892	183,892	183,892

Notes: The table shows the fuzzy SRD nonparametric estimates for average test scores on the KS3 by subject. Each coefficient comes from a separate regression, where the running variable is the residential distance to the invisible LE station boundary segment. The window size for the residential distance to the invisible LE station boundary segment is $\pm$ 300 meters, and only boundaries where the exchange distance jumps by at least 300 meters are included. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.5.2 By Student Background Characteristics

We analyze heterogeneity along the following student predetermined characteristics: student gender, prebroadband test scores, ethnicity and free school meal eligibility. Some interesting patterns emerge. First, the effect is not driven by only one gender. The effect per meter of distance is both more positive and more significant for girls than for boys (0.00125 vs. 0.00098, respectively). However, both estimates are positive. Second, the positive effects are strongest for high achievers based on a median split of the prebroadband KS1 test scores. The effect for high achievers is 0.00156 percentiles per exchange distance meter (column 4) and 0.00096 for low achievers (column 5). Third, nonwhite pupils have a larger estimate (0.00211 per meter, column 7), but this effect is also less precisely estimated than that for white pupils (0,00114, column 6). Last but not least, only students who are ineligible for free school meals benefit significantly, with an estimated effect of 0.00132 percentiles per meter in comparison to 0.00073 per meter (approximately half the size) for students with free school meal eligibility.

These resulting coefficients are not distinct from each other at conventional levels of statistical significance. However, the overall pattern is of interest, with the most positive effects for girls, high achievers and students who are not eligible for free school meals. This result is consistent with that of Dettling et al. (2018), who find that the impact of high-speed internet on college applications is concentrated among white students with more educated parents and mainly located in urban and high-income areas. Moreover, Malamud et al. (2019) find no significant effects of home internet access on student achievement for students enrolled in low-achieving primary schools in Peru. Potentially, the smaller effect on the below-median prebroadband KS1 test score and free school meal-eligible groups may partly reflect the trouble that struggling students have in developing effective study strategies for learning (Angrist & Lavy, 2009; Fryer Jr, 2011, e.g.,), in our case in a

home online environment. While none of the groups have negative effect estimates, these results still speak to the hypothesis that home broadband access might exaggerate existing educational inequality by achievement and family affluence.

Table 6: Fuzzy SRD Estimates: Impact of Exchange Distance on Student Performance by Student Characteristics.

Dependent variable: National exam scores at age 14 (KS3)

	Student Gender		KS1 Test Score		Ethnicity		Free School Meal Eligibility	
	Boys	Girls	High	Low	White	Nonwhite	Eligible	Ineligible
	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Nonparametric Estimates	0.00098	0.00125*	0.00156**	0.00096	0.00114**	0.00211	0.00073	0.00132***
	(0.00067)	(0.00068)	(0.00068)	(0.00066)	(0.00051)	(0.00133)	(0.00149)	(0.00050)
Observations	91,106	91,856	89,229	93,763	144,194	38,443	27,496	155,252

Notes: The table shows the fuzzy SRD nonparametric estimates for average test scores on the KS3 for different subgroups of students. Each coefficient comes from a separate regression, where the running variable is the residential distance to the invisible LE station boundary segment. The window size for the residential distance to the invisible LE station boundary segment is $\pm$ 300 meters, and only boundaries where the exchange distance jumps by at least 300 meters are included. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5 Broadband Internet Speed and Student Test Scores: Assessing the Magnitude of the Baseline Estimates

Up until this point, we have focused on using the fuzzy SRD to obtain a causal estimate of the effect of distance from the LE station on test scores, that is, β from equation (1). This is an LATE. However, we are ultimately interested in the relation between broadband speed and test scores. Unfortunately, postcode-level speed data for our study period do not exist. We use out-of-sample-period data to reverse engineer the distance-speed relation present during our study period.

To investigate this relation, we combine the linked postcode-to-exchange station telecom network data with data from the UK's telecommunication regulator (Ofcom) on the average internet speed experienced by households in 2012 and 2013. These data are explained in Section 2.2.3. The Ofcom dataset is based on the speeds of broadband connections operated by the main operators.[32] To replicate average internet broadband speeds that are plausible for our period of analysis (2005–2008), we use only postcodes with average measure speeds of up to 10 Mbit/s per second, thus excluding superfast broadband connections, which were not available before 2008.

The average internet speed in our resulting dataset is 5.5 Mbit/s and not far from UK consumers' actual average download speeds of 4.1 Mbit/s in 2008 (OfCom, 2009a).[33] The dataset

[32]This dataset includes over 13 million connections in 2012 and over 19 million connections in 2013. Annex 1 of the respective Ofcom infrastructure reports provides detailed descriptions (OfCom, 2012; OfCom, 2013).

[33]OfCom (2009a) reports that consumers living in urban areas received average download speeds of 4.3 Mbit/s in

provides us with the average speed (modem sync speeds) recorded in speed tests across individual households, which we can geolocate at the level of their full residential postcode. We use this information to link the speed microdata to the telecom network database discussed above to estimate the relationship between residential exchange distances and available internet speed using the fuzzy SRD design.[34]

Table 7: Fuzzy SRD Estimates: Impact of Exchange Distance on Average Broadband Speed.

Dependent variable: Average internet speed (Mbit/s)			
	(1)	(2)	(3)
Nonparametric Estimates	0.00083***	0.00080***	0.00089***
	(0.00015)	(0.00016)	(0.00018)
Months since ADSL Upgrade			0.0002
			(0.004)
Log Number of Premises			0.167*
			(0.085)
Log Area Square Meters			-0.190**
			(0.087)
Distance Bins	Yes	Yes	Yes
Area & Amenities Controls	No	Yes	Yes
Observations	20,274	20,274	20,274

Notes: The table shows the fuzzy SRD nonparametric estimates for the effect of exchange distance on average internet speed (Mbit/s). The variable "Months since ADSL Upgrade" refers to the number of months since an ADSL upgrade occurred in the LE station catchment area. The variable "Log Number of Premises" refers to the logarithm of the number of premises in the LE station catchment area. The variable "Log Area Square Meters" refers to the logarithm of the square meters of the LE station catchment area. Each coefficient is from a separate regression, where the running variable is the residential distance to the invisible LE station boundary segment. The window size for the residential distance to the invisible LE station boundary is $\pm$ 300 meters. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7 shows the fuzzy SRD estimates of the effect of exchange distance on postcode average (download) speed. These estimates are based on an identical empirical strategy as the one used to derive the effects of distance on test scores, as discussed in Section 3.3. Various specifications that include different sets of technological controls result in very similar estimates based on the boundary jumps: for each additional 100 meters in distance to the connected exchange, speed

comparison with 3.3 Mbit/s among those living in rural areas. Average internet speed varies greatly from country to country. The highest averages can be observed in Asian countries (e.g., South Korea, Hong Kong, Singapore, and Japan) and Scandinavian countries (e.g., Norway, Sweden, Finland, and Denmark). See Inc. (2015).

[34] After 2008, the broadband internet infrastructure in the UK was updated to significantly less distance-sensitive technologies, such as coaxial cables or fibers. Due to this feature of our dataset, we regard our estimates of the effect of crossing the invisible catchment area boundary from the *slow* to the *fast* side on the available internet connection speed as a lower bound (for the test score estimation period). However, in 2012, 75% of broadband subscriptions were still for ADSL technologies, which is sensitive to spatial decay.

decreases by 0.089 Mbit/s. [35] [36] This coefficient is statistically significant at the 1% level. This provides a causal estimate of the impact of proximity to the LE station on average speed, which we label β_{Speed}. By combining the estimate of the impact of proximity to the LE on test scores from equation (1), $\widehat{\beta}$ with $\widehat{\beta_{Speed}}$, we can compute a Wald estimate for the effect of available speed on test scores, ω:

$$\omega = \frac{\beta}{\beta_{Speed}} \tag{4}$$

Applying equation (4) to compute ω yields an effect of $\widehat{\omega} = 1.37$. This implies that for each increase in broadband speed of one Mbit/s, test scores increase on average by 1.37 percentile points. This average effect of one additional Mbit/s is equivalent to approximately 5% of a standard deviation in the national test score distribution. The standard error of $\widehat{\omega}$ is obtained using the Delta method and estimated with high precision (i.e., the p-value is < 0.01).[37] Even ω, which is the policy-relevant parameter, can still be interpreted as a reduced-form effect, as it reflects the effect of broadband availability – and not usage – on test scores.

Finally, to assess the magnitude of the effect of broadband usage on test scores, we complement this Wald estimate with survey data evidence on broadband usage for our study period and information about the extent to which students used the internet for school work. Using data from Oxford Internet Surveys (OxIS), we compute broadband takeup for our study period for the group of interest. We exploit this database because it allows us to be precise in the definition of the relevant population under study: English households with home broadband that had children aged between 14 and 17 years old.[38] We find that the weighted average of broadband usage among these families between 2005 and 2009 is 69.5%. These figures square with the finding reported by Livingstone & Bober (2005) that the vast majority of children used the internet at home, and most of them devoted the use time to do work for school or college. These authors' figures show that in the UK, 84% of 9- to 19-year-olds used the internet daily or weekly in 2005. Among those, 90%

[35] One of the channels through which we can surmise that broadband quality affects test scores is via broadband subscriptions takeup, which increased steadily during our period of analysis. This hypothesis is difficult to test due to the lack of appropriate data, a limitation shared by other papers studying the impact of broadband speed in the UK in this period. Using Ofcom postcode-level data for 2013 – the first year for which the data are available – and limiting the sample to postcodes with speeds below 10 Mbit/s, we test whether the postcode-level takeup rate (either the number of active broadband lines or this number divided by the number of premises) changes across LE boundary segments or if it was affected by the distance to the exchange. In 2013, 70% of active broadband connections were still using ADSL technology. We use the same strategy as that used for the main results in the paper (SRD), and we find precisely estimated zeroes. Even though this test has limitations, the results suggest that the impact of quality on student performance was mainly driven by speed changes rather than by different takeup rates on the *fast/slow* sides of the LE boundary segments.

[36] Note that local averages can be skewed towards extreme values. If the number of observations in the postcode is relatively small and some of the connections are fast or slow, the postcode median (download) speed might be a better measure. We find that the effect of exchange distance on postcode median (download) speed is 0.00126 (p-value < 0.01). Hence, the relevance of the results is unchanged when we use the median instead of the average (download) speed.

[37] See Appendix Section A.4 for more details.

[38] More aggregated datasets used in other studies, such as Eurostat or Ofcom data, lead to similar results. The drawback of using these databases is that they, unlike the OxIS database, do not allow us to refine our results to the specific population of interest.

declared that they used the internet to do work for school or college.

Using our OxIS calculations, which refer to the exact age group that we study, scaling our Wald estimate by broadband takeup shows that for each increase in broadband speed of one Mbit/s, test scores increase on average by 1.97 percentile points. This increase is equivalent to approximately 7% of a standard deviation in the national test score distribution.

6 Conclusion

This paper uses a fuzzy spatial regression discontinuity approach to present estimates of the causal effect of available broadband speed on the test scores of 14-year-old pupils in England. We estimate that an increase in available home broadband speed of 1 Mbit/s leads to an increase in student test scores of approximately 5% of a standard deviation, increasing to 7% when we scale the effects for the population of interest. We find that the effects are not driven by school characteristics and not biased by other mediating technological factors.

The effects that we estimate are important. An effect of 5% of a standard deviation in test scores is economically meaningful. Effects on wages and GDP are difficult to quantify, and this is certainly not possible in our setting. To gauge the potential economic importance, we note that Chetty et al. (2014a) and Chetty et al. (2014b) estimate that one standard deviation higher teaching quality increases student earnings by approximately 1.3%. Taking these effects of teachers on test scores as a reference, our headline estimate may translate into approximately $16,250 more in lifetime earnings per individual.[39]

A limitation of our paper is that we cannot precisely pin down the mechanisms driving the positive relationship between broadband speed and test performance. The net effect stems from the potential positive impact of increasing learning productivity and widening access to educational opportunity, less the potential negative effect of unproductive distraction. Our positive estimate could be the result of an impact on both the extensive margin – if higher speed encourages students to go online more often – and the intensive margin – once online, they spend more time connected. The data requirements to provide causal estimates on these underlying channels are extremely demanding, and available survey data is not rich enough to design a credible identification strategy.[40]

Ultimately, we can only suggest plausible adjustment channels. Higher internet speed increases the amount of information obtainable from the web or shared via the network per unit of

[39] This is based on combining estimates from Chetty et al. (2014b) that a one standard deviation improvement in teacher quality increases earnings at age 28 by 1.3%, equivalent to $39,000 if this effect remains constant over the life cycle, with the standardized effect of teachers on test scores from Chetty et al. (2014a), which they estimate at 0.10 and 0.14 for English and math. Averaging over subjects, we get $39,000 \times (0.05/(0.10 + 0.14/2)) = \$16.250$.

[40] Moreover, another limitation of survey data is that it generally provides answers with few categorical options, which are difficult to quantify.

time.[41] During our study period, multiple online resources offering educational support became available to teenagers. For instance, *YouTube* was launched in 2005, and soon after, educational institutions and individuals began uploading educational content, with keen interest from the platform to help educators.[42] The number of articles in encyclopedia-type sites, such as *Wikipedia*, also increased exponentially since 2004. In addition, students also could make use of early social media tools (such as *Microsoft Messenger*) to share information and study together, or visit sites such as *Sparknotes* to download essay-writing content. The existence of all these online resources coupled with the amount of time spent online at home for students in our age bracket indicates that the sign and magnitude of our results is relevant and credible.[43]

Our results provide new insights to inform policy decisions related to investments in high-speed broadband networks. Governments have used the argument that investing public funds into broadband infrastructure can boost firm performance and employment, while such arguments are hardly ever deployed for education-related policies (SQW, 2014). This is particularly salient in the UK with its recent ultrafast fiber broadband rollout, where there exist major political concerns about rural areas being left behind and missing growth opportunities due to the lack of private ICT investment incentives (see, for example, DCMS, 2011; DDCMS, 2018). Moreover, the vast majority of the programs focused on providing broadband for educational purposes have been targeted at schools.[44] However, this paper finds that the effects of home broadband on human capital formation are not trivial. Our results show that home broadband speed matters for student performance to an extent comparable to many more direct inputs of educational production. Broadband technology therefore increases the importance of the home environment for learning.[45]

The findings of this paper are even more pertinent in the context of the recent COVID-19 pandemic, in which many countries closed schools for several months. Hence, home online learning became a major substitute for in-classroom teaching. The pandemic brought to the spotlight

[41] The theoretical time taken to perform various online activities changes dramatically with the internet connection speed. For instance, downloading a 250 Kbit/s web-page takes 1 and 0.3 second with a provider of 2 Mbit/s and 8 Mbit/s, respectively. Besides, this difference grows exponentially with the online activity provided. Downloading a DVD quality film (4GB) takes 4hours and 48 minutes, and 1 hour and 11 minutes with a provider of 2 Mbit/s and 8 Mbit/s, respectively. See Figure 4.1 of OfCom (2010).

[42] Multiple popular education-oriented channels existed in this period. The founder of the *Khan Academy* started posting videos in 2006, and created the academy channel in 2008. His videos became very popular, attracting tens of thousands of viewers every month. *Smarter Every Day* started its educational videos in 2007, and *TED Talks* began sharing talks in 2006. The demand of educational videos led to the creation of *YouTube EDU* in 2009, as a repository for its educational content.

[43] According to OfCom (2008), the internet consumption for children aged 12-15 at home doubled between 2005 and 2007, from 7.1 to 13.8 average weekly hours. Access and consumption change remarkably by socio-economic category.

[44] For example, the Department for Education committed in 2019 to get gigabit (100 Mbit/s) capable speed broadband connections to over 100 rural primary schools.

[45] Interestingly, positive effects on educational outcomes have also been documented from the roll-out of TV (Gentzkow & Shapiro, 2008), suggesting that the take-up of new inputs into education production at home might have positive effects independent of the precise technology newly introduced.

the consequences of unequal access to technology, as some families with lower resources could not provide their children with the appropriate infrastructure to engage in digital school work (Bacher-Hicks et al., 2020). Overall, our results highlight the value added of broadband investments and the importance of ensuring universal access to mitigate increases in inequalities in educational opportunity.

References

Ahlfeldt, G., Koutroumpis, P., & Valletti, T. (2017). Speed 2.0: Evaluating access to universal digital highways. *Journal of the European Economic Association*, 15(3), 586–625.

Akerman, A., Gaarder, I., & Mogstad, M. (2015). The skill complementarity of broadband internet. *The Quarterly Journal of Economics*, 130(4), 1781–1824.

Amaral Garcia, S., Nardotto, M., Propper, C., & Valletti, T. (2019). *Mums Go Online: Is the Internet Changing the Demand for Healthcare?* CEPR Discussion Papers 13625, C.E.P.R. Discussion Papers.

Angrist, J. & Lavy, V. (2002). New Evidende on Classroom Computers and Pupil Learning. *Economic Journal*, 112(482), 735–765.

Angrist, J. & Lavy, V. (2009). The effects of high stakes high school achievement awards: Evidence from a randomized trial. *American economic review*, 99(4), 1384–1414.

Angrist, J. D., Lavy, V., Leder-Luis, J., & Shany, A. (2019). Maimonides' rule redux. *American Economic Review: Insights*, 1(3), 309–24.

Bacher-Hicks, A., Goodman, J., & Mulhern, C. (2020). Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time. *Journal of Public Economics*, 193, 104345.

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in india. *The Quarterly Journal of Economics*, 122(3), 1235–1264.

Barrera-Osorio, F. & Linden, L. L. (2009). *The use and misuse of computers in education: evidence from a randomized experiment in Colombia*. The World Bank.

Barro, R. J. (2001). Human capital and Growth. *American Economic Review*, 91(2), 12–17.

Barrow, L., Markman, L., & Rouse, C. E. (2009). The Educational Benefits of Computer-Aided Instruction. *American Economic Journal: Economic Policy*, 1(1), 52–74.

Battle, Paul Attewell, J. (1999). Home computers and school performance. *The information society*, 15(1), 1–10.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of political economy*, 70(5, Part 2), 9–49.

Bellou, A. (2015). The Impact of Internet Diffusion on Marriage Rates: Evidence from the Broadband Market. *Journal of Population Economics*, 28(2), 265–297.

Belo, R., Ferreira, P., & Telang, R. (2014). Broadband in school: Impact on student performance. *Management Science*, 60(2), 265–282.

Belo, R., Ferreira, P., & Telang, R. (2016). Spillovers from wiring schools with broadband: the critical role of children. *Management Science*, 62(12), 3450–3471.

Beuermann, D. W., Cristia, J., Cueto, S., Malamud, O., & Cruz-Aguayo, Y. (2015). One laptop per child at home: Short-term impacts from a randomized experiment in peru. *American Economic Journal: Applied Economics*, 7(2), 53–80.

Bhuller, M., Havnes, T., Leuven, E., & Mogstad, M. (2013). Broadband Internet: An Information Superhighway to Sex Crime? *Review of Economic Studies*, 8(4), 1237–1266.

Bulman, G. & Fairlie, R. W. (2016). Technology and Education: Computers, Software, and the Internet. *Handbook of the Economics of Education*, 5, 239–280.

Burgess, S. M. (2016). Human capital and education: The state of the art in the economics of education.

Campante, F., Durante, R., & Sobbrio, F. (2018). Politics 2.0: The multifaceted effect of broadband internet on political participation. *Journal of the European Economic Association*, 16(4), 1094–1136.

Cattaneo, M. D., Jansson, M., & Ma, X. (2018). Manipulation testing based on density discontinuity. *The Stata Journal*, 18(1), 234–261.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9), 2593–2632.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633–79.

Clement, J. (2020). Global digital population as of july 2020.

Cristia, J., Ibarrarán, P., Cueto, S., Santiago, A., & Severín, E. (2017). Technology and child development: Evidence from the one laptop per child program. *American Economic Journal: Applied Economics*, 9(3), 295–320.

Czernich, N., Falck, O., Kretschmer, T., & Woessmann, L. (2011). Broadband infrastructure and economic growth. *The Economic Journal*, 121(552), 505–532.

DCMS (2011). *Broadband Delivery Program: Delivery Model*. Policy report, Department for Culture, Media and Sport.

DDCMS (2018). *Future Telecoms Infrastructure Review*. Policy report, Department for Digital, Culture, Media and Sport.

Deming, D. J., Goldin, C., Katz, L. F., & Yuchtman, N. (2015). Can online learning bend the higher education cost curve? *American Economic Review*, 105(5), 496–501.

Dettling, L. J., Goodman, S., & Smith, J. (2018). Every little bit counts: The impact of high-speed internet on the transition to college. *Review of Economics and Statistics*, 100(2), 260–273.

Escueta, M., Quan, V., Nickow, A. J., & Oreopoulos, P. (2017). *Education technology: An evidence-based review*. Technical report, National Bureau of Economic Research.

Fairlie, R. W., Beltran, D. O., & Das, K. K. (2010). Home computers and educational outcomes: Evidence from the nlsy97 and cps. *Economic inquiry*, 48(3), 771–792.

Fairlie, R. W. & Robinson, J. (2013). Experimental Evidence on the Effects of Home Computers on Academic Achievement among Schoolchildren. *American Economic Journal: Applied Economics*, 5(3), 211–240.

Falck, O., Gold, R., & Heblich, S. (2014). E-Lections: Voting Behavior and the Internet. *American Economic Review*, 107(7), 2238–2265.

Falck, O., Mang, C., & Woessmann, L. (2018). Virtually no effect? different uses of classroom computers and their effect on student achievement. *Oxford Bulletin of Economics and Statistics*, 80(1), 1–38.

Fiorini, M. (2010). The effect of home computer use on children's cognitive and non-cognitive skills. *Economics of Education review*, 29(1), 55–72.

Fryer Jr, R. G. (2011). Financial incentives and student achievement: Evidence from randomized trials. *The Quarterly Journal of Economics*, 126(4), 1755–1798.

Gavazza, A., Nardotto, M., & Valletti, T. (2019). Internet and politics: Evidence from uk local elections and local government policies. *The Review of Economic Studies*, 86(5), 2092–2135.

Gelman, A. & Imbens, G. (2014). Why High-order Polynomials Should not be Used in Regression Discontinuity Designs. NBER Working Paper Num. 20405.

Gentzkow, M. & Shapiro, J. M. (2008). Preschool Television Viewing and Adolescent Test Scores: Historical Evidence from the Coleman Study. *The Quarterly Journal of Economics*, 123(1), 279–323.

Gibbons, S., Silva, O., & Weinhardt, F. (2013). Everybody needs good neighbours? evidence from students' outcomes in england. *The Economic Journal*, 123(571), 831–874.

Gibbons, S., Silva, O., & Weinhardt, F. (2017). Neighbourhood Turnover and Teenage Attainment. *Journal of the European Economic Association*, 15(4), 746–783.

Gibbons, S. & Telhaj, S. (2016). Peer effects: Evidence from secondary school transition in england. *Oxford Bulletin of Economics and Statistics*, 78(4), 548–75.

Goolsbee, A. & Guryan, J. (2006). The Impact of Internet Subsidies in Public Schools. *Review of Economics and Statistics*, 88(2), 336–347.

Hanushek, E. A. & Woessmann, L. (2011). The economics of international differences in educational achievement. In *Handbook of the Economics of Education*, volume 3 (pp. 89–200). Elsevier.

Hanushek, E. A. & Woessmann, L. (2012). Do better schools lead to more growth? cognitive skills, economic outcomes, and causation. *Journal of economic growth*, 17(4), 267–321.

Hanushek, E. A. & Woessmann, L. (2015). *The knowledge capital of nations: Education and the economics of growth*. MIT press.

Inc., A. T. (2015). *Akamai's [state of the internet]. Q1 2015 report*. Technical Report 1.

Jacob, B. A. (2004). Public housing, housing vouchers, and student achievement: Evidence from public housing demolitions in chicago. *American Economic Review*, 94(1), 233–258.

Lavy, V., Silva, O., & Weinhardt, F. (2012). The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools. *Journal of Labor Economics*, 30(2), 367 – 414.

Leuven, E. & Løkken, S. A. (2020). Long-term impacts of class size in compulsory school. *Journal of Human Resources*, 55(1), 309–348.

Livingstone, S. & Bober, M. (2005). Uk children go online: Final report of key project findings.

Machin, S., McNally, S., & Silva, O. (2007). New Technology in Schools: Is There a Payoff? *Economic Journal*, 117(522), 1145–1167.

Malamud, O., Cueto, S., Cristia, J., & Beuermann, D. W. (2019). Do children benefit from internet access? experimental evidence from peru. *Journal of Development Economics*, 138, 41–56.

Malamud, O. & Pop-Eleches, C. (2011). Home Computer Use and the Development of Human Capital. *Quarterly Journal of Economics*, 126(2), 987–1027.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, 142(2), 698–714.

McKeay, M. (2017). *"Q2 2017 state of the internet security report*. Technical report, Akamai, Tech. Rep., 2017.[Online]. Available: https://www. akamai. com/us/en ....

Mincer, J. (1974). Schooling, experience, and earnings. human behavior & social institutions no. 2.

Muralidharan, K., Singh, A., & Ganimian, A. J. (2019). Disrupting education? experimental evidence on technology-aided instruction in india. *American Economic Review*, 109(4), 1426–60.

Nardotto, M., Valletti, T., & Verboven, F. (2015). Unbundling the Incumbent: Evidence from UK Broadband. *Journal of the European Economic Association*, 13(2), 330–362.

OfCom (2008). *Media Literacy Audit: Report on UK children's media literacy*. Technical report.

OfCom (2009a). *UK broadband speeds 2008. Consumer experience of broadband performance: initial findings*. Research document, OfCom.

OfCom (2009b). *UK broadband speeds 2009. Consumers' experience of fixed-line broadband performance*. Research report, OfCom.

OfCom (2010). *UK fixed broadband speeds, November/December 2010*. Research document, OfCom.

OfCom (2012). *2012 UK Communications Infrastructure Report*. Report.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2), 417–458.

Rouse, C. E. & Krueger, A. B. (2004). Putting computerized instruction to the test: a randomized evaluation of a "scientifically based" reading program. *Economics of Education Review*, 23(4), 323–338.

Ryan, C. & Sibieta, L. (2010). Private schooling in the uk and australia.

Sanbonmatsu, L., Kling, J., Duncan, G., & Brooks-Gunn, J. (2006). Neighborhoods and academic achievement: Results from the moving to opportunity experiment. *The Journal of Human Resources*, XLI.

Schultz, T. W. (1961). Investment in human capital. *The American economic review*, (pp. 1–17).

Slater, H., Davies, N. M., & Burgess, S. (2012). Do teachers matter? measuring the variation in teacher effectiveness in england*. *Oxford Bulletin of Economics and Statistics*, 74(5), 629–645.

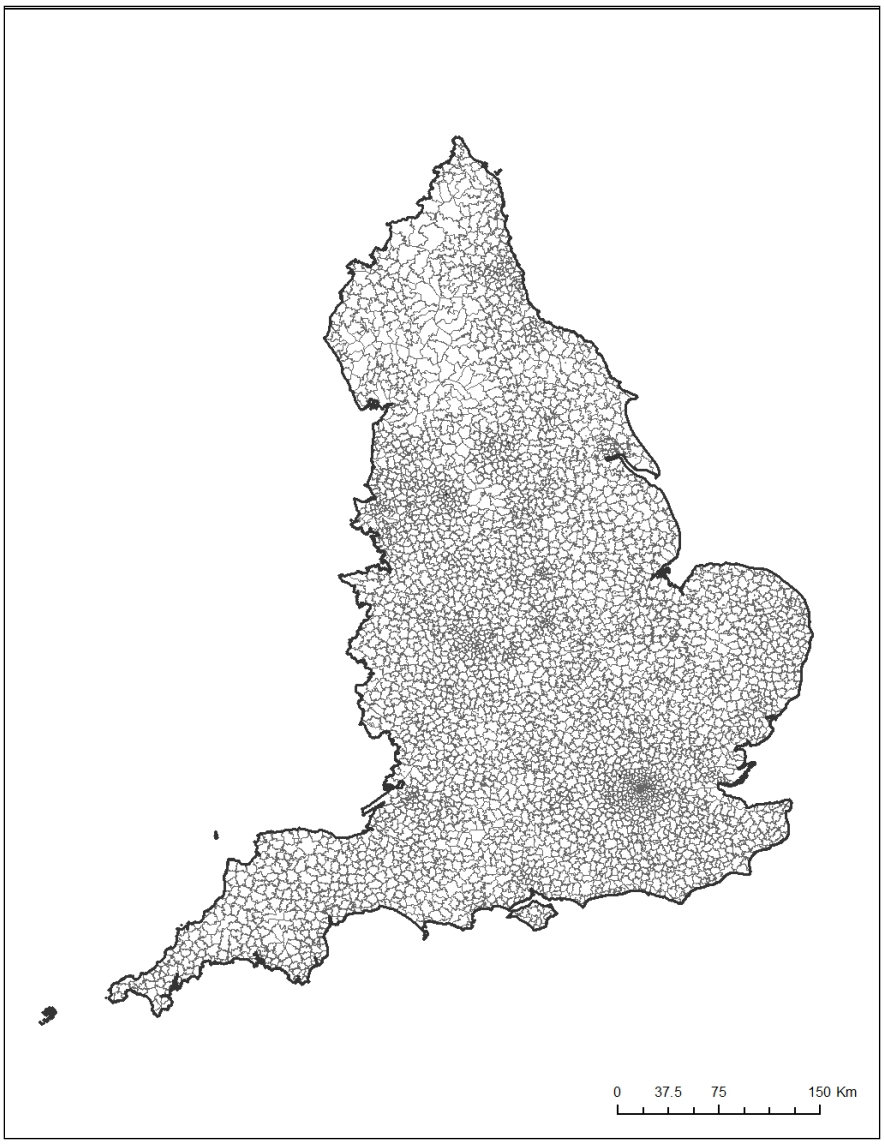
SQW (January 2014). UK Broadband Impact Study - Baseline Report - January 2014. Technical Report for the Department for Culture, Media and Sport.

Vigdor, J., Ladd, H., & Martinez, E. (2014). Scaling the Digital Divide: Home Computer Technology and Student Achievement. *Economic Inquiry*, 52(3), 1103–1119.

Appendix Figures and Tables

A.1 Telephone Local Exchange Areas in England

Figure A.1: Telephone Local Exchange Areas in England



Notes: There are 3,978 local exchange areas in England (of which 3,925 are completely contained in England). The average area is of 34 square kilometres, and serves an average of 5,830 premises (with 93% of them residential).

A.2 Validity of the Research Design

Table A.1: McCrary (2008) and Cattaneo et al. (2018) Test for Manipulation of the Forcing Variable for the Different Treatment subamples.

A. McCrary Test				
Distance (meters)	**Log Difference in frequency bins**	**Z-stat**	**Bandwidth**	**Bin size**
250 meters	.267 (0.206)	1.29	40.14	.556
300 meters	-.024 (0.068)	.347	44.92	.604
350 meters	-.043 (0.048)	.88	50.22	.65
400 meters	-.038 (0.04)	.936	54.5	.69
450 meters	-.038 (0.035)	1.07	59.69	.73
500 meters	-.041 (0.031)	1.33	66.7	.77
B. Cattaneo et al. (2018): RD Manipulation Test using local polynomial density estimation:				
Bias-corrected Density Estimate to the left of the cutoff	.00035 (0.000017)			
Bias-corrected Density Estimate to the right of the cutoff	.00033 (0.000018)			
T-test for bias-corrected density test	-0.7886			
P-value for bias-corrected density test	0.4303			

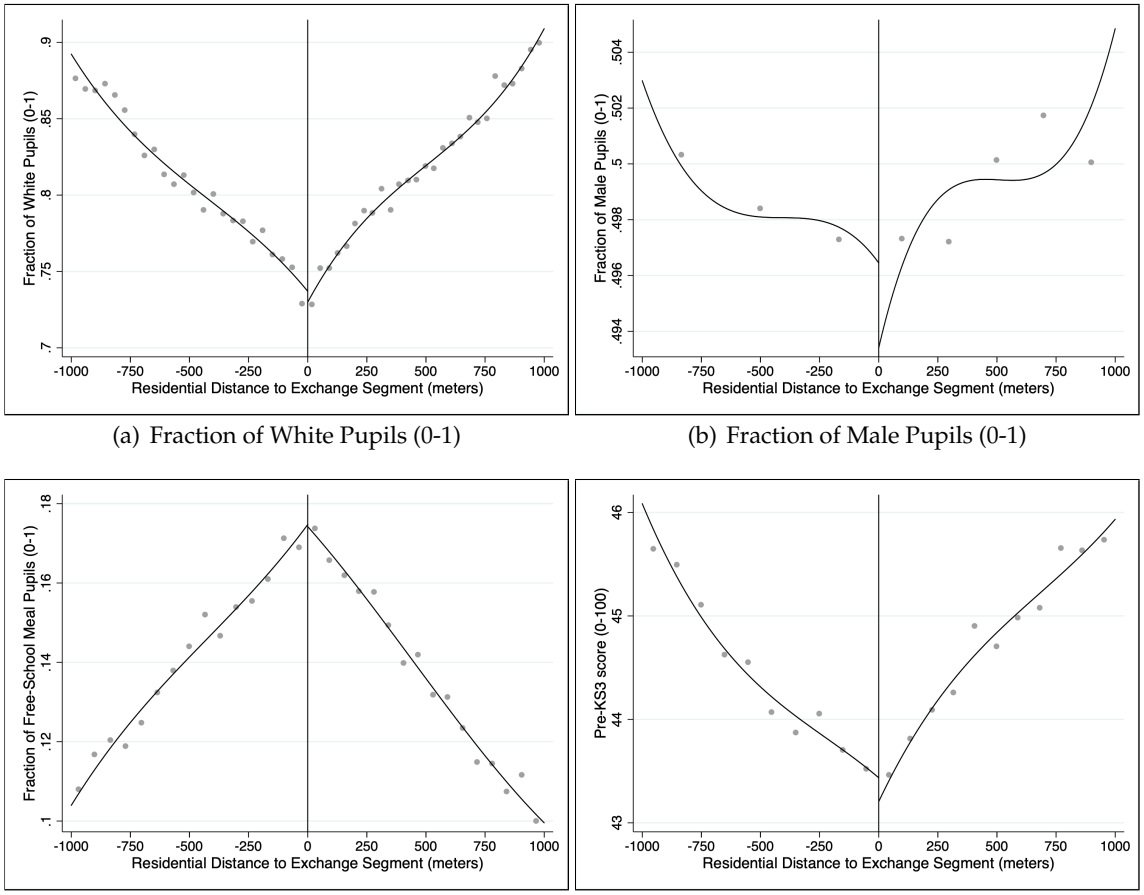
Notes: This table show the McCrary (2008) (Panel A) and Cattaneo et al. (2018) test for manipulation of the forcing variable. The McCrary (2008) test is performed separately for each treatment sample. The table columns of Panel A show the estimated discontinuity in the density function of the assignment variable at the threshold, its standard error (in parentheses), the associated z-statistic, the estimated optimal bandwidth, bin size and the number of observations. The optimal bandwidth and bin size are obtained using the selection procedure proposed by McCrary (2008). The table rows of Panel B present the bias corrected density estimate to the left and right of the invisible LE boundaries, the t-test for the bias-corrected density test and the p-value of the test.

Table A.2: Sharp-SRD Estimates: The Impact Exchange Distance on Student Performance

	(1)	(2)	(3)	(4)	(5)
Effect of exchange distance on national age-14 KS3 exam scores (in meters)					
Non-Parametric	0.00043***	0.00034**	0.00027**	0.00025*	0.00024*
Estimates	(0.00014)	(0.00014)	(0.00013)	(0.00013)	(0.00013)
Student Pre-Internet Score KS1	Yes	Yes	Yes	Yes	Yes
Student Controls	No	Yes	Yes	Yes	Yes
Time-Variant Area & Amenities	No	No	Yes	Yes	Yes
Time-Invariant Area & Amenities	No	No	No	Yes	Yes
School Controls	No	No	No	No	Yes
Observations	183,892	183,892	183,892	183,892	183,892

Notes: The table shows the sharp SRD non-parametric estimates for test scores in KS3. Columns report different specifications. Each coefficient comes from a separate regression, where the running variable is the residential distance to the LE station invisible boundary. The window size for the residential distance to the LE station invisible boundary is $\pm$ 300 meters. We only include boundaries where the distance jumps by at least 300 meters. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.2: Students Characteristics.



(a) Fraction of White Pupils (0-1)

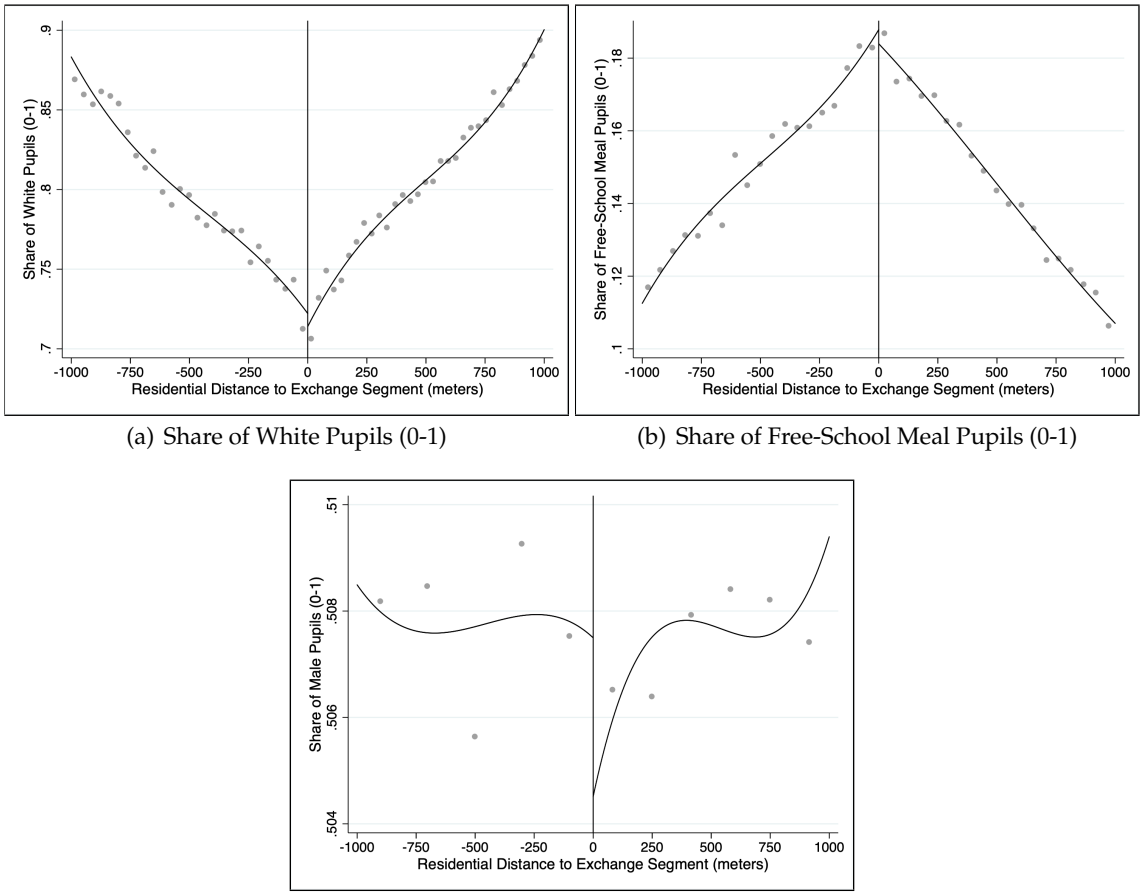
(b) Fraction of Male Pupils (0-1)

(c) Fraction of Free-School Meal Pupils (0-1)

(d) Pre-KS3 score (0-100)

Notes: The dots represent the fraction of white pupils, the fraction of male pupils, the fraction of free-school meal pupils and the pre-internet score per interval of residential distance to the exchange segment invisible boundaries. The solid lines are fitted values from a third order polynomial approximation, which is estimated separately on both sides of the cutoffs. "Residential Distance to Exchange Segment (meters)" refers to the residential distance to the exchange segment invisible boundaries. Black vertical lines identify the LE station invisible boundaries.

Figure A.3: Density & Area Socio-Economics.



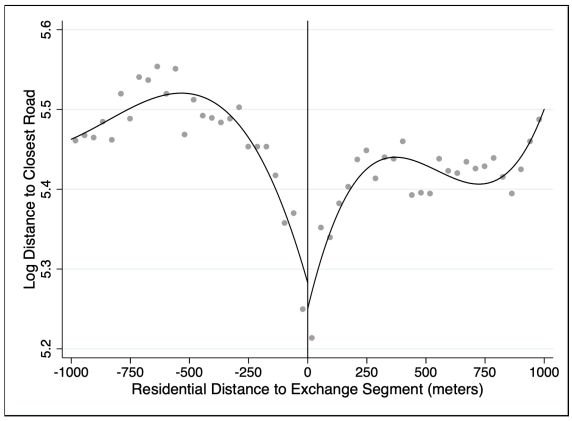
(a) Share of White Pupils (0-1)

(b) Share of Free-School Meal Pupils (0-1)

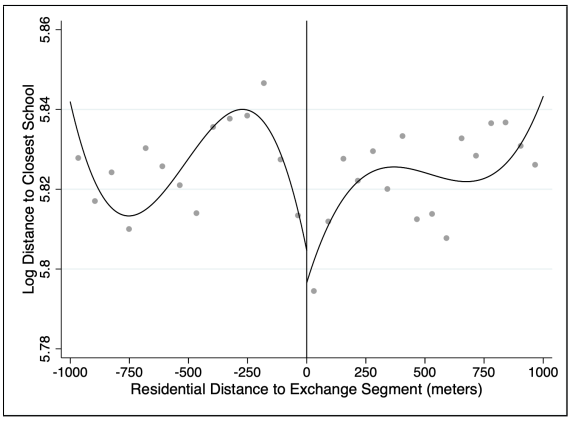
(c) Share of Male Pupils (0-1)

Notes: The dots represent the share of white pupils, share of free-school meal and share of male pupils per interval of residential distance to the exchange segment invisible boundaries.. The solid lines are fitted values from a third order polynomial approximation, which is estimated separately on both sides of the cutoffs. "Residential Distance to Exchange Segment (meters)" refers to the residential distance to the exchange segment invisible boundaries. Black vertical lines identify the LE station invisible boundaries

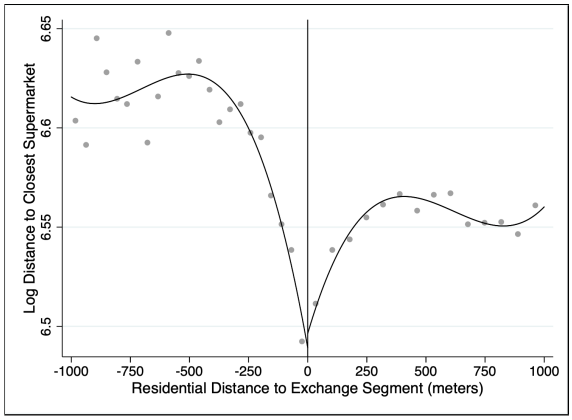
Figure A.4: Amenities.



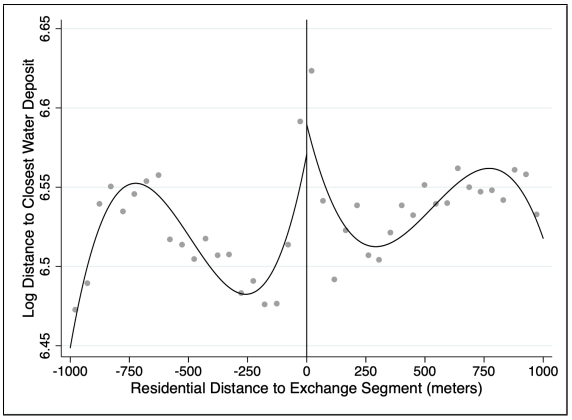
(a) Log Distance to Closest Road



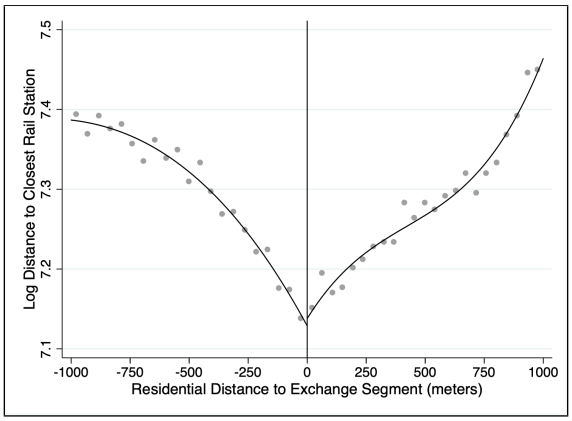
(b) Log Distance to Closest School



(c) Log Distance to Closest Supermarket



(d) Log Distance to Closest Water Deposit



(e) Log Distance to Closest Rail Station

Notes: The dots represent the logarithm distance to closest road, school, supermarket, water deposit and rail station per interval of residential distance to the exchange segment invisible boundaries. The solid lines are fitted values from a third order polynomial approximation, which is estimated separately on both sides of the cutoffs. "Residential Distance to Exchange Segment (meters)" refers to the residential distance to the exchange segment invisible boundaries. Black vertical lines identify the LE station invisible boundaries.

A.3 Robustness of Baseline Estimates

Table A.3: Sample Choice Based on Distance to Invisible Segment-Boundary

	Distance (meters)						
	100	**150**	**200**	**300**	**500**	**750**	**1000**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Avg. jump in distance to connected exchange when crossing boundary to the *Fast* side (in m)							
Non-Parametric	-779.4***	-794.1***	-805.9***	-816.6***	-820.3***	-828.5***	-832.9***
Estimates	(19.7)	(18.4)	(16.7)	(15.9)	(15.4)	(15.3)	(15.3)
B. Effect of exchange distance on national age-14 KS3 exam scores (in percentile ranks)							
Non-Parametric	0.00157**	0.00154**	0.00134***	0.00122**	0.00092**	0.00101**	0.00100**
Estimates	(0.00066)	(0.00060)	(0.00051)	(0.00047)	(0.00045)	(0.00043)	(0.00043)
Observations	49,303	85,182	119,55	183,892	305,404	431,709	530,773

Notes: The table shows the fuzzy SRD non-parametric estimates for the average distance to LE station (Panel A) and test scores in KS3 (Panel B) with different samples based on the distance to invisible segment-boundary. Columns report different specifications. Each coefficient comes from a separate regression, where the running variable is the residential distance to the LE station invisible boundary. We only include boundaries where the distance jumps by at least 300 meters. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: Sample Choice Based on Magnitude of Jump across Invisible Segment-Boundary

	Minimum Boundary Jump (in meters)					
	100	**200**	**250**	**300**	**400**	**500**
	(1)	(2)	(3)	(4)	(5)	(6)
A. Avg. jump in distance to connected exchange when crossing boundary to the *Fast* side (in m)						
Non-Parametric	-653.788***	-731.578***	-775.172***	-816.649***	-915.704***	-999.089***
Estimates	(13.477)	(14.451)	(15.254)	(15.927)	(17.194)	(18.996)
B. Effect of exchange distance on national age-14 KS3 exam scores (in percentile ranks)						
Non-Parametric	0.00064	0.00080*	0.00116**	0.00122**	0.00131***	0.00135***
Estimates	(0.00051)	(0.00049)	(0.00048)	(0.00047)	(0.00045)	(0.00046)
Observations	671,601	601,366	561,921	530,773	463,692	396,354

Notes: The table shows the fuzzy SRD non-parametric estimates for the average distance to LE station (Panel A) and test scores in KS3 (Panel B) with different samples based on the minimum boundary jump across the invisible segment-boundary. Columns report different specifications. Each coefficient comes from a separate regression, where the running variable is the residential distance to the LE station invisible boundary. The window size for the residential distance to the LE station invisible boundary is $\pm$ 300 meters. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Fuzzy-SRD Estimates: The Impact of Crossing the Invisible Broadband Boundary on Student Performance by Clustering choices for standard errors.

	(1)	(2)	(3)	(4)	(5)
Non-Parametric	0.00122**	0.00122**	0.00122**	0.00122***	0.00122**
Estimates	(0.00047)	(0.00032)	(0.00052)	(0.00047)	(0.00051)
Segment-Year Level	Yes	No	No	No	No
Segment and Year Level	No	Yes	No	No	No
Segment Level	No	No	Yes	No	No
Boudary-Year Level	No	No	No	Yes	No
Boundary Level	No	No	No	No	Yes
Observations	183,892	183,892	183,892	183,892	183,892

Notes: The table shows the fuzzy SRD non-parametric estimates for test scores in KS3 by different clustering choice for standard errors. Columns report different specifications. Each coefficient comes from a separate regression, where the running variable is the residential distance to the LE station invisible boundary. The window size for the residential distance to the LE station invisible boundary is $\pm$ 300 meters. We only include boundaries where the distance jumps by at least 300 meters. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.4 Delta Method Calculations

We use the test statistics:

$$\frac{\frac{\widehat{\beta}}{\widehat{\beta_{Speed}}}}{SD(\frac{\widehat{\beta}}{\widehat{\beta_{Speed}}})} \sim N(0,1) \tag{A.1}$$

Assuming that our estimates are independent, we can compute:

$$VAR(\frac{\widehat{\beta}}{\widehat{\beta_{Speed}}}) = \frac{1}{(\widehat{\beta_{Speed}})^2} * (VAR(\widehat{\beta}) + (\frac{\widehat{\beta}}{\widehat{\beta_{Speed}}})^2 * VAR(\widehat{\beta_{Speed}})) \tag{A.2}$$