

ICT and Education: Evidence from Student Home Addresses

Benjamin Faber (UC Berkeley and NBER)
Rosa Sanchis-Guarner (SERC, CEP and LSE)
Felix Weinhardt (Humboldt-University, CEP and LSE)

October 2015

This work is part of the research programme of the UK Spatial Economics Research Centre funded by a grant from the Economic and Social Research Council (ESRC). The views expressed are those of the authors and do not represent the views of the ESRC.

© B. Faber, R. Sanchis-Guarner and F. Weinhardt, submitted 2015

ICT and Education: Evidence from Student Home Addresses

Benjamin Faber*
Rosa Sanchis-Guarner**
Felix Weinhardt***

October 2015

* Department of Economics, UC Berkeley and NBER

** Spatial Economics Research Centre and Centre for Economic Performance, London School of Economics and Political Science.

*** Department of Economics, Humboldt-University, Berlin, Centre for Economic Performance, London School of Economics and Political Science, and IZA

We are grateful to the English Department for Education for allowing us access to the confidential microdata of the National Pupil Database (NPD) and the Longitudinal Study of Young People in England (LSYPE), to the technology consulting firm Point Topic Ltd for allowing us access to their proprietary data on the UK telecommunication network, to the staff at the UK Secure Data Service (SDS) for providing us access to the confidential microdata of the Understanding Society (US) and the British Household Panel Surveys (BHPS), and to the staff at the LSE Research Library for providing us access to the microdata of the English Land Registry. Amar Shanghavi provided excellent research assistance. We thank Steve Gibbons, Victor Lavy, Magne Mogstad, Jonathan Robinson, Andres Rodriguez-Clare, Olmo Silva and participants at multiple seminars and conferences for helpful comments. Weinhardt would like to thank Steve Machin for advice and support during his post-Doc at the LSE. This work contains statistical data from ONS which is Crown copyright and reproduced with the permission of the controller of HMSO and Queen's Printer for Scotland. The use of the ONS statistical data does not imply the endorsement of the ONS in relation to the interpretation or analysis of the data. This work uses research datasets which may not exactly reproduce National Statistics aggregates. Copyright of the statistical results may not be assigned, and publishers of this data must have or obtain a license from HMSO. The ONS data in these results are covered by the terms of the standard HMSO "click-use" license. SERC/LSE provided research funding and Weinhardt acknowledges DFG funding (WE5750/1-1). All errors are our own.

Abstract

Governments are making it a priority to upgrade information and communication technologies (ICT) with the aim to increase available internet connection speeds. This paper presents a new empirical strategy to estimate the causal effects of these policies, and applies it to the questions of whether and how ICT upgrades affect educational attainment. We draw on a rich collection of microdata that allows us to link administrative test score records for the population of English primary and secondary school students to the available ICT at their home addresses. To base estimations on exogenous variation in ICT, we notice that the boundaries of usually invisible telephone exchange station catchment areas give rise to substantial and essentially randomly placed jumps in the available ICT across neighboring residences. Using this design across more than 20,000 boundaries in England, we find that even very large changes in available broadband connection speeds have a precisely estimated zero effect on educational attainment. Guided by a simple model we then bring to bear additional microdata on student time and internet use to quantify the potentially opposing mechanisms underlying the zero reduced form effect. While jumps in the available ICT appear to increase student consumption of online content, we find no significant effects on student time spent studying online or offline, or on their learning productivity.

Keywords: Education, information and communication technology, internet

JEL Classifications: F66; I20; D83

1 Introduction

Governments around the world are currently in the process of committing substantial amounts of public funds to upgrade information and communication technologies (ICT) with the aim to increase available broadband internet connection speeds.¹ These policies are motivated by claims that upgrades to high-speed broadband play a vital role in achieving policy objectives in two key areas: education and growth. Policy proposals have emphasized the link between broadband upgrades and educational attainment in light of both the astounding amounts of time that students of all age groups devote to online content every day, and abundant online educational resources ("e-learning") that range from longstanding platforms, such as the thousands of homework support channels on YouTube or Wikipedia, to more recent developments such as massive open online courses (MOOCs).²

However, despite bold presumptions about the effects of ICT upgrades, these policies are currently based on limited empirical evidence. This paper proposes a new empirical strategy to estimate the causal effects of these policies, and applies it to the questions of whether and how ICT upgrades affect educational attainment. In doing so, the paper makes three main contributions to the existing literature. First, we draw on a new and uniquely rich collection of English microdata that allows us for the first time to link administrative test score records of the population of primary and secondary school students to the available ICT at student home addresses.³ Second, to base estimations on exogenous variation in the available ICT, we notice that capacity constraints at telephone exchange stations lead to invisible and essentially randomly placed boundaries of station-level catchment areas that give rise to substantial and discontinuous jumps in the available ICT across neighboring residences. Third, guided by a simple theoretical framework we bring to bear additional microdata on student internet and time use to quantify the channels underlying the estimated reduced form effects. As part of these estimations, we also provide empirical evidence on the elasticity of demand for online content with respect to its time cost, which is a key parameter for ICT policy decisions.

At the center of the analysis lies the construction of an extremely rich collection of microdata that allows us to estimate the causal effect of changes in the available ICT on educational outcomes, and to quantify the underlying channels. We combine several sources of microdata that we are able to geo-reference at a very disaggregated spatial scale: i) Administrative test score records for the population of English primary and secondary school students in national "Key Stage" exams over the period 2002-2008; ii) telecommunication network data including the position of the universe (≈ 3900) of English telephone exchange stations, their assignments to each of roughly 1.5

¹In the US, the FCC launched the National Broadband Plan in 2010 with the objective to make 100Mbit/s connections available to 100 million citizens by 2020. Cost estimates have ranged between USD 20bn and 250bn. In 2012, the EU approved EUR 7bn (USD 7.9bn) for increases in internet speeds across member countries. The UK plans to cover 90% of its population with "superfast" broadband by the end of 2016 at an estimated cost of GBP 3.9bn (USD 6.1bn).

²In 2007/08 UK students spent on average between five (ages 5-7) and fourteen hours (ages 12-15) per week using the internet from home (OfCom, 2012). Common examples of policy claims about the link between high-speed internet and education and growth include the White House Broadband Report (June 2013), the FCC's National Broadband Plan (March 2010), the White House Report on Obama's most recent ConnectHome initiative (July 2015), or the UK Broadband Delivery Scheme (February 2013).

³See discussion of related literature at the end of this section.

million full postcodes (up to eight characters), and household internet connection speeds; iii) survey microdata on student time use and internet use over the period 2002-2013; and iv) a rich vector of geo-referenced control variables including the universe of property transaction values over the period 2002-2008 and residential proximity to a comprehensive list of local (dis-)amenities.

Using these data, we present a new strategy to exploit a well-known feature of DSL-broadband technology –that the length of the copper wire that connects residences to the local exchange station is a key determinant of the available internet connection speeds.⁴ One possible strategy in this context would be to compare the outcomes of students whose residences are located at different distances to connected exchanges. However, neither observed differences in ICT across households (which include package choices), nor residential distances to the connected exchange stations are randomly assigned across space. Today’s telephone exchange stations in the UK were mainly installed before and during the Second World War and while it may seem unlikely that residents actively sort as a function of usually unknown copper wire distances to the local telephone exchange station, planners targeted central locations that were also close to road junctions for hosting the switchboard infrastructure. To address a number of endogeneity concerns that arise due to the non-random placement of the exchange stations, we propose an empirical strategy that exploits boundary discontinuities across usually unobserved exchange station catchment areas.

We notice that each telephone exchange station has an invisible catchment area of residential addresses that it serves in its surrounding. The extent and shape of this catchment area is a by-product of history: Rapid growth in fixed-line telephony during and after the Second World War in combination with capacity constraints at the exchange switchboards have led to invisible and essentially randomly placed station-level boundaries. We use the telecom network microdata in order to construct these catchment areas for each of several thousand telephone exchange stations in England. As depicted in Figure 1, neighboring residences on different sides of the boundaries can face substantial differences in the available ICT at their home addresses. This variation stems from jumps in the copper wire distances between residences and their connected exchange stations on the slower (longer distances to connected exchange) relative to the faster side (shorter distances to connected exchange) across a given boundary segment. We exploit this feature in a spatial regression discontinuity (RD) design in which the distance to the nearest boundary segment is the running variable. The identifying assumption is that no other characteristics that affect educational outcomes, except for the available ICT, change discontinuously between neighboring residences when crossing the invisible boundary.

This empirical strategy has a number of appealing features and wide applicability. First, rather than focusing on zero/one variation in first-time broadband access or penetration rates, the design is explicitly targeted at evaluating the large number of currently ongoing policy proposals to upgrade already existing ICT. Given close to universal internet coverage in developed countries, and the vast amounts of public funds currently earmarked for the purpose of increasing available internet connection speeds, this design adds a useful tool that can be applied to a variety of policy questions in addition to the effects on education. Second, the design provides a rich empirical

⁴Telephone exchange stations are also referred to as the central office. Online Appendix Figure A.1 provides an illustration from the UK telecom regulator OfCom.

setting that allows for the flexible estimation of the effect of interest. Rather than exploiting a single discontinuity, we can estimate the effect of jumps in the available ICT on education outcomes across more than 20,000 individual boundary segments in England, with rich variation in both the intensity of the jump, as well as in the levels of ICT on the slower side across boundaries.⁵ Third, because the design exploits cross-sectional variation, it allows us to estimate the medium to long-run effects of ICT upgrades instead of shorter-term adjustments when using variation in the timing of ICT roll-outs.⁶ Finally, the identifying assumption is readily assessable in any one cross-section of data, even when panel data covering the pre-internet period are not available. In our analysis, we first report how point estimates are affected by the inclusion of school fixed effects and observable student characteristics.⁷ We then also test directly to what extent property values, student characteristics as well as local amenities vary when moving from the slow side of a boundary segment to the fast side.

Taking this design to the collected microdata, the empirical analysis proceeds in three steps that exploit different combinations of the data. In the first step, we use the telecom network data to quantify the causal effect of residential distances to their connected exchange stations on the available ICT. In the second step, we then combine information on student test scores and their home addresses with the telecom network database to estimate the causal effect of residential exchange connection distances on educational outcomes, and combine these estimates with the estimation in step 1 to quantify the effect of changes in the available ICT on education outcomes. In the final step, guided by a simple theoretical framework, we use additional geo-referenced survey microdata on student time use and internet use in combination with the telecom network data to quantify the mechanisms underlying the estimated reduced form effect in step 2.

In step 1, we find that the average jump in available ICT across invisible exchange station boundaries is substantial. We estimate that the average difference in residential distances to their connected exchange station between neighboring residences on different sides of the boundary is 725 meters, or 65 percent of a standard deviation. This estimate increases to 2,250 meters, or two standard deviations, when restricting the estimation to the top third of boundary segments with the largest mean difference in residential connection distances across the boundary. These discontinuous jumps in copper wire connection lengths translate into substantial differences in the available ICT across space. We find that the average jump in the time cost of accessing a given amount of online content rises by close to one quarter when moving from the slow side of a boundary segment to the faster side of the invisible boundary. This effect increases to close to one half when restricting attention to the top third of boundary segments with the largest jumps in connection distances. Furthermore, because broadband technology has been updated to less distance sensitive technologies such as coaxial or fiber after 2008, and the available UK internet

⁵This also relates to recent work by Angrist and Rokkanen (forthcoming) on the identification of causal effects away from a single cutoff, and Bertanha (2015) on RD designs with multiple cutoffs.

⁶Especially in the context of educational attainment, where effects may take time to accumulate, this is an important feature. Given broadband coverage in England was already close to universal at the beginning of our sample in 2002, and the fraction of students who move residences during secondary school is only 0.5 percent, the design captures a wide range of length of exposure (>7 years towards the end of the sample). As discussed below, we also further explore the heterogeneity of the effects across years and age groups.

⁷Because school attendance is not directly linked to residential areas in the UK, boundary segment fixed effects and school fixed effects are not collinear.

speed data only start in 2012, we regard these estimates as lower bounds of the effect of connection distances on the available ICT at student residences during the period that we observe their test scores (2002-2008).

In steps 2 and 3, we guide the estimation of the effect of these jumps in the available ICT on education outcomes by writing down a simple theoretical framework that decomposes the observed reduced form effect into several distinct channels. At the broadest level, we distinguish between an effect on student time supply for studying and an effect on student learning productivity per unit of time spent studying.⁸ To fix ideas, we refer to the hypothesis that reductions in the time cost of accessing online content lead to a reduction in student time spent studying as the “Facebook effect”. On the other hand, we refer to the hypothesis that reductions in the time cost of accessing online content lead to an increase in learning productivity as the “MOOC effect”.⁹ Naturally, both the effect on study time supply and on learning productivity could in principle also go in the opposite direction.¹⁰

The main results can be summarized as follows. We find that even substantial changes in the available internet connection speeds have a precisely estimated zero effect on educational attainment, and that the estimates are causally identified: house prices, student socioeconomic characteristics and access to local (dis-)amenities are flat across the boundaries. We further explore the extent to which this average effect may mask significant heterogeneity, and find that the result holds for all age, gender, ethnic and socioeconomic groups in the student population, for all years in the estimation sample 2002-2008, for different subjects (English, Mathematics, Science), across different treatment intensities, and across different initial levels of the available ICT. Using the additional information on student internet and time use to quantify the channels underlying the zero reduced form effect, we find that jumps in the available ICT have no significant effect on student time spent studying online or offline, or on the productivity of time spent studying. Finally, while ICT improvements appear to at least weakly increase student consumption of online content, we find that the elasticity of student demand for online content with respect to its time cost is negative but bounded by -1.

The empirical strategy of this paper is related to a growing empirical literature on the socioeconomic consequences of the internet.¹¹ Bhuller et al. (2013) and Akerman et al. (forthcoming) exploit the timing of the roll-out of broadband internet across Norwegian municipalities to estimate the effect on sex crime and skilled-to-unskilled wages and productivity respectively. Campanante et al. (2013) exploit time variation from the broadband roll-out across Italian municipalities to estimate the effect on political participation, and Bellou (2015) exploits time variation in internet penetration across US states to estimate the effect on marriage rates. Forman et al. (2012) use US county-level measures of infrastructure costs, industry mix, or characteristics of nearby loca-

⁸These channels are very general and have been discussed before. See related literature below and for example Beltran et al. (2010) for a related model on the effect of home computers on education outcomes.

⁹So called massive open online courses (MOOC), such as EdX or the Khan Academy, increasingly target primary and high school students. One of the earliest platforms for educational videos was YouTube, with thousands of channels for homework support. Recent work on MOOCs in economics includes Deming et al. (2015) and Hoxby (2014).

¹⁰For study supply, websites such as Lumosity.com provide online learning tools/games that could make studying more attractive. For productivity, online distractions could in principle also decrease learning per time spent studying.

¹¹Our analysis is also related to existing work in the education literature using RD designs, as for example Angrist and Lavy (1999), Black (1999), Clark (2009) and Gibbons et al. (2013).

tions as instruments for internet investment to estimate the effect on local wages. [Bauernschuster et al. \(2014\)](#) use the fact that some regions in Eastern Germany use telephony networks that are not compatible with DSL-broadband to estimate the effect on social capital. Most closely related to our empirical strategy is [Falck et al. \(2014\)](#) who exploit municipality-level variation in distances to telephone exchange stations for a sample of German municipalities that do not host a telephone exchange station within their own boundary to estimate the effect on political participation.¹²

Relative to existing work, this paper presents a research design with a number of appealing features that we discuss above: It provides a rich empirical setting with tens of thousands of boundaries even in a relatively small country setting such as England; rather than focusing on zero/one variation in broadband access, it is explicitly targeted at the evaluation of currently on-going policy proposals to provide faster internet connections in a setting with close to universal broadband access; it exploits cross-sectional variation that can be applied to estimate medium to long-run effects; and the identifying assumption is readily assessable in any one cross-section of data, even when panel data covering the pre-internet period are not available.

Thematically, the paper relates to an existing literature on the effects of home computer use on education outcomes.¹³ This literature has found some mixed results. On one hand, [Fairlie \(2005\)](#) and [Beltran et al. \(2010\)](#) for the US, [Schmitt and Wadsworth \(2006\)](#) for the UK, and [Fiorini \(2010\)](#) for Australia have documented positive associations between home computers and educational outcomes. On the other hand, [Fuchs and Wossmann \(2004\)](#) who use data from 31 countries, and [Vigdor et al. \(2014\)](#) who exploit student panel data in North Carolina, find evidence for negative effects.¹⁴ In addition to effects of home computers, [Vigdor et al. \(2014\)](#) also find negative associations between expansions in broadband availability at the ZIP code level and student maths and reading test scores in middle grades as well as reported computer use for homework. One difference of this study is that we trace the available ICT at a much finer geographical scale: the average ZIP code in North Carolina contains more than 10,000 people, whereas postcode units in England are equivalent to a single building in cities or segments of the same street in less urban areas and capture on average about 15 households. This level of detail allows us to exploit discontinuous changes in the available ICT across neighboring residences in a spatial RD design.

In this literature we relate most closely to a number of recent studies that examine the effects of home computer use on education outcomes using randomized controlled trials (RCTs) or quasi-experimental variation. [Malamud and Pop-Eleches \(2011\)](#) find negative effects on school grades but positive effects on computer skills and on a separate measure of cognitive ability. To estimate these effects, they exploit a regression discontinuity from a voucher policy that subsidized computers for low-income households in Romania. [Fairlie and London \(2012\)](#) find modestly positive effects on education outcomes from an RCT that provided free computers to lower-income college

¹²In concurrent and independent work on the effect of ICT on house prices [Ahlfeldt et al. \(2015\)](#) also use information on exchange catchment areas. In contrast to the spatial RD design that we propose here, their approach conditions on cross-sectional variation in residential distances to the connected exchanges, and instead exploits variation in the timing of broadband upgrades across exchanges and/or households.

¹³Another recent literature examines effects of ICT in schools (e.g. [Angrist and Lavy 2002](#); [Barrow et al. 2009](#); [Goolsbee and Guryan 2006](#); [Machin et al. 2007](#); [Falck et al. 2015](#)). In contrast, our results abstract from school-level factors by including school fixed effects in our main specifications.

¹⁴See [Bulman and Fairlie \(forthcoming\)](#) for a comprehensive review of this literature.

students in California. More recently, Fairlie and Robinson (2013) present evidence from a similar RCT that randomly assigned home computers to more than one thousand secondary school students in California who did not previously have access to PCs, and find no evidence for effects on test scores or homework time.

This paper aims to shed light on a related but different question of policy interest. Motivated by the recent wave of government proposals to invest in ICT upgrades for providing ever faster broadband internet, our analysis investigates the likely effect of such policies on education outcomes. Furthermore, by empirically linking administrative test score records to the available ICT at student home addresses, we are able to estimate the effect of ICT improvements on education outcomes in the population of primary and secondary school students, rather than restricting attention to students without prior access to ICT. This also allows us to explore the heterogeneity of these effects along multiple dimensions of the student population.

The remainder of the paper proceeds as follows. Section 2 describes the background and data. Section 3 presents the theoretical framework that guides the empirical analysis. Section 4 presents the empirical strategy. Section 5 presents estimation results on the effect of jumps in the available ICT on education outcomes. Section 6 presents additional results to quantify the underlying channels. Section 7 concludes.

2 Background and Data

This section provides a brief description of the background and the main datasets used in the empirical analysis. Table 1 presents descriptive statistics for each of the estimation samples.

2.1 Administrative Student Records and the English Education Context

The English education curriculum is organized around five Key Stages (KS) at ages seven, eleven, fourteen, sixteen and eighteen, corresponding to the grades –“academic years” in England– two, six, nine, eleven and thirteen. Generally, students transfer from primary to secondary school after completing grade six, where they stay at least until the end of compulsory education in grade eleven. Towards the end of each of these Key Stages students are assessed in national exams that are either administered by their teachers (KS1), or through externally marked tests (KS2, KS3, KS4 and KS5).¹⁵ There is no grade skipping or repeating and compulsory education ends at age sixteen with the fourth Key Stage, which is also called “GCSE”. Until and including KS3 at age fourteen, students have very limited options in choosing subjects or specializing according to interest or ability. Consequently, students are tested in the three main compulsory subjects English, Mathematics and Science in the national and externally assessed KS2 and KS3 tests. After KS3 students are able to specialize and choose among a large number of possible subjects.

We obtain access to the administrative datasets covering the population of English students which allow us to observe education outcomes for Key Stages 1-4 alongside their residential addresses. To match the student information to the telecom network data that we describe below, we

¹⁵External KS3 tests were abolished in 2009.

first extract the full (up to eight characters) residential postcode for each registered student in a given year from the National Pupil Database (NPD). Full postcodes in England on average include about 15 households and can be thought of as identifying a single building in cities or identifying a segment of a street in less urban areas. In the second step, we use the unique student identifiers to link this information to individual test score results, which are also provided as part of the NPD.¹⁶

As described in the next paragraphs, our main estimation sample are observations from postcodes within 1 km distance of an exchange station boundary segment.¹⁷ This covers slightly more than 7 million test score observations for over 4.5 million students living in over 450,000 postcodes in England over the period 2002-2008. These students attend over 20,000 primary and secondary schools. Following the education literature, we percentilize the national test scores by Key Stage exam, subject and cohort in order to make them comparable to (national) GPAs in the US. These subject-percentiles are then added into a total score, which we percentilize to obtain an average total score ranging from 1 to 100. In addition to test scores, the administrative data give us access to a series of observable student characteristics, such as gender, age, ethnicity (eight groups), and whether the student is eligible for a free school meal. Table 1 provides descriptive statistics of the student estimation sample in the second panel.

2.2 Telecom Network Data

The roll-out of broadband technology in the UK started in the major urban centers at the end of the 1990s and proceeded rapidly. By the start of our estimation period in the year 2002, more than 90% of English households had access to broadband enabled telephone exchange stations. A salient feature of DSL-broadband technology is that the available connection speed depends on the distance of the copper wire connection between the router in the residence and the local telephone exchange station.¹⁸

As a new strategy for exploiting this technological feature empirically, we obtain information about the location of all English telephone exchange stations (roughly 3,900), and link the exchange station identifiers to telecom network records that provide us with information on each English postcode's assignment (in total, there are slightly more than 1.46 million English postcodes) to one of the 3,900 exchange stations.¹⁹ This geo-referenced dataset allows us to construct exchange station-level catchment areas and their usually unobserved exact boundaries using a geographical information system (GIS).²⁰ The main estimation sample consists of more than 580,000 individual

¹⁶The Department for Education 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 about seven percent of the English student population.

¹⁷Distance is defined as the minimum Euclidean (straight line) distance between the centroid of a full postcode and the nearest boundary segment. As will become apparent from the non-parametric estimation results of the boundary discontinuity effect, none of the presented findings are sensitive to increasing or decreasing this 1 km sample threshold. For completeness, we also report estimation results for a wider 2 km distance band around the boundaries covering more than 97 percent of the student population in England, as discussed in Section 5.

¹⁸Online Appendix Figure A.1 provides an illustration from the UK telecom regulator OfCom.

¹⁹We are grateful to the technology consulting firm Point Topic Ltd for providing us access to their proprietary databases.

²⁰We use exact vector (polygon) data from the Ordnance Survey.

postcodes that are within 1 km of an exchange catchment area boundary.²¹ The left panel of Figure 1 provides a map of these catchment areas in England, and Table 1 shows descriptive statistics using the linked postcode-exchange station geographical dataset.

The location of telephone exchange stations was determined during the roll-out of the English land line telephony network which mainly occurred before and during the Second World War. Importantly, distance to the local exchange station did not affect the quality of traditional telephone services. However, while we believe it to be unlikely that households in the past or present actively sort on the basis of distances to the connected exchange station, there are several reasons to believe that exchange station location is far from random and may be correlated with other local neighborhood characteristics that do matter for household sorting. For example, telephone exchange 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. To address a series of endogeneity concerns that arise in the context of this “passive sorting”, we notice that capacity constraints for any given exchange station coupled with the rapid growth of land line telephony subscriptions after the Second World War have given rise to an empirical setting in which the invisible boundaries of exchange station-level catchment areas are essentially randomly drawn on a map. These boundaries give rise to substantial cross-sectional variation in the available ICT due to discontinuous jumps in the length of the copper wire that connects residences on either side of the boundary to their assigned exchange station. The right panel of Figure 1 provides an example of the boundary placements from one of the author’s residential areas in London. Within the catchment areas we can also see the fine geographical resolution of the English postcodes, which contain about 15 households on average and are often as small as a single building.

In addition to the linked postcode-exchange station telecom network data, we also obtain microdata on the available internet speed from the UK’s telecommunication regulator (OfCom) across roughly 1.3 million English postcodes during the years 2012 and 2013. This is important as it allows us to not just rely on the theoretical or laboratory-tested relationship between copper wire lengths and the available internet speed depicted in Online Appendix Figure A.1. This dataset provides us with the average, maximum, and minimum speed recorded across individual households that we can geo-locate at the level of their full residential postcode. We use this information to link the speed microdata to the telecom network GIS database discussed above in order to estimate the relationship between residential exchange distances and available internet speed in England. One important caveat of these microdata on internet speed is that they significantly postdate our student test score observations between 2002-2008. In particular, broadband internet infrastructure in the UK was updated to significantly less distance sensitive technologies after 2008 such as coaxial or fiber, so that we regard our estimates of the effect of crossing a 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). Table 1 provides additional descriptive statistics about this dataset.

²¹See footnote 17.

2.3 Survey Microdata on Student Time Use and Internet Use

To further investigate the channels that underlie the reduced form effect of variation in the available ICT at student home addresses on educational outcomes, we obtain access to the restricted-use version of the microdata of the British Household Panel Survey (BHPS) covering the years 2002-2010 and its successor, the Understanding Society (US) Survey covering the years 2011 to 2013.²² We complement these data with the restricted-use versions of the first and second waves of the Longitudinal Study of Young People in England (LSYPE) covering the years 2004 and 2005.

The confidential data columns in these datasets allow us to link the survey microdata on student time use and internet use to the residential postcode that the surveyed household resides in. Each round of the BHPS data covers around 15,000 geo-referenced households, with an increase in the number of surveyed households in the US successor surveys to around 25,000 in each round. The LSYPE is a separate panel survey that includes about 14,000 students in each round. In the first wave, the sampled students of the LSYPE were age 14 in 2004, and one year older in the second wave in 2005. As for the other datasets, we define our main estimation samples as observations within residential postcodes that are within a 1 km distance from an exchange station catchment area boundary.

Taken together, the youth surveys that are included in the BHPS and the US in combination with the survey data from the LSYPE provide us with geo-referenced information on student time spent on homework broken up into whether this is online or offline, as well as time spent using computers or the internet more generally, and coincide with our test score estimation sample. Descriptive statistics for the BHPS, US and LSYPE estimation samples from these datasets are shown in Table 1. English students during this period on average spent about 9 hours per week using the internet for email and online social media and about 5.5 hours per week for doing homework. And more than three quarters of students report using online resources for doing their homework.

2.4 Data on House Prices and Access to Local (Dis-)Amenities

We use transaction-level data on property sales in England over the estimation period 2002-2008. The data are administrative records from the English Land Registry and report the universe of property transactions during the estimation period. We use the reported property address information in order to link these property transaction values to individual residential postcodes. The estimations are based on several millions of individual property transactions that occurred in England over the period 2002-2008. In addition to property values, we also observe property characteristics, including the type of property (e.g. 2 bedroom house versus studio), whether it is a new building, and the ownership type (lease or free hold).

We complement the house price microdata with a number of direct measures for local amenities and dis-amenities. Even though our empirical analysis is based on a spatial discontinuity design that compares only very proximate outcomes, 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 have a higher likelihood to host a given type of local (dis-)amenity–

²²We use the answers from the Youth Questionnaire (children aged 11-15) which is available in both surveys.

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 company CBRE and the English Department for Education, we compute Euclidean distances between each English post-code and the following features: nearest school (primary or secondary), nearest road (class A, B and motorways) and nearest supermarket.²³ As for the other datasets, we define our main estimation samples as observations within residential postcodes that are within a 1 km distance from an exchange station catchment area boundary. Descriptive statistics in kilometers for these variables are shown in the lower panel of Table 1.

3 Theoretical Framework

This section presents a simple model to guide the empirical estimation. We analyze the effect of changes in the time cost of accessing online information and communication on learning outcomes through the lens of a basic knowledge production function. Improvements in the available ICT can affect learning outcomes through two mechanisms: either through changing the productivity of a given amount of time spent studying, or by affecting the supply of time spent studying relative to other student activities. In any given period, a student i 's knowledge production function is given by:

$$H_i = A_i L_i^\alpha \quad (1)$$

where H_i is educational achievement, A_i is a student-specific learning productivity shifter, L_i is the student supply of time spent studying, and $\alpha > 0$ is the elasticity of learning achievement with respect to time spent studying. In the following, we let both the productivity shifter A_i and the student study time supply L_i be functions of student-specific characteristics (λ_i^A and λ_i^L), neighborhood characteristics (ϕ_n^A and ϕ_n^L), school characteristics (μ_s^A and μ_s^L), and a residual term (ϵ_{ins}^A and ϵ_{ins}^L).

We introduce the available ICT as the inverse of the per-unit time cost (c_i) of accessing online content: $\frac{1}{c_i}$. We allow this inverse cost to affect both student productivity and the supply of time spent studying:

$$\begin{aligned} A_{ins} &= \left(\frac{1}{c_i}\right)^\delta \lambda_i^A \phi_n^A \mu_s^A e^{\epsilon_{ins}^A} \\ L_{ins} &= \left(\frac{1}{c_i}\right)^\eta \lambda_i^L \phi_n^L \mu_s^L e^{\epsilon_{ins}^L} \end{aligned} \quad (2)$$

where δ is the elasticity of learning productivity with respect to available internet connection speed $\frac{1}{c_i}$, and η is the elasticity of study time supply with respect to $\frac{1}{c_i}$. The first elasticity captures the potential effect of changes in the time cost of accessing online content on the learning productivity for a given amount of time spent studying. The second elasticity captures the potential

²³Data on supermarket locations are for the first year of our test score analysis (2002), and we obtain them from the Retail Research Unit of the real estate advisory company CBRE in the UK. Supermarkets refer to all establishments in the category "grocery/supermarket".

effect of student time allocation between studying and other activities. Analogous to a basic labor supply equation, η can capture both a relative price effect whereby changes in $\frac{1}{c_i}$ affect the relative attractiveness of studying compared to other activities online or offline, as well as an endowment effect whereby changes in $\frac{1}{c_i}$ affect study time supply by endowing the student with more or less online content for a given amount of time spent studying.

Substituting (2) into (1) and taking logs we get the following estimation equation:

$$\ln H_{ins} = \beta \ln \left(\frac{1}{c_i} \right) + \phi_n + \mu_s + \lambda_i + \epsilon_{ins} \quad (3)$$

where $\beta = (\delta + \alpha\eta)$, $\phi_n = \ln \phi_n^A + \alpha \ln \phi_n^L$, $\mu_s = \ln \mu_s^A + \alpha \ln \mu_s^L$, $\lambda_i = \ln \lambda_i^A + \alpha \ln \lambda_i^L$, and $\epsilon_{ins} = \epsilon_{ins}^A + \alpha \epsilon_{ins}^L$. This basic theoretical setting gives rise to several interesting hypotheses for empirical estimation. To fix ideas, we refer to the hypothesis that $\eta < 0$ as the “Facebook effect”. That is, lower information and communication costs decrease effective study time supply by either making other online or offline activities relatively more attractive (positive relative price effect) and/or by letting students access a fixed amount of learning content in less time (endowment effect). For learning productivity, we refer to the hypothesis that $\delta > 0$ as the “MOOC effect”.²⁴ That is, lower information and communication costs increase the productivity of a given amount of study time for educational achievement. Of course, both the effect on study supply and learning productivity can in principle also go in the opposite direction. Websites such as Lumosity.com offer online learning and brain-training games that could increase the perceived attractiveness of learning relative to other activities (negative relative price effect) to an extent that outweighs the endowment effect ($\eta > 0$). For productivity, online distractions could in principle also decrease learning per unit of time spent studying ($\delta < 0$).

The empirical analysis in Section 5 (steps 1 and 2 below) combines the administrative microdata on English student test score results with the exchange station boundary discontinuity design to estimate the reduced form causal effect of discontinuous jumps in the available ICT on education outcomes: $\beta = \delta + \alpha\eta$. The empirical analysis in Section 6 (step 3 below) then brings to bear additional microdata on student time use and internet use to further disentangle and quantify the underlying mechanisms.

4 Empirical Strategy

In the empirical analysis the objective is to estimate the effect of exogenous variation in the available ICT ($\ln \left(\frac{1}{c_i} \right)$) on a number of different outcomes including student test score performances in national exams. Our empirical analysis proceeds in three main steps that exploit different combinations of the microdata. In step 1 we estimate the causal effect of student home distances to connected telephone exchange stations on the available ICT at the students’ residences.

²⁴So called massive open online courses (MOOC) provide educational video streaming. They have grown rapidly and increasingly target primary and high school students in addition to post-secondary education. One of the earliest platforms for educational videos was YouTube which features thousands of educational channels for homework support. Even before the take-over by Google in 2006, YouTube had over 100 million daily video views in the UK and over 65,000 daily uploads according to press articles at the time.

In step 2 we then estimate the causal effect of residential distances to exchange stations on student test score performances, and use these estimates in combination with those from step 1 to quantify the effect of changes in available internet speed on educational outcomes (the $\beta = \delta + \alpha\eta$ in Section 3 above).²⁵ In the final step 3, we use the theoretical framework of the previous section in combination with additional geo-referenced survey microdata on student time use and internet use to quantify the potentially opposing mechanisms underlying the reduced form effects in step 2.

The identifying assumption to estimate causal effects in specification (3) is that, conditional on disaggregated neighborhood and school fixed effects, variation in $\ln\left(\frac{1}{c_i}\right)$ is orthogonal to the error term. That is, variation in available residential ICT is uncorrelated with unobserved student characteristics or locational factors that also affect test scores. To this end, we propose an empirical strategy that is based on comparing student outcomes across exchange station catchment area boundaries. We estimate variations of the following two alternative specifications:

$$y_{ipnkt} = \beta FastSide_p + \sum_{\tau=1}^4 \left(\beta_{\tau}^{slow} DistBoundary_p^{\tau} + \beta_{\tau}^{fast} FastSide_p * DistBoundary_p^{\tau} \right) + \gamma X_{it} + \phi_n + \mu_s + \kappa_{kt} + \epsilon_{ipnkt} \quad (4)$$

$$y_{ipnkt} = \sum_{b=1}^{10} \left(\beta_b^{slow} I(DistBoundary_p = b) \right) + \sum_{b=1}^9 \left(\beta_b^{fast} FastSide_p * I(DistBoundary_p = b) \right) + \gamma X_{it} + \phi_n + \mu_s + \kappa_{kt} + \epsilon_{ipnkt} \quad (5)$$

where i indexes students, p the full postcode where the student lives, n neighborhoods, s schools, k Key Stage exam types, and t test score years. The estimation sample are roughly 580 thousand residential postcodes that are within 1 km of an exchange station boundary as discussed in the data section. As discussed in the next section, we also report estimation results for a wider 2 km distance band around the boundaries covering more than 97 percent of the student population in England.

We define neighborhoods as a group of residential postcodes that are nearest to the same boundary segment. To ensure that we define neighborhoods at a disaggregated spatial resolution that extends on both sides of a given boundary segment, we partition each exchange station boundary outline into individual segments by overlaying a randomly placed “cookie cutter” grid of 1x1 km squares covering the land area of England in order to split and define individual boundary segments (i.e. the maximum length is the square’s diagonal ≈ 1.4 km). This definition of

²⁵This is analogous to a two-stage instrumental variable estimation where the first step represents the first stage and the second step the reduced form, which are then used to quantify the second stage. We approach the estimation in separate steps to use the universe of available information for each of them, as they are of interest in their own right. Having said this, the results are virtually identical when restricting the first-stage estimation and the reduced form to a sample of over-lapping postcodes, which is not surprising as our second stage uses the universe of postcodes containing student residences. For sample comparisons also see Table 1. Finally, note that we refer to “reduced form effects” in the text implying a model-based definition (i.e. β is a combination of structural parameters) rather than the econometric definition (where step 2 is the reduced form of the IV estimation).

local neighborhoods leads to the inclusion of roughly 20,000 boundary segment fixed effects, ϕ_n , pertaining to approximately 3,900 exchange station catchment areas.²⁶ X_{it} are observable student socioeconomic controls that we obtain from the administrative school records: age, gender, free school meal eligibility, and student ethnicity.²⁷ In some of our specifications discussed below, we also include school fixed effects μ_s , which is feasible because residential location is not deterministically linked to school attendance in England. Finally, year-by-Key Stage fixed effects κ_{kt} allow for any potential differences in testing or annual shocks. To address potential autocorrelation in the error term, ϵ_{ipnskt} , across residences within the same local neighborhood, we cluster standard errors at the level of individual boundary segments.

Turning to the explanatory variables of interest in the first rows of specifications (4) and (5), we follow two different estimation approaches. The first approach in (4) is to fit a fourth-order polynomial function on either side of the boundary with respect to the running variable of the discontinuity design: residential distances to the nearest exchange station boundary segment. We define the fast and slow sides of each individual boundary segment in terms of the difference in the mean of residential distances to their connected exchange station on either side of the same boundary segment.²⁸ The causal effect of the boundary discontinuity on an outcome variable of interest, y_{ipnskt} , is thus estimated using only the discontinuous jump in levels that occurs when crossing a given boundary segment from the slower side to the faster side, captured by $\beta_{FastSide_p}$, after conditioning on the flexible functional form with respect to boundary distance on either side in addition to neighborhood fixed effects, school fixed effects and a rich set of socioeconomic student controls. As discussed in Section 2, the variation that this design exploits are discontinuous jumps in the length of the copper wire that connect residences on either side of the invisible boundary to their assigned telephone exchange stations. The identifying assumption is that other factors that affect student test score performances vary smoothly across the boundary, which can in principle be tested on observables as we discuss below.

To address potential concerns about the parametric assumptions imposed by specification (4) (e.g. [Gelman and Imbens, 2014](#)), we exploit the richness of the collected data to also estimate a less parametric specification. To empirically assess how well the fourth-order polynomial functional form captures the relationship between exchange station distances and the outcomes of interest, the first row of specification (5) replaces the two fourth-order polynomials with 19 100-meter distance bins, where the reference category (the 20th bin) are residential postcodes within 900-1,000 meter distance from the boundary segment that are located on the fast side (i.e. these are the residences with the shortest distances to their connected exchange on the faster side). I is an indicator function that takes the value of one if a student's residential postcode p is within a b distance bin from the boundary, of which there are 10 on each side of the boundary in total. In specification

²⁶Our estimation sample are boundary segments with nearest postcodes assigned on both sides (from two different exchange station catchment areas). Some boundary segments along the coast have nearest residences assigned from only one exchange station catchment area, and 0.38 percent of boundary segments had nearest residences assigned from more than two exchange station catchment areas. We exclude these latter two cases.

²⁷Since there is no skipping or repeating, the effect of age is already controlled for by the test type-by-year fixed effects (κ_{kt}).

²⁸To define this consistently for each boundary segment as well as over time, we use the universe of residential postcodes to define the slow and fast sides for each boundary segment.

(5), we label the b 's to be increasing in distance to the boundary on either side. Overlaying the point estimates of this less parametric specification onto the estimation results of specification (4) allows us to assess to what extent the polynomial functional forms capture the actual relationship between boundary distances and various outcomes of interest that we consider in the following sections. As discussed in the next section, we also report additional results after replacing the fourth-order polynomials with 39 50 m bins instead of the 100 m bins.

In addition to fitting flexible functional forms, the empirical design in specifications (4) and (5) offers three advantages. The first is that we can exploit variation in ICT in the cross-section of students, which allows us to estimate the medium to long-run effects of ICT upgrades rather than shorter-term adjustments.²⁹ Especially in the context of educational attainment, where effects may take time to accumulate, this is an important feature. Given broadband coverage in England was already close to universal at the beginning of our sample in 2002, and the fraction of students who move residences during secondary school is only 0.5 percent, the design captures a wide range of length of exposure (more than 7 years towards the end of the sample).³⁰ Second, we can provide direct empirical evidence on the validity of the identifying assumption: We report how point estimates are affected by the inclusion of a rich set of additional observable control covariates, such as school fixed effects and student socioeconomic characteristics (that is before and after including X_{it} above). Furthermore, we can test to what extent property values as well as local amenities vary when moving from the slow side of a boundary segment to the fast side. Third, we can estimate the effect of changes in available connection speeds on educational outcomes across a wide range of treatment intensities/doses, as well as across different initial positions in the distribution of available ICT (different levels of ICT on the slower side). That is, rather than exploiting a single discontinuity, the exchange station boundary design allows us to estimate the effect of ICT on education across more than 20,000 individual boundary segments, with rich variation in both the intensity of the jump in available ICT as well as initial positions in the ICT distribution on the slow side of different segments.

5 Estimation

This section combines the microdata described in Section 2 with the empirical strategy outlined in Section 4 to estimate the reduced form causal effect of ICT improvements for educational attainment ($\beta = \delta + \alpha\eta$) that we derived in Section 3.

5.1 The Boundary Effect on Connection Distances and Available ICT

This subsection uses the telecom network data to estimate the size and statistical significance of discontinuous jumps in the available ICT across exchange station catchment boundaries (step 1).

²⁹On a more subtle note, this also allows us to disentangle the effect of ICT on the outcome of interest while holding constant the economic environment. For example, when using time variation, connecting an area to the internet may have general equilibrium effects on the returns to skills that could also affect student effort. In contrast, our research design tests for the effect of ICT improvements while holding constant the economic environment for students on either side of the invisible exchange station boundary.

³⁰As discussed below, we also further explore the heterogeneity of the estimated effects across years and age groups.

The two key pieces of information that we exploit in these estimations are data on the assignment of each of the roughly 1.5 million residential postcodes in England to one of several thousand telephone exchange stations in combination with additional microdata on realized internet connection speeds across residential postcodes.

To estimate the average effect of crossing an exchange station boundary on the residential distance to the connected exchange switchboard, we estimate the following specification:

$$DistExchange_{pn} = \beta FastSide_p + \sum_{\tau=1}^4 \left(\beta_{\tau}^{slow} DistBoundary_p^{\tau} + \beta_{\tau}^{fast} FastSide_p * DistBoundary_p^{\tau} \right) + \phi_n + \epsilon_{pn} \quad (6)$$

where p indexes individual residential postcodes and n indexes neighborhoods defined as boundary segments within 1x1 km squares as described above. $DistExchange_{pn}$ is the distance in meters between residential postcode centroids and their connected telephone exchange station. Alternatively, we replace the outcome variable of interest by $\ln(Speed_{pnt})$, which is the logarithm of the average realized connection speed in residential postcode p located in neighborhood n and year t , and include year fixed effects in addition to the neighborhood fixed effects in specification (6). As discussed above, we also estimate these two specifications after replacing the polynomial functional forms with the less parametric version in specification (5) in order to assess the fit of the functional forms in specification (6) on either side of the boundary.

Figure 2 and Table 2 report the estimation results of the effect of boundary discontinuities on residential connection distances. The the left hand side (to the left of the vertical line and with negative distances on the x-axis) in each of the graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each graph depicts estimation results from two separate specifications: The plotted functional forms present results after estimating the fourth-order polynomials conditional on neighborhood (boundary segment) fixed effects. The shaded area depicts 95% confidence intervals. The black bars on the other hand depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means, which also implies that the 20th reference bin is not at zero in these graphs. The confidence intervals are based on standard errors that are clustered at the level of boundary segments.

Several findings emerge. The fourth-order polynomials appear to capture the functional form of exchange distances with respect to boundary distances on either side of the boundary extremely well: The less parametric point estimates on the 20 100m location bins trace out the smooth polynomial function almost exactly. The average jump in residential distances to the connected exchange is substantial: 725m, or 65 percent of a standard deviation in connection distances according to the descriptive statistics in Table 1. Restricting attention to the top 30 percent of the 20,000 English boundary segments that are characterized by the largest jumps in mean residential connection dis-

tances, this average effect increases from 725m to close to 2,250m, or two standard deviations. In conclusion, the boundary discontinuity design appears to provide substantial and discontinuous variation in residential distances to the connected telephone exchange stations, as well as a wide range of different treatment intensities across the more than 20,000 boundary segments.

In Figure 3, we replace residential connection distances with log average realized internet connection speeds on the left hand side of specification (6). As suggested by the technological relationship in Online Appendix Figure A.1, the jump in connection distances translates into substantial variation in the available internet connection speeds on different sides of the boundary. The average jump in connection speed is 22 percent across all boundary segments. When restricting the estimation to the top 30 percent of boundary segments with the highest mean differences in connection distances across the boundary, this average jump increases to 47 percent. These effects suggest that a household on the slower side of an invisible catchment area boundary incurs 20 or 50 percent more time in order to access the same amount of online content as a neighboring household on the faster side of the boundary. In more practical terms, OfCom –the UK telecom regulator– reports that in 2008 (the end of our student test score data below) more than one quarter of English internet connections still delivered actually observed *maximum* speeds below 1.8 Mbit/s, which would be significantly below common requirements for many online activities including video streaming.³¹ At such levels, the documented jumps in the available ICT across exchange boundaries give rise to significant variation in the time cost of accessing online content. Related to this, and as discussed in the data section, we note that these estimates of the effect on connection speeds are likely to be lower bounds for the estimation period of our test score analysis. The reason is that the available speed data cover a period (2012-2013) with significantly less distance sensitive broadband technology compared to what was in place during our estimation period 2002-2008 in what follows below.

5.2 The Boundary Effect on Student Test Scores

After estimating the boundary effect on jumps in residential connection distances and the available ICT, we proceed to our second step, and estimate the effect of exchange station boundaries on student educational attainment:

$$\begin{aligned} TestScore_{ipnkt} = & \beta FastSide_p + \sum_{\tau=1}^4 \left(\beta_{\tau}^{slow} DistBoundary_p^{\tau} + \beta_{\tau}^{fast} FastSide_p * DistBoundary_p^{\tau} \right) \quad (7) \\ & + \gamma X_{it} + \phi_n + \mu_s + \kappa_{kt} + \epsilon_{ipnkt} \end{aligned}$$

where i indexes students, p postcodes, n neighborhoods as defined above, s schools, k Key Stage exam types, and t test score years. As described in Section 2, $TestScore_{ipnkt}$ is the student's national test score percentile in Key-Stage exams. We report results both for a baseline specification that includes neighborhood fixed effects (ϕ_n) and test type-by-year fixed effects (κ_{kt}), as well as after additionally including controls for student socio-economic characteristics (X_{it}) and school

³¹See OfCom Broadband Report 2008. Video streaming requires stable speeds of around 3 Mbit/s.

fixed effects (μ_s) in what we refer to as the full specification in the figures. The observable student characteristics from the administrative records are age and dummies for gender, free school meal eligibility and eight different groups of ethnic background.³² Once again, we also estimate specification (7) after replacing the polynomial functions with the less parametric specification in (5) to estimate the effect across 100m distance bins. In addition, we also estimate specification (7) both for all English boundary segments as well as after restricting attention to the 30 percent of boundary segments with the highest mean differences in residential connections distances within 1 km on both sides.³³

Figure 4 and Table 2 report the results of these eight estimations. In addition, the left panel of Figure 5 reports results after extending the estimation sample to 2 km on either side of the boundary (capturing more than 97% of the student population), and the right panel of Figure 5 after replacing the 100 m bins by more granular 50 m bins instead.

Several findings emerge. First, the effect of variation in the available ICT on test scores is a precisely estimated zero effect.³⁴ Second, this result holds both before and after including additional student controls and school fixed effects in the full specification. Third, even for substantial variation in available internet connection speeds, the effect is a precisely estimated zero as depicted in the lower panels of Figure 4. Fourth, the estimated level and curvature of the distance relationships on either side of the boundaries are interesting in their own right. In the full 2 km sample reported in the left panel of Figure 5 test scores within 2 km of a given boundary segment are on average statistically significantly lower on the slow (left) side compared to the fast side of the boundary when estimated in terms of simple mean differences conditional on covariates. Furthermore, even within a given exchange catchment area on either side of the boundary, test scores are significantly correlated with available internet connection speeds. In Figure 5, test scores are at first rising as we move away from the boundary on both sides, and then falling especially on the slow side of the boundary segment. These results point to granular and unobserved locational factors and household sorting that are correlated with distances to the connected exchange stations both across catchment areas as well as within. As depicted in Figures 4 and 5, the fact that even substantial discontinuous jumps in available ICT at the boundary do not have a significant effect on test scores leads to this conclusion in a clear and transparent way.

Finally, how do our estimated effect sizes compare to existing findings in the education literature? In our main specification also reported in Table 2, we can reject positive effects greater than 1.9 percent and negative effects greater than 0.6 percent of a standard deviation in test scores from crossing the invisible boundary to the fast side.³⁵ Focusing on the top 30 percent of boundaries in terms of speed jumps, we can still reject small positive (negative) effects of 4.2 (-1.7) percent of a standard deviation in test scores. In comparison, for example Malamud and Pop-Eleches (2011)

³²Since there is no skipping or repeating, the effect of age is already controlled for by the test type-by-year fixed effects (κ_{kt}).

³³As part of the next subsection, we also estimate results across four quartiles of the intensity of jumps across boundaries.

³⁴In reference to footnote 25, note that the estimated zero point estimate for the reduced form of a two-stage IV estimation also implies that the 2nd stage estimate is a zero.

³⁵As discussed in the previous subsection, this treatment is associated with a (lower bound) increase in available connection speeds of roughly one quarter.

find that winning a voucher to subsidize computer ownership has negative effects of 25-30 percent of a standard deviation on school grades in Math, English and Romanian. Fairlie and London (2012) estimate positive effects of 0.14 standard deviations in a summary index of educational outcomes using experimental variation. Fairlie and Robinson (2013) report insignificant effects on academic test scores from a random provision of home computers to students in California, rejecting positive (negative) effects greater than 8 (-4) percent of a standard deviation in GPAs.³⁶ Finally, a teacher who is one standard deviation better than average improves student test scores by 10-20 percent of a standard deviation (Aaronson et al., 2007; Rivkin et al., 2005). Viewed in this context, we would argue that the richness of the empirical setting allows us to estimate reasonably precisely estimated (zero) effects.

5.3 Heterogeneity of the Test Score Effect

The estimated average effect reported above may mask significant heterogeneity. One important consideration is that the effect of differences in the available ICT across student residences may take time to accumulate and materialize into test score outcomes, or that skill formation may be more or less sensitive to ICT at certain stages/ages (e.g. Cunha et al., 2010). While our empirical strategy exploits cross-sectional variation precisely with the aim to estimate longer-term effects (and broadband was close to universally available in England by the start of our seven-year study period in 2002), it could still be the case that the effects only start materializing after an extended period of exposure. If this is the case, one would expect the effect to become stronger over time between 2002-2008.³⁷ One might also expect to see on average larger effects among older student cohorts who have been exposed to variation in ICT for longer and/or make more intensive use of online educational resources (OfCom, 2012). On the other hand, it could also be the case that students at younger ages are particularly sensitive to changes in their learning environment.

To further explore these possibilities, we report to what extent the effect of jumps in the available ICT differs across survey years between 2002-08 as well as across four student age groups (7, 10, 14, and 16 year olds). We also exploit the richness of the empirical setting to explore a number of other potential sources of heterogeneity. To this end, we estimate specification (7) separately across two student gender groups, two income groups (eligibility for free school meals), two ethnic groups (White and Non-White), three subjects (English, Mathematics, Science), four quartiles of treatment intensities at different boundaries, as well as across three groups of mean residential connection distances on the slow side of each segment. One of the strengths of the exchange station boundary discontinuity design is that it allows us to explore the effect of interest across a range of different initial positions of the available ICT on the slower side of the boundary. This is in contrast to typical RD designs that focus on one local effect. In our setting, this allows us to explore potential non-linearities due to, for example, speed thresholds for certain activities such as video streaming.

³⁶Based on the reported confidence intervals and the standard deviation of GPAs in Table 3 of the paper.

³⁷Note that the fraction of students who move residential address during secondary school is about 0.5 percent. Even around the compulsory primary to secondary school transition at age 11, only less than 2 percent of students change residence.

We estimate these specifications and report the results across the different specifications in exactly the same way as we do for the average effect reported in Figure 4. As reported in Figures A.2-A.7 in the Online Appendix, we find that the estimated zero average effect that we report above holds in each of these 24 sub-samples of the database. These additional results provide further support that the cross-sectional variation that we exploit in our empirical strategy appears to effectively capture longer-term effects (of more than seven years at the end of our sample), and that the estimated average effect does not appear to mask significant heterogeneities across either different groups of the student population or non-linearities with respect to initial levels of ICT or the intensity of the jump.

5.4 Robustness

To further assess the validity of the identification strategy we run the following specifications:

$$\ln(\text{PropValue})_{jpnt} = \beta \text{FastSide}_p + \sum_{\tau=1}^4 \left(\beta_{\tau}^{\text{slow}} \text{DistBoundary}_p^{\tau} + \beta_{\tau}^{\text{fast}} \text{FastSide}_p * \text{DistBoundary}_p^{\tau} \right) + \gamma X_{jt} + \phi_n + \kappa_t + \epsilon_{jpnt} \quad (8)$$

$$\text{StudentChar}_{ipnkt} = \beta \text{FastSide}_p + \sum_{\tau=1}^4 \left(\beta_{\tau}^{\text{slow}} \text{DistBoundary}_p^{\tau} + \beta_{\tau}^{\text{fast}} \text{FastSide}_p * \text{DistBoundary}_p^{\tau} \right) + \phi_n + \kappa_{kt} + \epsilon_{ipnkt} \quad (9)$$

$$\ln(\text{DistAmen})_{pn} = \beta \text{FastSide}_p + \sum_{\tau=1}^4 \left(\beta_{\tau}^{\text{slow}} \text{DistBoundary}_p^{\tau} + \beta_{\tau}^{\text{fast}} \text{FastSide}_p * \text{DistBoundary}_p^{\tau} \right) + \phi_n + \epsilon_{pn} \quad (10)$$

where as before p indexes full residential postcodes and n indexes neighborhoods defined as individual boundary segments as described above. Specifications (8)-(10) provide estimation results on six additional outcome variables. These additional estimations are aimed at testing for statistically significant jumps when moving from the slow side of a given boundary segment to the faster side for the following outcomes: log house prices, student propensities to be eligible for free school meals or to have a non-white ethnic background, and log residential distances to supermarkets, major roads, and schools. The identification strategy outlined in Section 4 requires all of these observable locational characteristics to vary smoothly across the invisible exchange station boundaries.

In specification (8) t indexes individual quarters (e.g. 1st quarter of 2005) during which property transactions j took place over the estimation period 2002-2008. $\ln(\text{PropValue})_{jpnt}$ are the universe of log property transaction values from England's Land Registry reported for a property j , and X_{jt} is a vector of property characteristics including the type of property (e.g. 2 bedroom house

versus studio), whether it is a new building, and the ownership type (lease or free hold). Next, in specification (9) t indexes years in which exam type k was taken by student i . $StudentChar_{ipnkt}$ is either free school meal eligibility or ethnic background (White or Other) of student i . Finally, in equation (10) $\ln(DistAmen)_{pn}$ is log distance from the centroid of postcode p to the nearest supermarket, major road, or primary or secondary school. As noted at the end of Section 2, even though the research design is based on comparisons between only very nearby outcomes, 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 have a higher likelihood to host a given type of local (dis-)amenity –the combination of which could lead to bias in the boundary effect. Specification (10) is aimed to assess such concerns.

As before, we alternatively estimate specifications (8)-(10) after replacing the polynomial functions with the 100m distance bins in order to assess the fit of the imposed flexible functional forms. We also once again estimate each specification both for all English boundary segments as well as after restricting attention to the top 30 percent of boundary segments with the highest mean differences in residential connections distances within 1 km on both sides.

Figure 6 reports the estimation results across four specifications for the boundary effect on house prices, and Tables 2 and 3 report estimation results for all specifications (8)-(10). In the interest of space, we report the figures for the five additional outcomes in specifications (9) and (10) in the Online Appendix Figures A.8-A.12.

House prices, student socio-economic characteristics as well as local amenities remain virtually flat across exchange station boundaries.³⁸ Thanks to the richness of the collected data, these results are precisely estimated. The zero effects are also confirmed in each of the six cases after restricting attention to the boundary segments with the largest observed variation in connection distances and realized internet speeds.³⁹ These additional estimation results provide reassurance for the identification strategy outlined in Section 4.

Finally, as above for the test score results, the figures also depict interesting correlations with respect to exchange station distances. For example, within a given catchment area we see a statistically significant increase in access to local schools or supermarkets as exchange connection distances become smaller, and we see that free school meal eligibility or the propensity for non-white ethnic backgrounds is significantly higher when taking simple averages on the side with higher connection distances (left) compared to the side with shorter connection distances (right). Once again, these correlations indicate endogeneity concerns that operate at a very granular spatial scale when projecting test scores or other outcomes on residential exchange distances, which the exchange station boundary discontinuity design appears to effectively address.

³⁸Note that the estimated zero effect on property values depicted in Figure 6 is at odds with the positive effect of speed on house prices that Ahlfeldt et al. (2015) report. In addition to using a different identification strategy (they use variation in the timing of technology upgrades across exchanges and/or households), another difference is that we use data on the universe of property transactions in England over the period 2002-08, whereas Ahlfeldt et al. (2015) use information from one financial institution that finances roughly 10 percent of the mortgage-based transactions.

³⁹Notice that there is a very tiny but statistically significant jump in the propensity for free school meal eligibility depicted in the left panel of Online Appendix Figure A.9 and reported in Table 2. This becomes a statistical zero once we control for other student characteristics and/or school fixed effects. And as we document in Figure 4 and Table 2, the test score effect remains a precisely estimated zero both before and after conditioning on student observables.

6 Channels

Guided by the theoretical framework in Section 3, this section (step 3) brings to bear additional geo-referenced survey microdata on student time use and internet use to further investigate the potentially opposing mechanisms underlying the estimated reduced form effect. To disentangle the different components of the reduced form effect $\beta = \delta + \alpha\eta$ in equation (3), we estimate the following specification with a number of different student outcomes on the left hand side:

$$\begin{aligned} TimeUse_{ipnt} = & \beta FastSide_p + \sum_{\tau=1}^4 \left(\beta_{\tau}^{slow} DistBoundary_p^{\tau} + \beta_{\tau}^{fast} FastSide_p * DistBoundary_p^{\tau} \right) \\ & + \gamma X_{it} + \phi_n + \kappa_t + \epsilon_{ipnt} \end{aligned} \quad (11)$$

where as before i indexes students, p indexes individual residential postcodes, n indexes neighborhoods and t indexes rounds of survey data and X_{it} is a vector of basic student characteristics (second-order polynomial for age and a gender dummy). Again, we estimate these specifications both before and after replacing the polynomial functions with the less parametric specifications discussed in Section 4. $TimeUse_{ipnt}$ in specification (11) is one of several measures for either student internet use or study time supply. For internet use, this is measured either as the reported number of days per week that a student in the British Household Panel Survey (BHPS) or in the Understanding Society survey (US) uses the internet, or as the reported number of hours per day spent using email and online social media from the US survey. For study time, we use one of four different measures. The first is the number of weekdays doing homework in the Longitudinal Study of Young People in England (LSYPE) survey. The second is the number of hours spent doing homework per weekday in the US survey.⁴⁰ The third is the propensity of using internet resources for homework from the LSYPE survey. And the fourth is the number of hours spent doing homework using a computer also from the LSYPE survey.⁴¹

How do these additional empirical moments on student time use and internet use relate to the theoretical framework that we presented in Section 3? The reduced form effect $\beta = \delta + \alpha\eta$ that we estimate in Section 5 can be decomposed into the sum of two potentially opposing effects. The first is the effect of lower time costs in accessing online content on the the learning productivity per unit of time spent studying, which is captured by the elasticity δ . The second is the effect of reducing the time cost of online content consumption on the supply of student time spent studying, which is captured by the elasticity η . In the theory section, we refer to the hypothesis that $\delta > 0$ as the “MOOC effect”. That is, lower information and communication costs increase the productivity of a given amount of study time for educational achievement. In turn, we refer to the hypothesis that $\eta < 0$ as the “Facebook effect”. That is, lower information and communication

⁴⁰We compute this variable by combining answers to the following two separate survey questions: “During an average week in term time, on how many evenings do you do any homework? Please just think about Monday to Friday evenings during term time.”; and “When you do homework on a week-day evening during term time, how many hours do you usually spend doing your homework?”

⁴¹We compute this variable by combining answers from the following two questions: “Days a week the young person usually uses home computer for school work, during term time”; and “Number of hours the young person uses home computer for school work, on a day when use computer for school work during term time”.

costs decrease effective study time supply by either making other online or offline activities relatively more attractive (positive relative price effect) or by letting students access a fixed amount of learning content in less time (endowment effect). Conversely, we also discussed that both the effect on learning productivity and on study supply could in principle also go in the opposite direction.

The additional specifications summarized in (11) allow us to empirically test for the sign, size and statistical significance of η (the effect on study time supply). Furthermore, the combination of η with our previous estimates of the effect on student test scores (the β) allow us to infer the sign and significance of δ (the effect on educational attainment holding constant study times). Finally, by observing student responses in their time spent on the internet, we can in addition learn about the changes in student consumption of online content underlying these effects, and provide empirical evidence on the elasticity of student demand for online content with respect to its time cost.

6.1 Estimation Results

Before estimating the specifications in (11), we re-run specification (6) from Section 5 on the sample of residential postcodes for which we have student survey microdata in order to verify that the treatment in the available ICT is present in the same way as reported above in these estimations. Appendix Figure A.13 presents the estimation results which document that the boundary treatment jumps for surveyed students in these samples are of similar magnitude compared to the population data presented in Figure 2. As before, crossing the average boundary is associated with approximately 65 percent of a standard deviation reduction in distance to the connected exchange, which corresponds to sizable changes in the available ICT as shown using the telecom network data in Section 5.

Despite these significant jumps in the available ICT, Figure 7 and Table 3 present estimation results that suggest a zero effect on student time spent on the internet.⁴² Figures 8 and 9 then move attention to the effect on student time spent studying and time spent studying using online resources. Once again the estimation results, which we also report again in Table 3, suggest zero effects both on total student time spent studying as well as on student time spent studying using online resources.

These results provide a number of additional insights. First, we find that exogenous changes in the available ICT do not significantly affect student supply of study time online or offline, or their learning productivity per time spent studying. Relating back to the terminology we use to fix ideas in the model, the empirical analysis does not provide evidence in favor of the Facebook effect or in favor of the MOOC effect of ICT upgrades on student educational outcomes. Given that this is the first empirical study to relate administrative test score performances to large and exogenous variation in the available ICT in the cross-section of a student population, we believe these results serve to inform currently ongoing policy debates on ICT infrastructure investments that are targeted at providing ever faster broadband internet connections.

Second, the presented results are also informative about the elasticity of student consumption of online information and communication with respect to the per-unit (e.g. Mbit) time cost of on-

⁴²Because we have to switch from administrative census data to using survey microdata in these estimations, point estimates are naturally somewhat less precise than in the previous section.

line consumption, which is a key parameter for ICT policy decisions for which we are currently not aware of existing empirical estimates. In particular, our findings suggest that the elasticity of student demand for online content is negative, but bounded by -1. The empirical moments underlying this conclusion are that i) there are large and significant jumps in the available ICT across catchment area boundaries as documented in Section 5, but ii) these jumps do not appear to lead to significant increases (or reductions) in the time students spend on the internet. These results suggest that upgrades in the available ICT at student residences are at least weakly increasing their consumption of online content.⁴³ However, in absence of a significant effect on internet time use, the maximum increase in the consumption of online content can only be proportional to the decrease in the time cost of accessing that content, which effectively bounds the elasticity of demand for online consumption with respect to its time cost to lie between 0 and -1.

7 Conclusion

A large wave of currently ongoing policy proposals are motivated by strong claims about the benefits of increases in the available broadband connection speeds for education and growth. However, these policies are so far based on limited empirical evidence. This paper proposes a new empirical strategy to estimate the causal effects of these policies, and applies it to the questions of whether and how ICT improvements affect educational attainment. To base estimations on exogenous variation in ICT, we notice that capacity constraints at individual telephone exchange stations give rise to invisible and essentially randomly placed boundaries of exchange station catchment areas that in turn give rise to large and discontinuous jumps in the available ICT across space. Because this spatial RD design exploits cross-sectional variation in ICT across thousands of boundary segments, it can be applied to estimate the medium to long-run effects of ICT upgrades, and to flexibly explore the heterogeneity of these effects along multiple dimensions. We then combine this methodology with a new collection of microdata that allows us to link administrative records on educational achievement for the population of English primary and secondary school students to the available ICT at their home addresses.

The analysis presents a number of new results. We find that even substantial increases in the available broadband connection speeds have a precisely estimated zero effect on educational attainment, and that these effects are causally identified: house prices, student socio-economic characteristics and local amenities remain virtually flat across the boundaries. We explore to what extent this average effect may mask significant heterogeneity, and find that the precisely estimated zero effect holds for all age, gender, ethnic and socioeconomic groups in the student population, for all years in the estimation sample 2002-2008, for different subjects (English, Mathematics, Science), across different treatment intensities, and across different initial levels of the available ICT.

Guided by a simple theoretical framework, we further explore to what extent the zero reduced form effect may be driven by the interplay of opposing effects on students' learning productivity as opposed to student supply of time spent studying. While ICT upgrades appear to at least weakly

⁴³The ideal data to precisely estimate this parameter would be geo-referenced microdata on the amount of megabytes downloaded or uploaded, which we could combine with the exchange boundary RD design. We are unaware that such data have been made available for research in the English context or elsewhere.

increase student consumption of online content, we find no significant effects on student time spent studying online or offline, or on the productivity of time spent studying. The results also suggest that the elasticity of student demand for online consumption with respect to its time cost is negative, but bounded at -1 in our empirical setting.

The paper serves to inform currently ongoing policy debates on ICT in two main ways. The first is that the analysis provides policy makers with credible empirical estimates for an important outcome of interest (education) that has served as one of the central motivations for ICT upgrades and has featured prominently in recent debates about the so-called “digital divide” (Norris, 2001). Second, the empirical strategy that we present in this paper provides a useful tool that can be applied to study a number of additional policy-relevant questions on the effects of ICT upgrades, and in a variety of different empirical contexts.

References

- Daniel Aaronson, Lisa Barrow, and William Sander. Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics*, 25:95–135, 2007.
- Gabriel Ahlfeldt, Pantelis Koutroumpis, and Tommaso Valletti. Speed 2.0 Evaluating Access to Universal Digital Highways. CESifo Working Paper 5186, 2015.
- Anders Akerman, Ingvil Gaarder, and Magne Mogstad. The Skill Complementarity of Broadband Internet. *Quarterly Journal of Economics*, forthcoming.
- Joshua Angrist and Victor Lavy. Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics*, 114(2):533–575, 1999.
- Joshua Angrist and Victor Lavy. New Evidence on Classroom Computers and Pupil Learning. *Economic Journal*, 112(482):735–765, 2002.
- Joshua Angrist and Miikka Rokkanen. Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff. *Journal of the American Statistical Association*, forthcoming.
- Lisa Barrow, Lisa Markman, and Cecilia Elena Rouse. The Educational Benefits of Computer-Aided Instruction. *American Economic Journal: Economic Policy*, 1(1):52–74, 2009.
- Stefan Bauernschuster, Oliver Falck, and Ludger Woessmann. Surfing alone? The Internet and Social Capital: Evidence from an Unforeseeable Technological Mistake. *Journal of Public Economics*, 117(C):73–89, 2014.
- Andriana Bellou. The Impact of Internet Diffusion on Marriage Rates: Evidence from the Broadband Market. *Journal of Population Economics*, 28(2):265–297, 2015.
- Daniel O. Beltran, Kuntal K. Das, and Robert W. Fairlie. Home Computers and Educational Outcomes: Evidence from NSLY97 and CPS. *Economic Inquiry*, 48(3):771–792, 2010.
- Marinho Bertanha. Regression Discontinuity Design with Many Thresholds. manuscript, 2015.
- Manudeep Bhuller, Tarjei Havnes, Edwin Leuven, and Magne Mogstad. Broadband Internet: An Information Superhighway to Sex Crime? *Review of Economic Studies*, 8(4):1237–1266, 2013.

- Sandra E. Black. Do Better Schools Matter? Parental Valuation of Elementary Education. *Quarterly Journal of Economics*, 114(2):577–599, 1999.
- George Bulman and Robert W. Fairlie. Technology and Education: Computers, Software, and the Internet. In *Handbook of the Economics of Education*, volume 5. Elsevier, forthcoming.
- Filipe R. Campante, Ruben Durante, and Francesco Sobbro. Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation. NBER Working Paper 19029, 2013.
- Damon Clark. The Performance and Competitive Effects of School Autonomy. *Journal of Political Economy*, 117(4):745–783, 2009.
- Flavio Cunha, James J. Heckman, and Susanne M. Schennach. Estimating the Technology of Cognitive and Non-cognitive Skill Formation. *Econometrica*, 78(3):883–931, 2010.
- David J. Deming, Claudia Goldin, Lawrence F. Katz, and Noam Yuchtman. Can Online Learning Bend the Higher Education Cost Curve? NBER Working Paper 20890, 2015.
- Department for Culture, Media and Sport. The UK Broadband Delivery Scheme: Overview. Technical Report, February 2013.
- Robert W. Fairlie. The Effects of Home Computers on School Enrollment. *Economics of Education Review*, 24(5):533–547, 2005.
- Robert W. Fairlie and Rebecca A. London. The Effects of Home Computers on Educational Outcomes: Evidence from a Field Experiment with Community College Students. *Economic Journal*, 122(561):727–753, 2012.
- Robert W. Fairlie and Jonathan Robinson. Experimental Evidence on the Effects of Home Computers on Academic Achievement among Schoolchildren. *American Economic Journal: Applied Economics*, 5(3):211–240, 2013.
- Oliver Falck, Robert Gold, and Stephan Heblich. E-Lectures: Voting Behavior and the Internet. *American Economic Review*, 107(7):2238–2265, 2014.
- Oliver Falck, Constatin Mang, and Ludger Woessmann. Virtually No Effect? Different Uses of Classroom Computers and their Effect on Student Achievement. CESifo Working Paper No. 5266, March 2015.
- Federal Communications Commission. Connecting America: The National Broadband Plan. Technical report, Federal Communications Commission, March 2010.
- Mario Fiorini. The Effect of Home Computer Use on Children’s Cognitive and Non-Cognitive Skills. *Economics of Education Review*, 29(1):55–72, 2010.
- Chris Forman, Avi Goldfarb, and Shane Greenstein. The Internet and Local Wages: A Puzzle. *American Economic Review*, 102(1):556–75, 2012.
- Thomas Fuchs and Ludger Wossmann. Computers and Student Learning: Bivariate and Multivariate Evidence on the Availability and Use of Computers at Home and at School. *Brussels Economic Review*, 47(3-4):359–386, 2004.
- Andrew Gelman and Guido Imbens. Why High-order Polynomials Should not be Used in Regression Discontinuity Designs. NBER Working Paper 20405, 2014.

- Stephen Gibbons, Stephen Machin, and Olmo Silva. Valuing School Quality Using Boundary Discontinuities. *Journal of Urban Economics*, 75:15–28, 2013.
- Austan Goolsbee and Johathan Guryan. The Impact of Internet Subsidies in Public Schools. *Review of Economics and Statistics*, 88(2):336–347, 2006.
- Caroline M. Hoxby. The Economics of Online Postsecondary Education: MOOCs, Nonselective Education, and Highly Selective Education. *American Economic Review P&P*, 104(5):528–533, 2014.
- Stephen Machin, Sandra McNally, and Olmo Silva. New Technology in Schools: Is There a Payoff? *Economic Journal*, 117(522):1145–1167, 2007.
- Ofer Malamud and Cristian Pop-Eleches. Home Computer Use and the Development of Human Capital. *Quarterly Journal of Economics*, 126(2):987–1027, 2011.
- Pippa Norris. *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide*. Cambridge University Press, Cambridge, 2001.
- UK OfCom. Children and Parents: Media Use and Attitudes Report. Report, 2012.
- Steven G. Rivkin, Eric A. Hanushek, and John F. Kain. Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2):417–458, 03 2005.
- John Schmitt and Jonathan Wadsworth. Is There an Impact of Household Computer Ownership on Children’s Educational Attainment in Britain? *Economics of Education Review*, 25(6):659–673, 2006.
- The White House. FACT SHEET: ConnectHome: Coming Together to Ensure Digital Opportunity for All Americans. Technical report, Office of the Press Secretary, July 2015.
- The White House. Four Years of Broadband Growth. Technical report, Office of Science and Technology Policy and the National Economic Council, June 2013.
- Jacob Vigdor, Helen Ladd, and Erika Martinez. Scaling the Digital Divide: Home Computer Technology and Student Achievement. *Economic Inquiry*, 52(3):1103–1119, 2014.

8 Figures and Tables

Figures

Figure 1: Two Maps of Exchange Station Catchment Area Boundaries in England
Figure 1.1: Exchange Station Catchment Areas

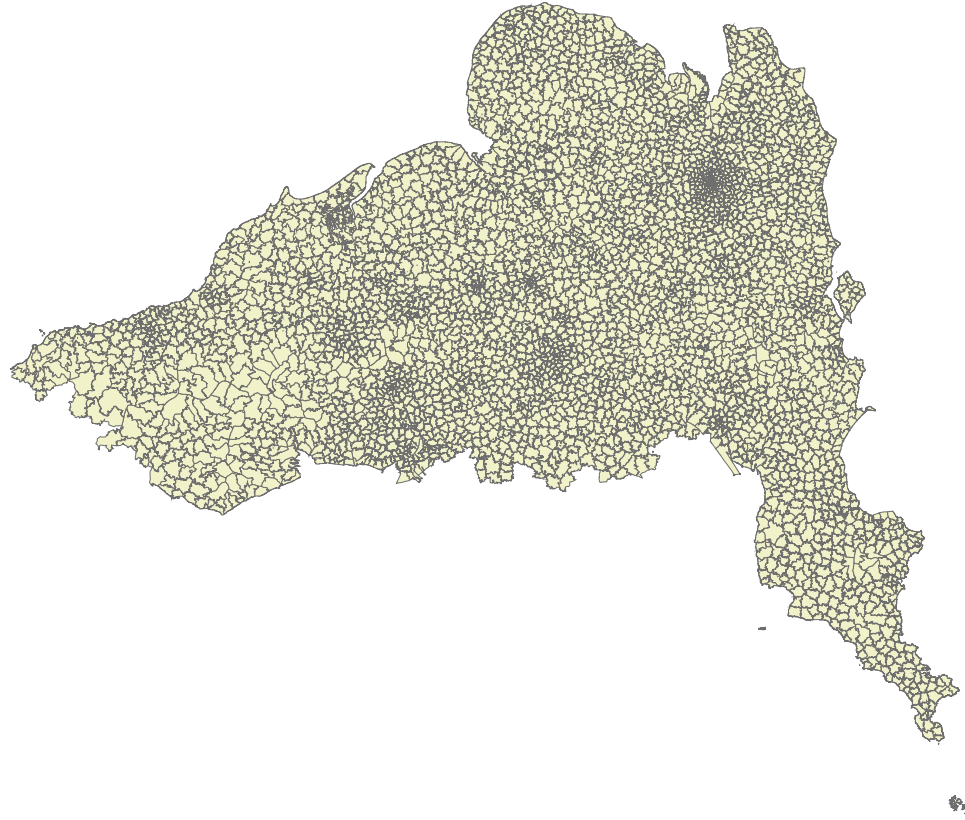


Figure 1.2: Zoom Into North London



Notes: Both figures are based on the UK telecom network data. The left panel depicts exchange station catchment area boundaries in England. The right panel depicts these catchment boundaries (thick black lines) in addition to individual postcodes (thinner gray lines) in one of the author's residential areas in North London. The black points indicate the locations of exchange stations. Descriptive statistics on the postcode and boundary estimation samples are reported in [Table 1](#).

Figure 2: The Boundary Effect on Residential Distances to the Connected Exchange Station

Figure 2.1: Full Specification - All Boundaries

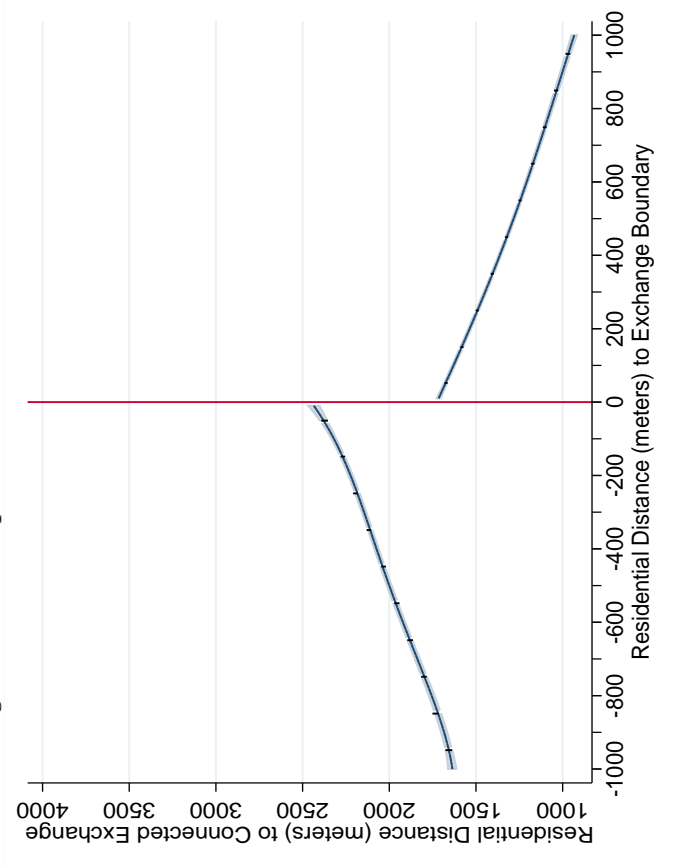
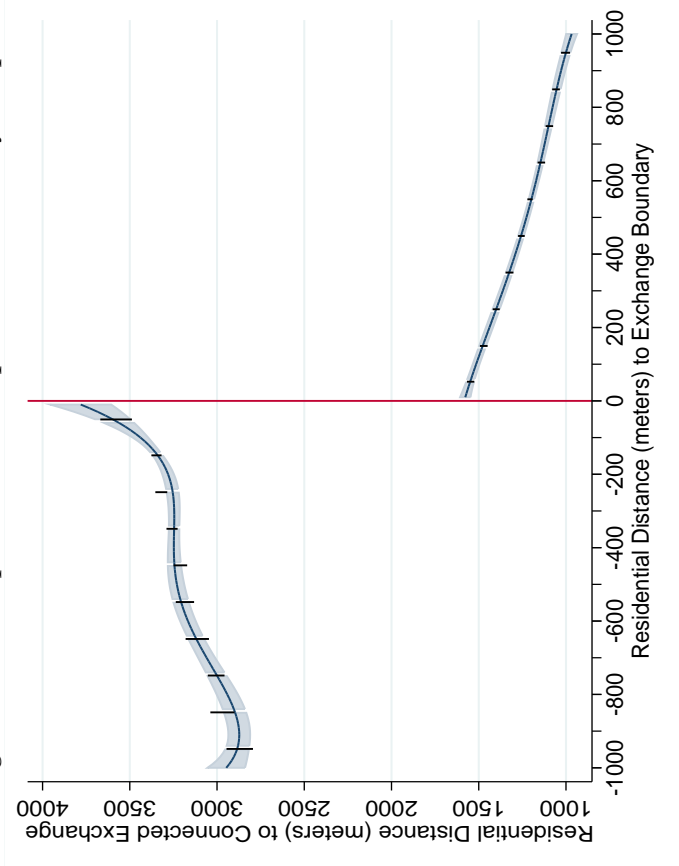


Figure 2.2: Full Specification - Top 30% of Boundary Jumps



Notes: The left hand side in both graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 5. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Descriptive statistics on the estimation samples are reported in Table 1.

Figure 3: The Boundary Effect on Observed Internet Connection Speed

Figure 3.1: Full Specification - All Boundaries

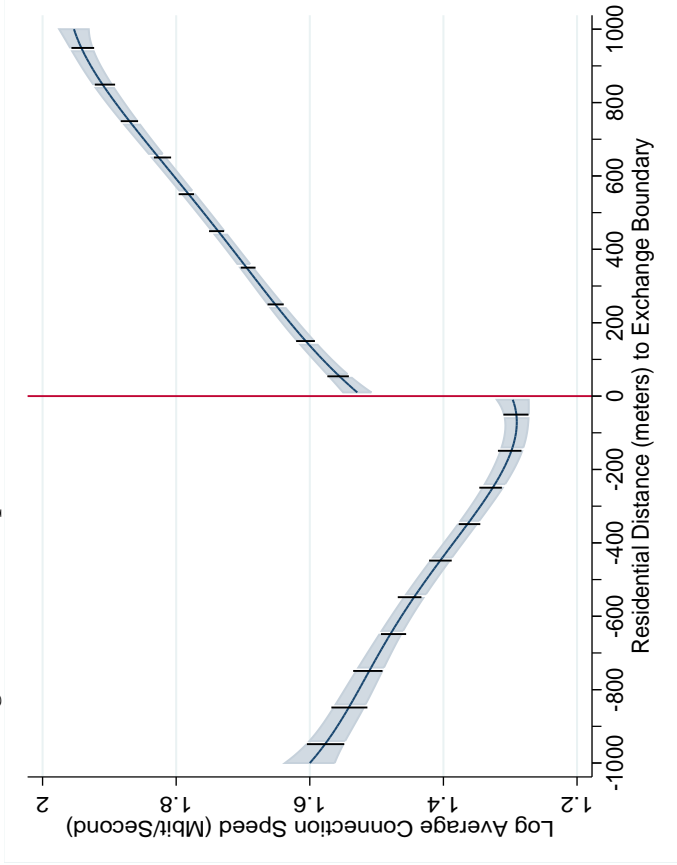
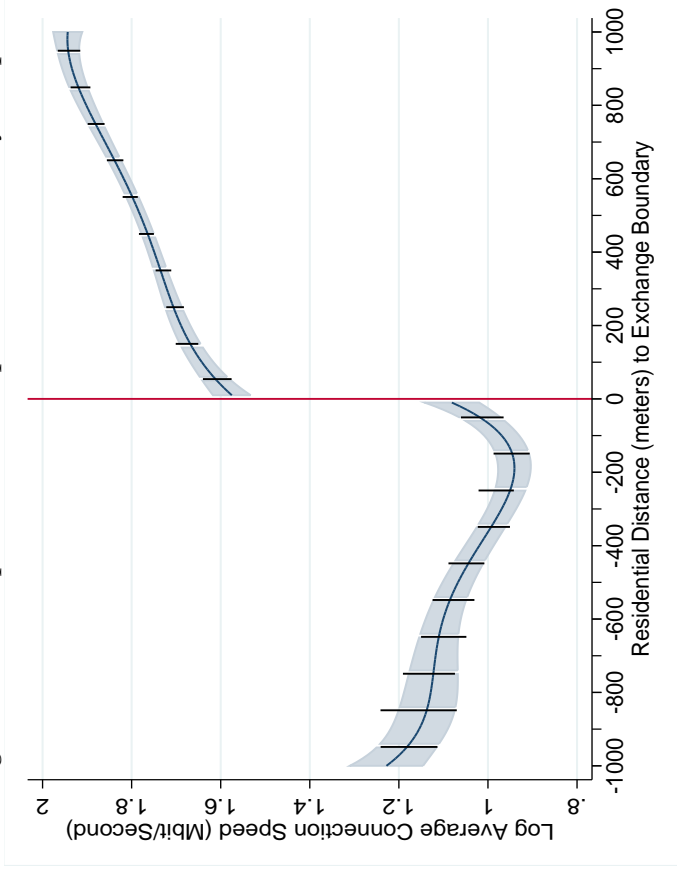


Figure 3.2: Full Specification - Top 30% of Boundary Jumps



Notes: The left hand side in both graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 5. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Descriptive statistics on the estimation samples are reported in Table 1.

Figure 4: The Boundary Effect on Educational Attainment

Figure 4.1: Baseline Specification - All Boundaries

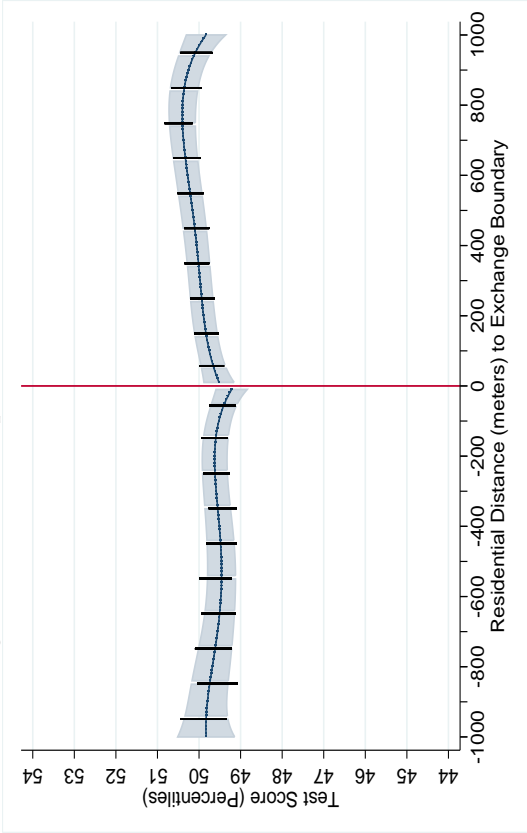


Figure 4.2: Full Specification - All Boundaries

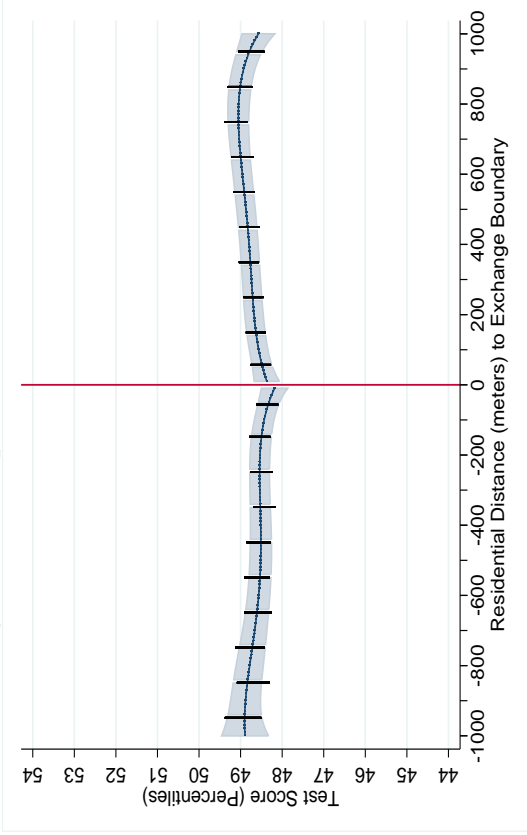


Figure 4.3: Baseline Specification - Top 30% of Boundary Jumps

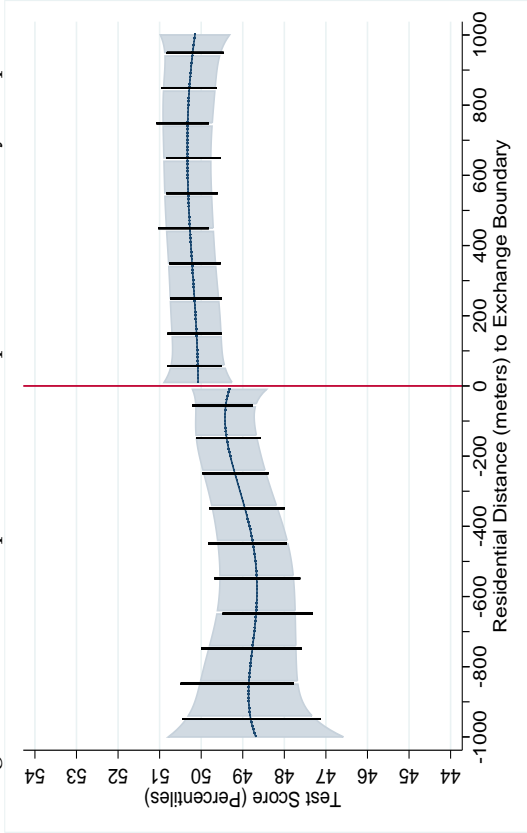
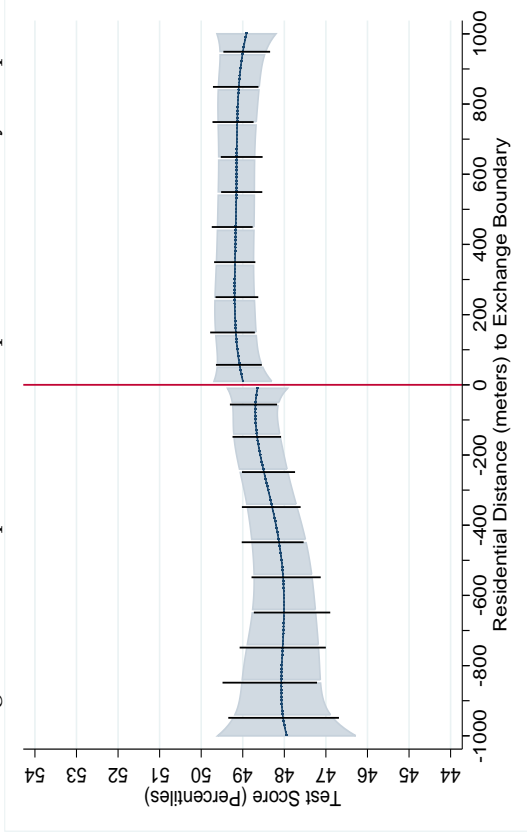


Figure 4.4: Full Specification - Top 30% of Boundary Jumps



Notes: The left hand side in all graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 5. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Descriptive statistics on the estimation samples are reported in Table 1.

Figure 5: The Boundary Effect on Educational Attainment – Additional Results

Figure 5.1: Full Specification - Extended Estimation Sample

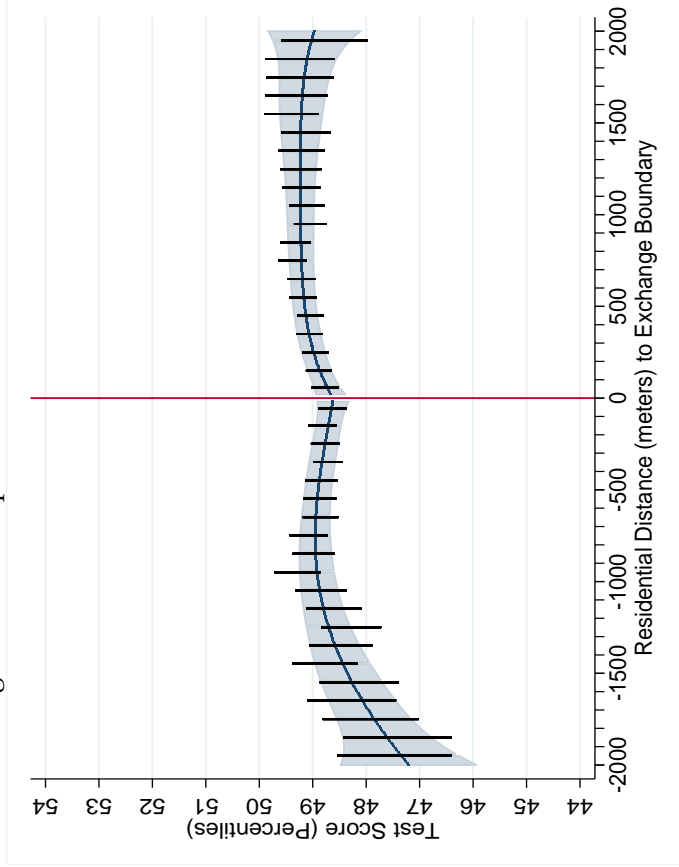
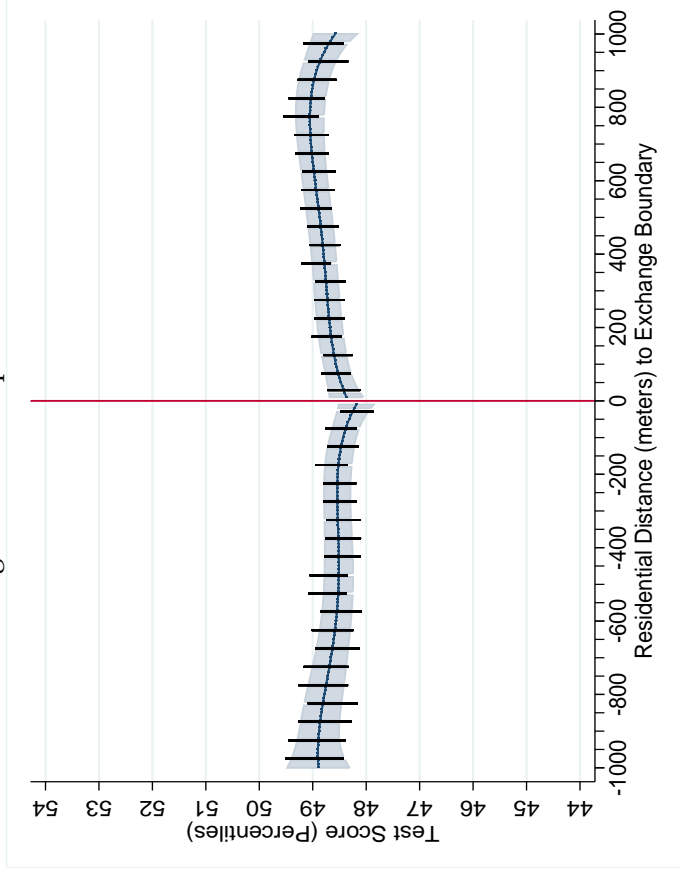


Figure 5.2: Full Specification - 50 m Bins



Notes: The left hand side in both graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 5. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100 or 50 m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Figure 5.1 is estimated using around 8.7m test score observations of 5.5m students in 561,584 residential postcodes, across 19,586 boundary segments and 3,894 exchange stations. Descriptive statistics on the main estimation sample of Figure 5.2 are reported in Table 1.

Figure 6: The Boundary Effect on Property Values

Figure 6.1: Full Specification - All Boundaries

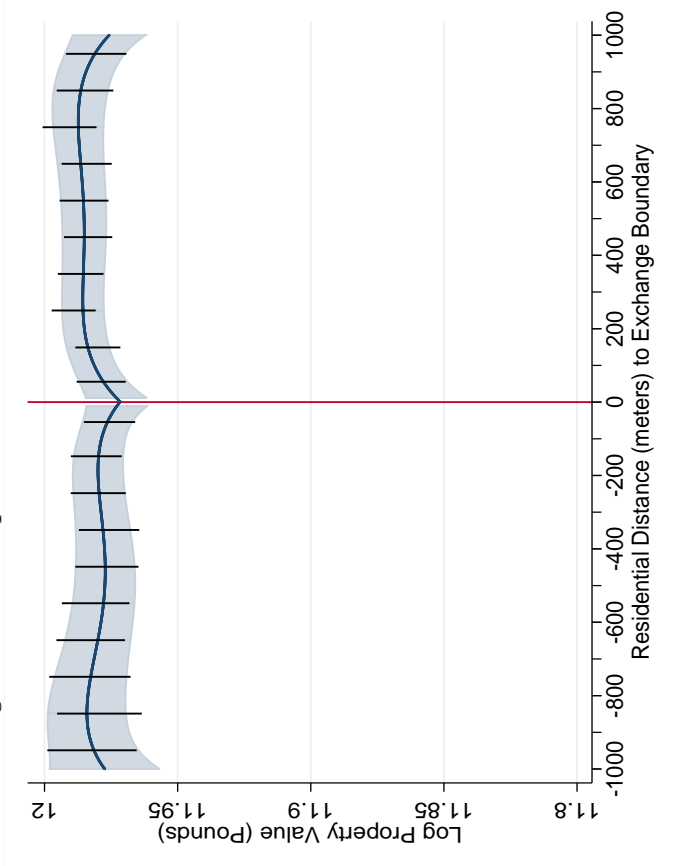
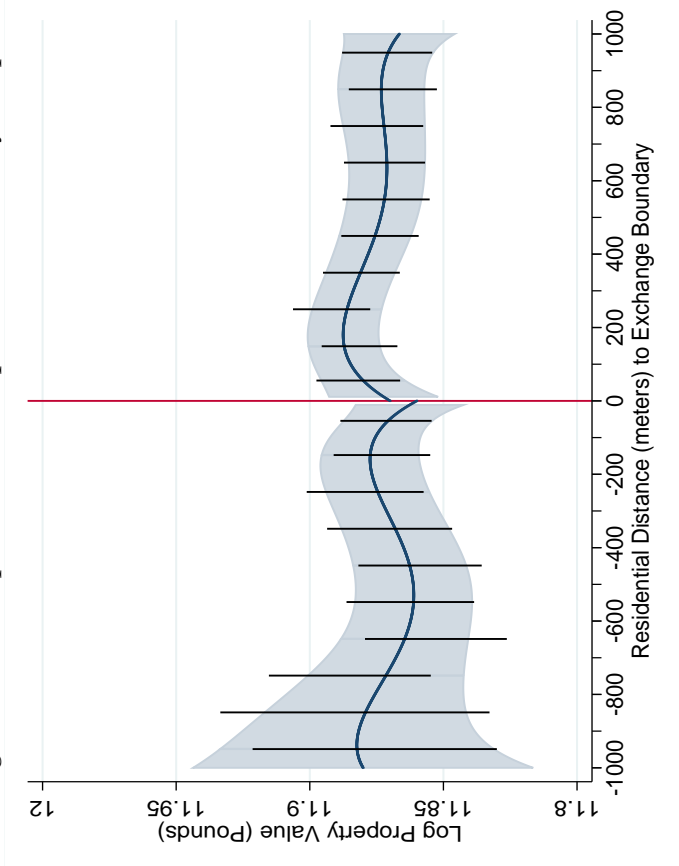


Figure 6.2: Full Specification - Top 30% of Boundary Jumps



Notes: The left hand side in both graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 5. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Descriptive statistics on the estimation samples are reported in Table 1.

Figure 7: The Boundary Effect on Student Time Spent Online

Figure 7.1: Days Per Week Using the Internet

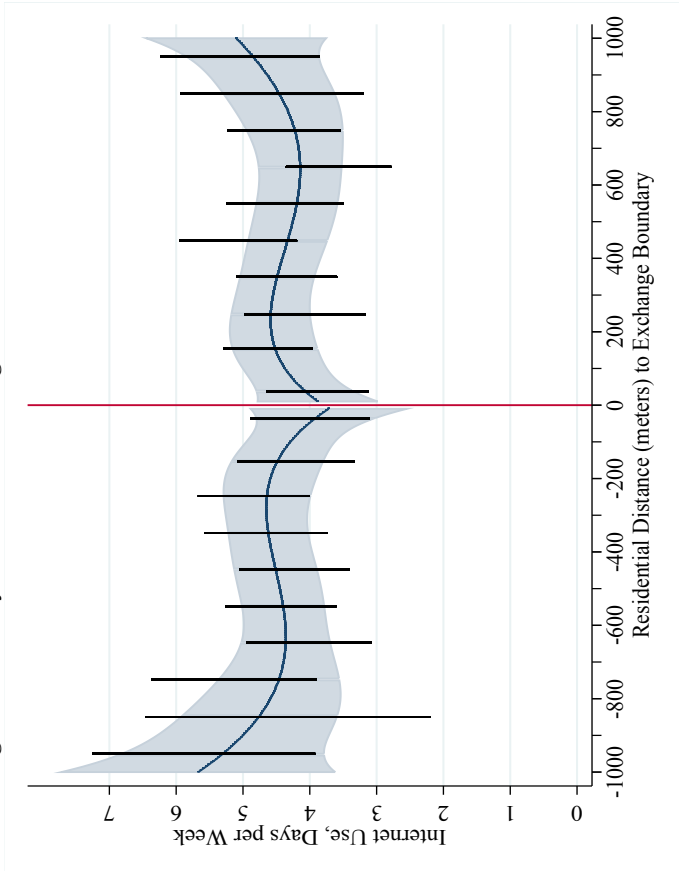
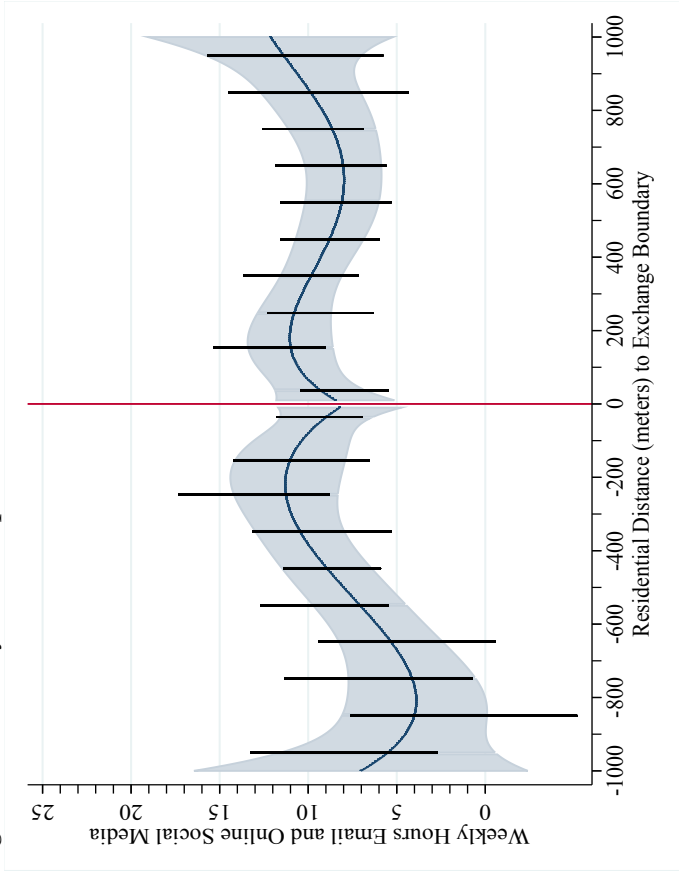


Figure 7.2: Weekly Hours Spent on Email and Online Social Media



Notes: The left hand side in both graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 6. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Descriptive statistics on the estimation samples are reported in Table 1.

Figure 8: The Boundary Effect on Student Time Spent Studying

Figure 8.1: Weekly Hours Spent on Homework

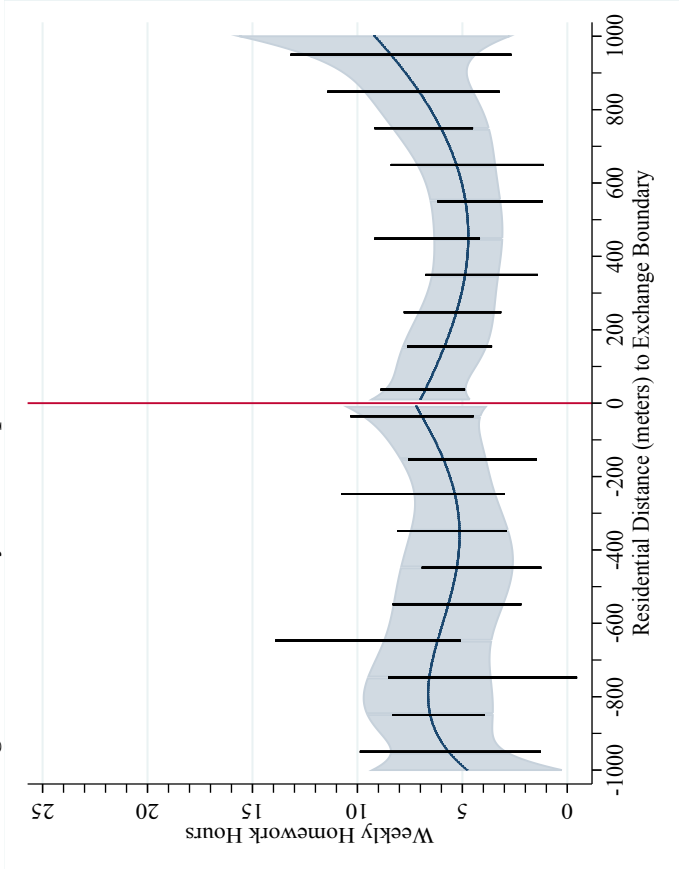
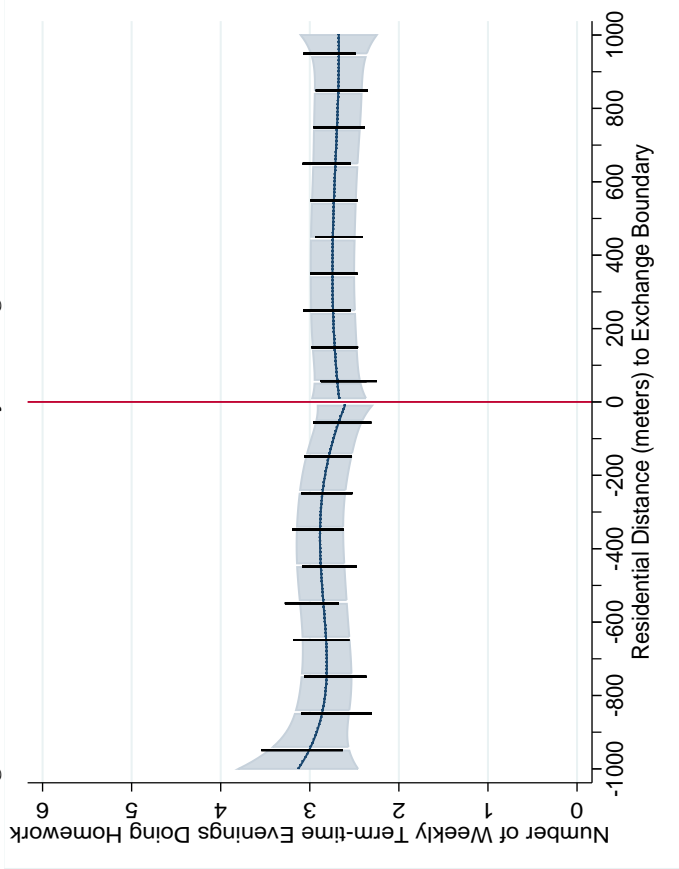


Figure 8.2: Number of Weekdays Doing Homework



Notes: The left hand side in both graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 6. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Descriptive statistics on the estimation samples are reported in Table 1.

Figure 9: The Boundary Effect on Student Time Spent Studying Online

Figure 9.1: Propensity Using Online Resources for Homework

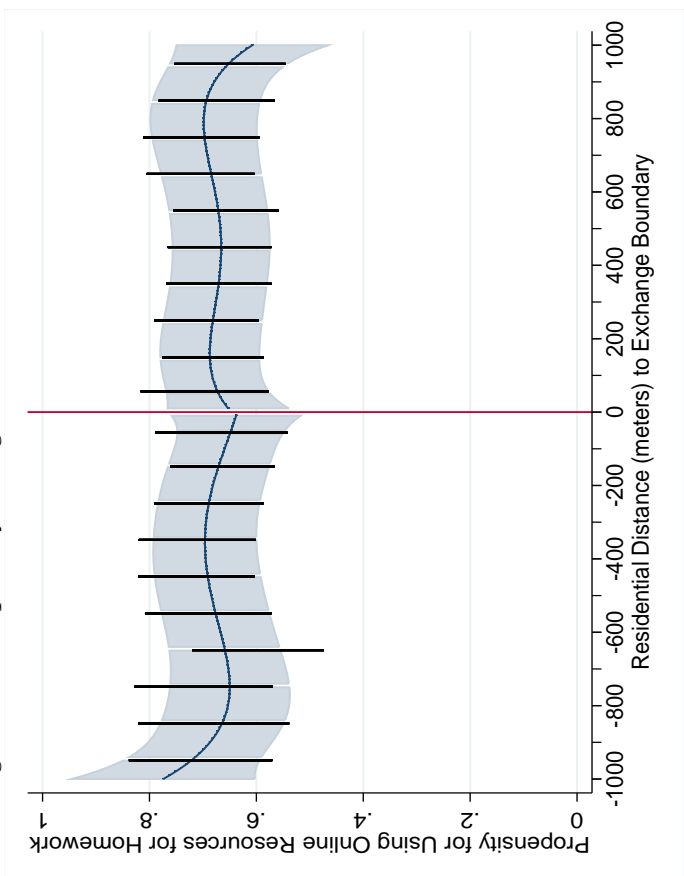
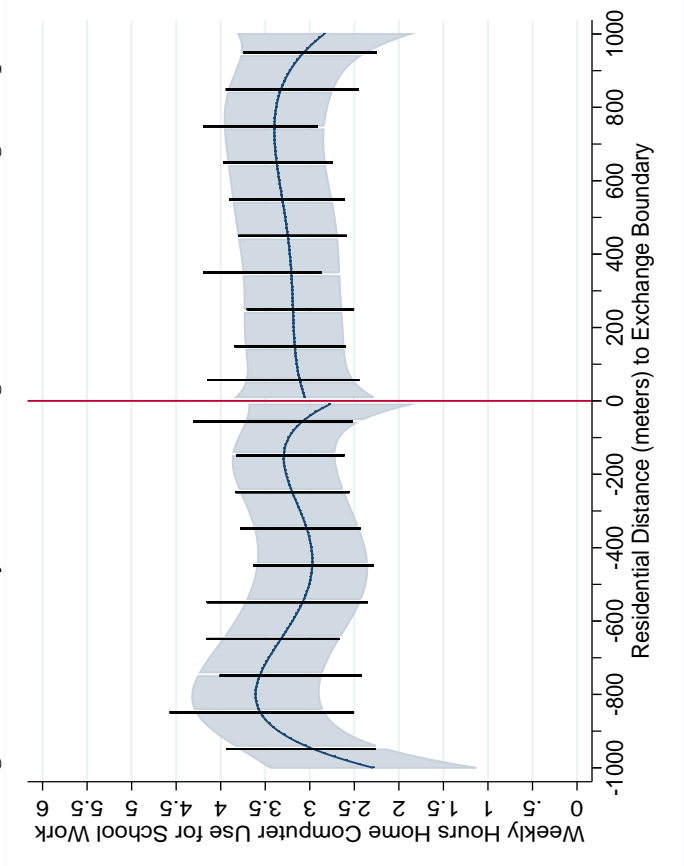


Figure 9.2: Weekly Hours Doing Homework Using Computer



Notes: The left hand side in both graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 6. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Descriptive statistics on the estimation samples are reported in Table 1.

Tables

Table 1: Descriptive Statistics

Data Set	Variable	Number of Observations	Mean	Standard Deviation	Min	Max
Telecom Network Data	Distance to Connected Exchange Station (m)	581,594	1,638	1,152	0	240,266
	Distance to Nearest Boundary Segment (m)	581,594	415.3	276.9	0.000274	1,000
	Fast Side Dummy	581,594	0.614	0.487	0	1
	Number of Exchange Stations	3,920
	Number of Boundary Segments	20,248
Administrative Test Score Records 2002-08	Test Score Percentile	7,021,667	48.05	28.52	1	100
	Year	7,021,667	2,005	1.990	2,002	2,008
	Male	7,021,667	0.504	0.500	0	1
	Free School Meal	7,021,667	0.177	0.382	0	1
	Ethnic Background White British	7,021,667	0.741	0.438	0	1
	Distance to Connected Exchange Station (m)	7,021,667	1,622	944.9	0	62,599
	Distance to Nearest Boundary Segment (m)	7,021,667	419.8	272.2	0.000620	1,000
	Fast Side Dummy	7,021,667	0.607	0.488	0	1
	Number of Exchange Stations	3,883
	Number of Boundary Segments	19,472
Internet Speed Data 2012-13	Average Internet Speed	178,027	6.179	3.771	0.200	15
	Log Average Internet Speed	178,027	1.607	0.698	-1.609	2.708
	Year	178,027	2,012	0.499	2,012	2,013
	Distance to Connected Exchange Station (m)	178,027	1,960	1,283	0	80,033
	Distance to Nearest Boundary Segment (m)	178,027	436.4	277.6	0.0389	1,000
	Fast Side Dummy	178,027	0.619	0.486	0	1
	Number of Exchange Stations	3,800
	Number of Boundary Segments	15,997
UK Land Registry Records of Property Transactions 2002-08	Property Value (British Pound)	3,559,352	191,463	186,276	5,250	19,750,000
	Log Property Value	3,559,352	11.93	0.662	8.566	16.80
	Year	3,559,352	2,005	1.915	2,002	2,008
	Apartment Dummy (as opposed to houses)	3,559,352	0.238	0.426	0	1
	Newly Built Dummy	3,559,352	0.0937	0.291	0	1
	Leasehold Dummy (as opposed to freehold)	3,559,352	0.299	0.458	0	1
	Distance to Connected Exchange Station (m)	3,559,352	1,535	927.4	0	61,828
	Distance to Nearest Boundary Segment (m)	3,559,352	411.5	275.1	0.000620	1,000
	Fast Side Dummy	3,559,352	0.618	0.486	0	1
	Number of Exchange Stations	3,815
	Number of Boundary Segments	18,308
GIS Data on Local Amenities	Distance to the Nearest Supermarket (m)	581,594	1,200	1,430	0	21,805
	Distance to the Nearest Major Road (m)	581,594	370.4	474.6	0	9,314
	Distance to the Nearest Primary or Secondary School (m)	581,594	549.4	567.6	0	8,624
	Distance to Connected Exchange Station (m)	581,594	1,637	1,152	0	240,266
	Distance to Nearest Boundary Segment (m)	581,594	415.3	276.9	0	1,000
	Fast Side Dummy	581,594	0.614	0.487	0	1
	Number of Exchange Stations	3,920
	Number of Boundary Segments	20,248
Geo-referenced Survey Data from the BHPS/US 2002-13	Days per Week Using the Internet	5,046	4.668	2.496	0	7
	Weekly Hours Email and Social Media	5,308	8.777	9.791	0	48
	Weekly Homework Hours	3,135	5.319	5.332	0	52
	Distance to Connected Exchange Station (m)	10,066	1,569	924.3	0	12,331
	Distance to Nearest Boundary Segment (m)	10,066	413.9	273.9	0	999.8
	Fast Side Dummy	10,066	0.608	0.488	0	1
	Number of Exchange Stations	984
	Number of Boundary Segments	1,999
Geo-referenced Survey Data from the LSYPE 2004-05	Number of weekdays doing homework	11,020	2.937	1.348	0	5
	Use online resources for homework	10,345	0.760	0.427	0	1
	Weekly homework hours using ICT	10,345	3.280	3.657	0	60
	Distance to Connected Exchange Station (m)	10,345	1,558	910.2	0	7,669
	Distance to Nearest Boundary Segment (m)	10,345	403	267.5	0	999.5
	Fast Side Dummy	10,345	0.601	0.490	0	1
	Number of Exchange Stations	966
	Number of Boundary Segments	2,709

Notes: The reported statistics refer to our estimation samples that are based on all observations within 1 km of an exchange station boundary segment in England.

Table 2: Point Estimates

Dependent Variable	(1) Exchange Distance (meters) Full	(2) Exchange Distance (meters) Top 30% Jumps	(3) Log Speed Full	(4) Log Speed Top 30% Jumps	(5) Test Score Percentile Full	(6) Test Score Percentile Top 30% Jumps	(7) Log Property Value Full	(8) Log Property Value Top 30% Jumps	(9) FSME Dummy Full	(10) FSME Dummy Top 30% Jumps	(11) White Dummy Full	(12) White Dummy Top 30% Jumps
Fast Side Dummy (Jump)	-725.2*** (19.75)	-2.252*** (96.04)	0.224*** (0.0178)	0.465*** (0.0425)	0.196 (0.181)	0.349 (0.432)	-0.000255 (0.00652)	0.0100 (0.0116)	-0.00896** (0.00389)	0.000846 (0.00949)	0.00359 (0.00365)	0.0105 (0.00777)
Boundary Dist X Slow Side Dummy	-1.549*** (0.224)	-5.622*** (1.223)	-0.000246 (0.000163)	-0.00201*** (0.000411)	0.00423** (0.00205)	0.00204 (0.00543)	0.000110 (7.40e-05)	0.000259* (0.000156)	-0.000146*** (4.50e-05)	3.14e-05 (0.000113)	0.000142*** (4.23e-05)	0.000165* (8.84e-05)
Boundary Dist^2 X Slow Side Dummy	0.00277*** (0.000827)	0.0191*** (0.00456)	1.99e-06*** (6.79e-07)	7.98e-06*** (1.68e-06)	-1.53e-05* (8.89e-06)	-1.60e-05 (2.39e-05)	-4.81e-07 (3.16e-07)	-1.19e-06* (6.74e-07)	6.54e-07*** (1.90e-07)	-1.00e-08 (4.72e-07)	-6.04e-07*** (1.71e-07)	-6.90e-07* (3.57e-07)
Boundary Dist^3 X Slow Side Dummy	-3.85e-06*** (1.17e-06)	-2.67e-05*** (6.46e-06)	-2.53e-09*** (1.05e-09)	-1.05e-08*** (2.59e-09)	2.15e-08 (1.41e-08)	2.54e-08 (3.82e-08)	7.57e-10 (4.97e-10)	1.77e-09* (1.06e-09)	-1.02e-09*** (2.97e-10)	-6.02e-11 (7.37e-10)	8.64e-10*** (2.63e-10)	9.09e-10* (5.51e-10)
Boundary Dist^4 X Slow Side Dummy	1.81e-09*** (5.52e-10)	1.23e-08*** (3.07e-09)	0** (0)	0*** (0)	-0 (0)	-0 (0)	-0 (0)	-0 (0)	0*** (0)	0 (0)	-0*** (0)	-0 (0)
Boundary Dist X Fast Side Dummy	-0.965*** (0.0643)	-0.615*** (0.194)	0.000653*** (0.000128)	0.000978*** (0.000232)	0.00315* (0.00184)	0.00218 (0.00425)	0.000142** (6.57e-05)	0.000231* (0.000120)	-3.57e-05 (3.83e-05)	-5.47e-05 (7.90e-05)	0.000120*** (3.33e-05)	-2.13e-05 (6.57e-05)
Boundary Dist^2 X Fast Side Dummy	5.34e-05 (0.000250)	-0.000866 (0.000711)	-8.07e-07 (5.00e-07)	-2.44e-06*** (8.64e-07)	-1.03e-05 (7.51e-06)	-7.43e-06 (1.67e-05)	-4.98e-07* (2.61e-07)	-9.59e-07*** (4.72e-07)	3.15e-08 (1.55e-07)	1.73e-07 (3.10e-07)	-3.92e-07*** (1.32e-07)	1.67e-07 (2.53e-07)
Boundary Dist^3 X Fast Side Dummy	3.09e-07 (3.76e-07)	1.81e-06* (1.01e-06)	1.20e-09 (7.35e-10)	3.53e-09*** (1.24e-09)	1.72e-08 (1.14e-08)	1.01e-08 (2.46e-08)	7.17e-10* (3.92e-10)	1.32e-09* (6.99e-10)	-0 (2.33e-10)	-2.63e-10 (4.57e-10)	4.99e-10** (1.99e-10)	-2.87e-10 (3.72e-10)
Boundary Dist^4 X Fast Side Dummy	-1.89e-10 (1.88e-10)	-9.45e-10* (4.90e-10)	-0* (0)	-0*** (0)	-0* (0)	-0 (0)	-0* (0)	-0* (0)	0 (0)	0 (0)	-0** (0)	0 (0)
Boundary Segment FX	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Key Stage-by-Year FX	✗	✗	✗	✗	✓	✓	✗	✗	✓	✓	✓	✓
School FX	✗	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗
Controls (X-vectors in specs)	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗
Quarter FX	✗	✗	✗	✗	✗	✗	✓	✓	✗	✗	✗	✗
Observations	581,594	128,994	178,027	53,077	7,021,667	1,512,011	3,559,352	750,618	7,021,667	1,512,011	7,021,667	1,512,011
R-squared	0.786	0.705	0.819	0.838	0.272	0.266	0.749	0.741	0.138	0.125	0.389	0.314
Number of Boundary Segment Clusters	20,248	6,074	15,997	4,848	19,472	5,832	18,308	5,338	19,472	5,832	19,472	5,832

Notes: The table depicts the estimation results for the full specifications using the fourth-order polynomial functional form. See Sections 4-6 for discussion. The zero estimates are due to rounding limits when saving the point estimates. * 10%, ** 5%, *** 1% significance levels.

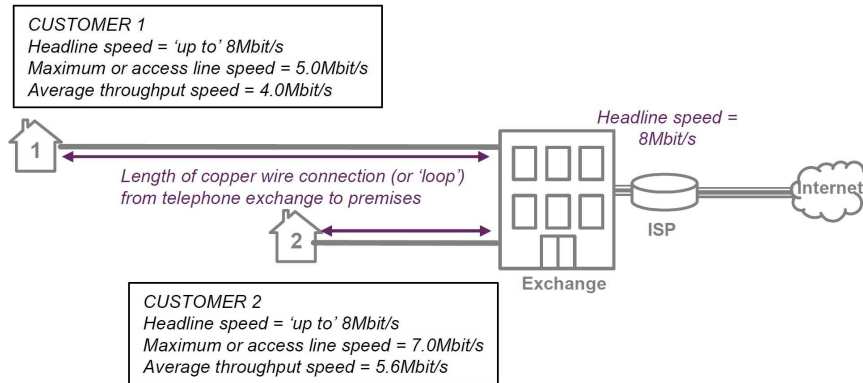
Table 3: Point Estimates (continued)

Dependent Variable	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
Estimation Sample	Supermarket Distance (meters)	Supermarket Distance (meters)	Road Distance (meters)	Road Distance (meters)	School Distance (meters)	School Distance (meters)	Days per Week Online	Weekly Hours Social Media	Weekly Homework Hours	Number of Weekdays Doing Homework	Propensity of Online Resources for H'work	Weekly Study Hours using ICT
Estimation Sample	Full	Top 30% Jumps	Full	Top 30% Jumps	Full	Top 30% Jumps	Full	Full	Full	Full	Full	Full
Fast Side Dummy (Jump)	2.468 (3.352)	7.233 (8.791)	-1.310 (2.602)	-8.955 (6.795)	-2.172 (2.686)	-6.666 (7.184)	0.173 (0.783)	0.204 (2.741)	-0.224 (2.198)	0.0774 (0.164)	0.348 (0.573)	0.0132 (0.0679)
Boundary Dist X Slow Side Dummy	0.0609 (0.0571)	0.144 (0.160)	0.174*** (0.0411)	0.141 (0.117)	-0.0235 (0.0436)	-0.0426 (0.123)	0.00876 (0.00951)	0.0331 (0.0376)	-0.0101 (0.0220)	0.00168 (0.00189)	0.00934 (0.00626)	0.000915 (0.000688)
Boundary Dist^2 X Slow Side Dummy	-0.000510* (0.000260)	-0.000833 (0.000772)	-0.000293 (0.000186)	0.000243 (0.000553)	4.10e-05 (0.000199)	0.000112 (0.000593)	-2.41e-05 (4.05e-05)	-8.97e-05 (0.000167)	2.21e-06 (0.000106)	-0.24e-06 (8.49e-06)	-4.65e-05* (2.57e-05)	-3.04e-06 (2.92e-06)
Boundary Dist^3 X Slow Side Dummy	4.30e-07 (4.28e-07)	7.86e-07 (1.31e-06)	1.12e-07 (3.02e-07)	-6.67e-07 (9.20e-07)	-3.76e-07 (3.26e-07)	-2.82e-07 (1.01e-06)	2.20e-08 (6.27e-08)	4.01e-08 (2.64e-07)	3.73e-08 (1.80e-07)	-1.17e-09 (1.39e-08)	7.97e-08** (3.90e-08)	3.65e-09 (4.57e-09)
Boundary Dist^4 X Slow Side Dummy	-1.80e-10 (2.27e-10)	-3.05e-10 (7.07e-10)	-0 (1.59e-10)	3.06e-10 (4.92e-10)	2.55e-10 (1.72e-10)	9.82e-11 (5.49e-10)	-0 (0)	0 (1.36e-10)	-0 (9.58e-11)	0 (0)	-0** (0)	-0 (0)
Boundary Dist X Fast Side Dummy	-0.00407 (0.0466)	-0.0141 (0.0988)	0.132*** (0.0332)	0.170** (0.0749)	0.0263 (0.0372)	0.0318 (0.0786)	0.00876 (0.00951)	0.0383 (0.0306)	-0.00864 (0.0215)	0.000472 (0.00163)	0.00166 (0.00477)	0.000854 (0.000619)
Boundary Dist^2 X Fast Side Dummy	-0.000298 (0.000195)	-0.000343 (0.000400)	-0.000396*** (0.000140)	-0.000455 (0.000304)	-0.000442*** (0.000157)	-0.000654*** (0.000324)	-2.63e-05 (2.99e-05)	-0.000155 (0.000134)	2.84e-06 (9.94e-05)	-7.56e-07 (6.80e-06)	-7.20e-06 (1.96e-05)	-4.10e-06 (2.51e-06)
Boundary Dist^3 X Fast Side Dummy	1.53e-07 (3.06e-07)	2.82e-07 (6.10e-07)	3.84e-07* (2.16e-07)	4.25e-07 (4.53e-07)	5.22e-07** (2.43e-07)	7.75e-07 (4.91e-07)	2.83e-08 (4.63e-08)	1.98e-07 (2.09e-07)	1.28e-08 (1.62e-07)	5.99e-11 (1.06e-08)	1.40e-08 (2.99e-08)	6.71e-09* (3.81e-09)
Boundary Dist^4 X Fast Side Dummy	-0 (1.56e-10)	-8.82e-11 (3.06e-10)	-1.55e-10 (1.09e-10)	-1.64e-10 (2.22e-10)	-2.45e-10** (1.23e-10)	-3.32e-10 (2.45e-10)	-0 (0)	-7.72e-11 (1.06e-10)	-0 (8.57e-11)	0 (0)	-0 (0)	-0* (0)
Boundary Segment FX	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Survey Round FX	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓
Controls (X-vectors in specs)	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓
Observations	581,594	128,994	581,594	128,994	581,594	128,994	5,046	5,308	3,135	11,020	10,345	10,345
R-squared	0.962	0.970	0.832	0.857	0.866	0.882	0.494	0.455	0.524	0.437	0.358	0.373
Number of Boundary Segment Clusters	20,248	6,074	20,248	6,074	20,248	6,074	1,720	1,597	1,372	2,772	2,709	2,709

Notes: The table depicts the estimation results for the full specifications using the fourth-order polynomial functional form. See Sections 4-6 for discussion. The zero estimates are due to rounding limits when saving the point estimates. * 10%, ** 5%, *** 1% significance levels.

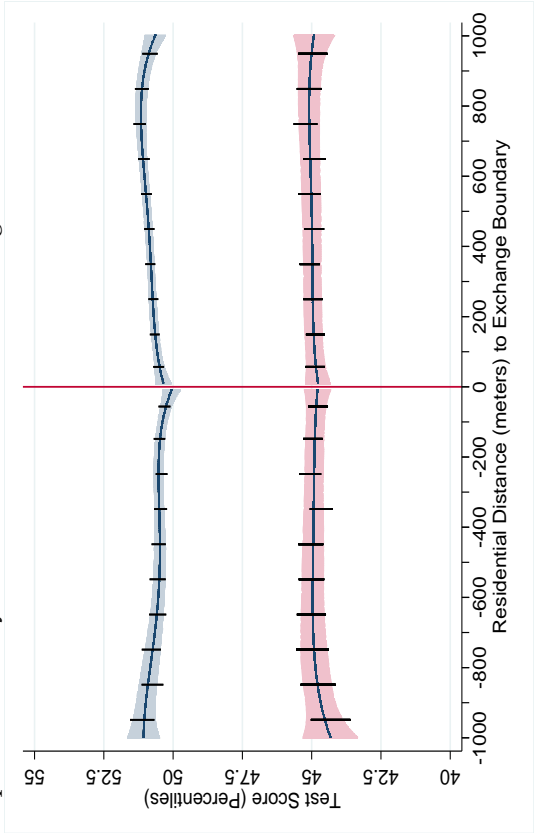
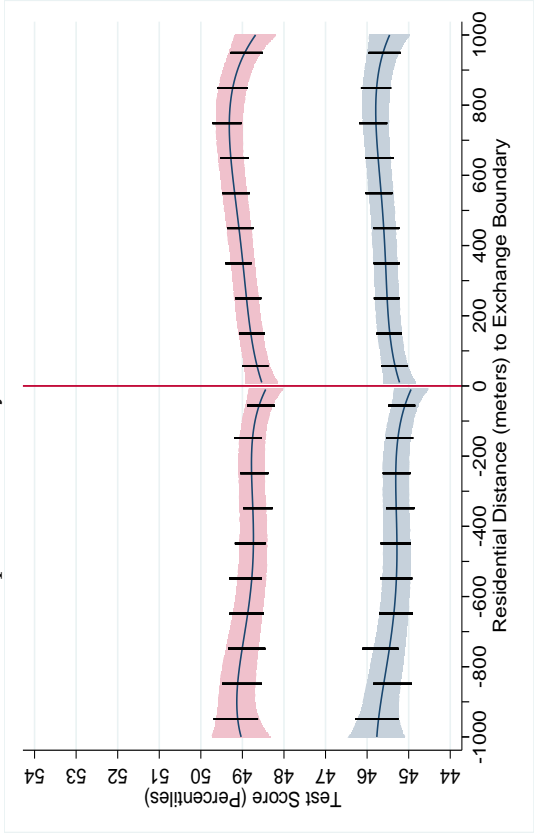
9 Online Appendix

Figure A.1: Technological Relationship between Exchange Distance and Internet Speed

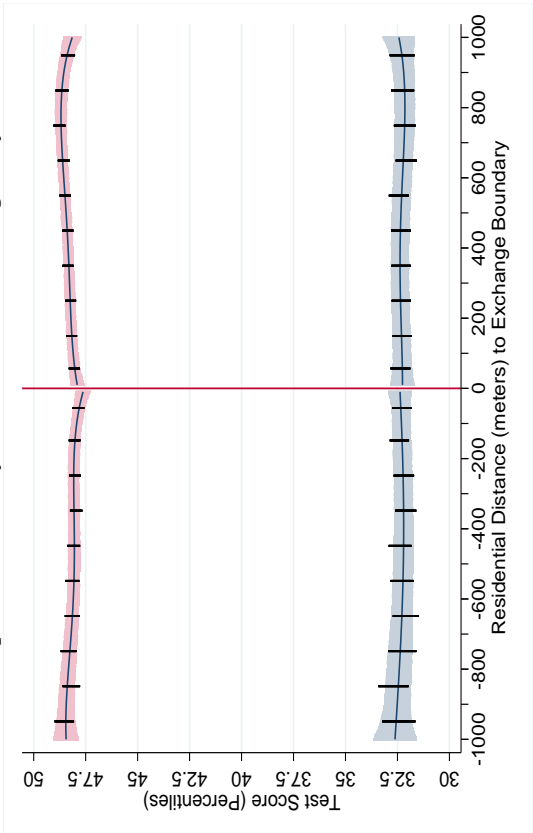


Notes: The Source is England's Telecommunication Regulation Authority OfCom.

Figure A.2: Heterogeneity across Gender, Ethnic Background and Free School Meal Eligibility
Full Specification - By Gender



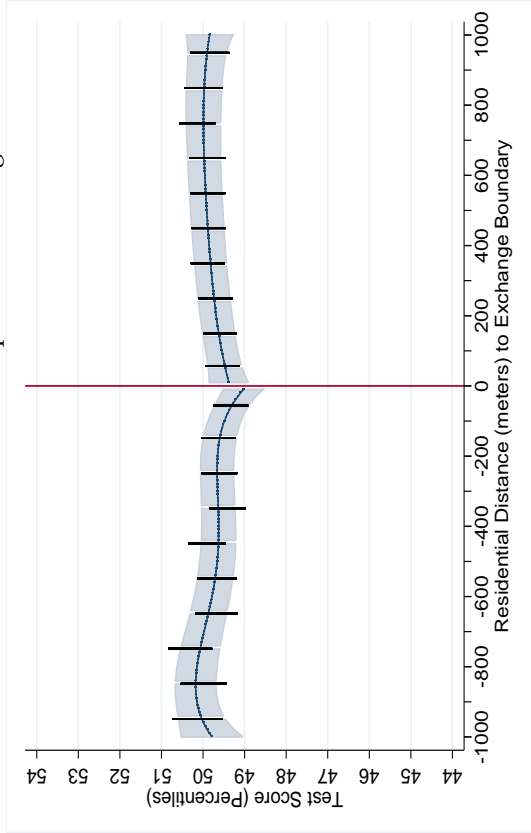
Full Specification - By Free School Meal Eligibility



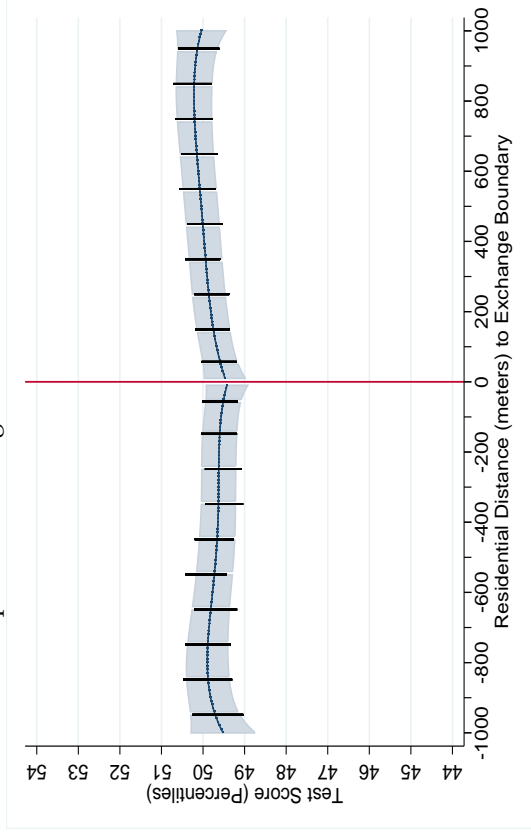
Notes: The left hand side in all graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from four separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 5, and by each group. Within each graph, the groups with the higher average test scores present estimates for (1) females, (2) Whites and (3) students not eligible for free school meals. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Descriptive statistics on the estimation samples are reported in Table 1.

Figure A.3: Heterogeneity across Age Groups

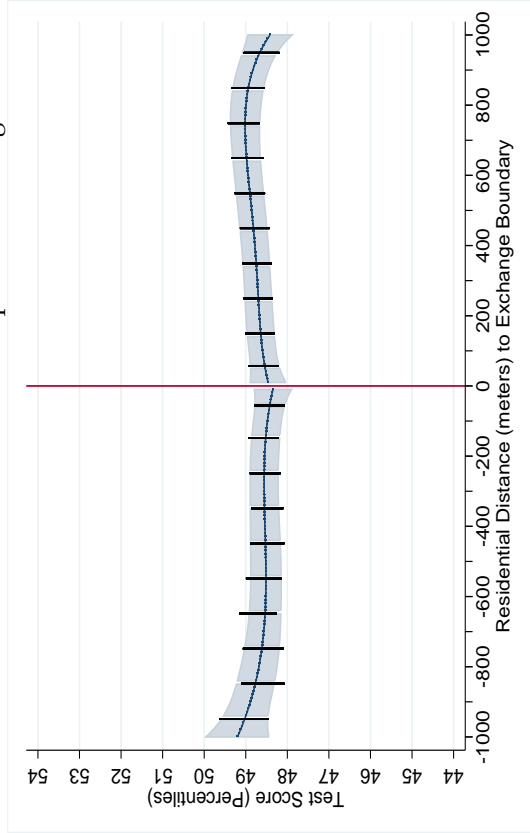
Full Specification - Age 7



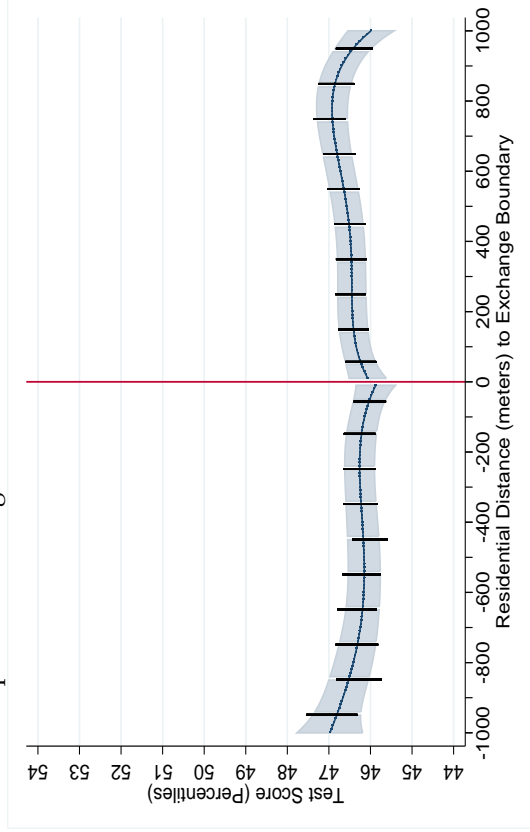
Full Specification - Age 10



Full Specification - Age 14



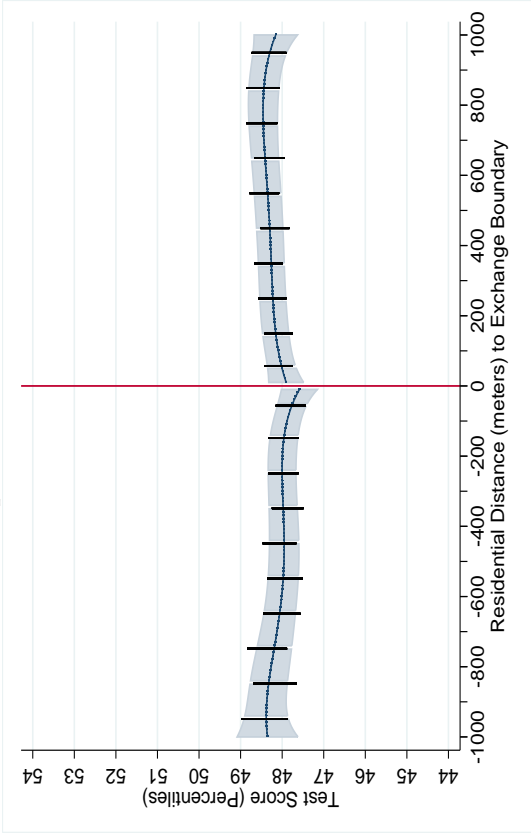
Full Specification - Age 16



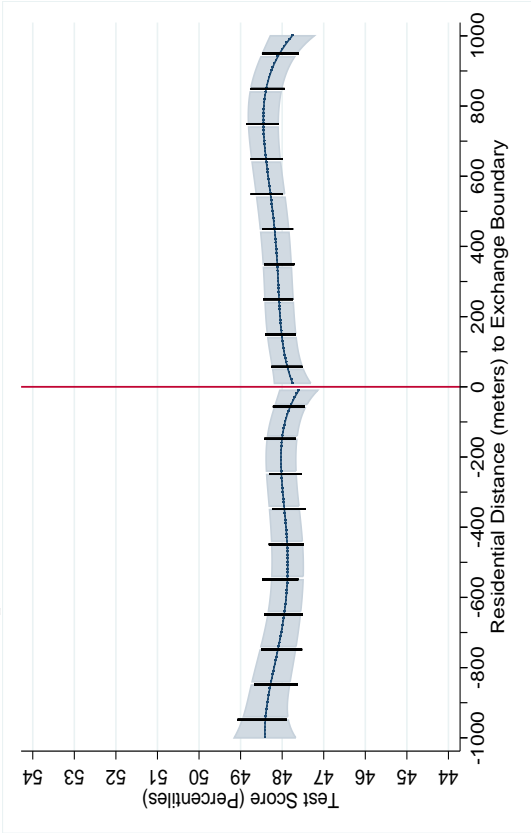
Notes: The left hand side in all graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 5. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Descriptive statistics on the estimation samples are reported in Table 1.

Figure A.4: Heterogeneity over Time

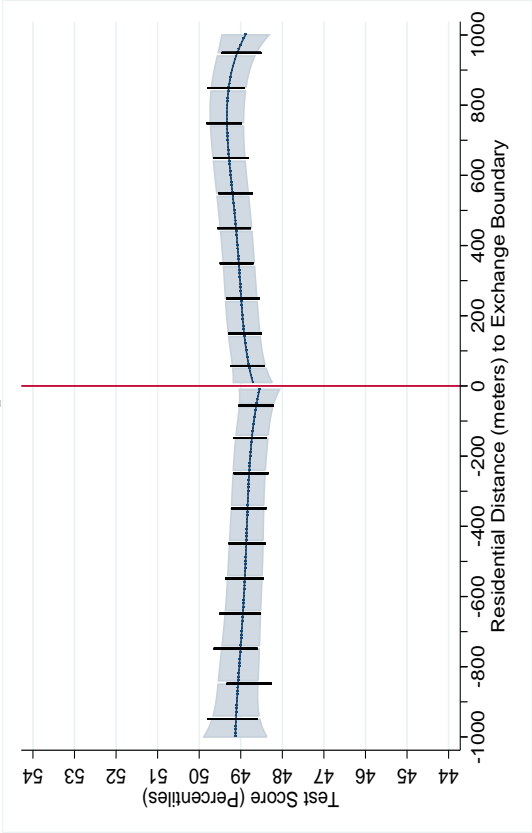
Full Specification - Years 2002 and 2003



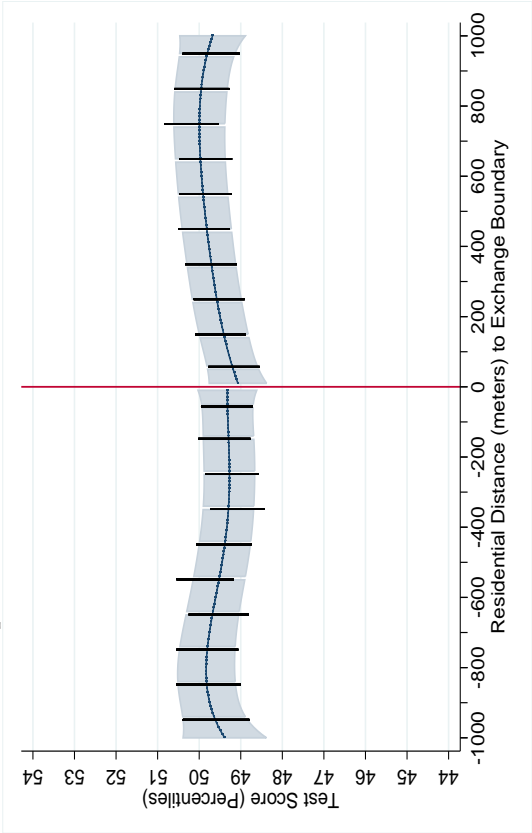
Full Specification - Years 2004 and 2005



Full Specification - Years 2006 and 2007

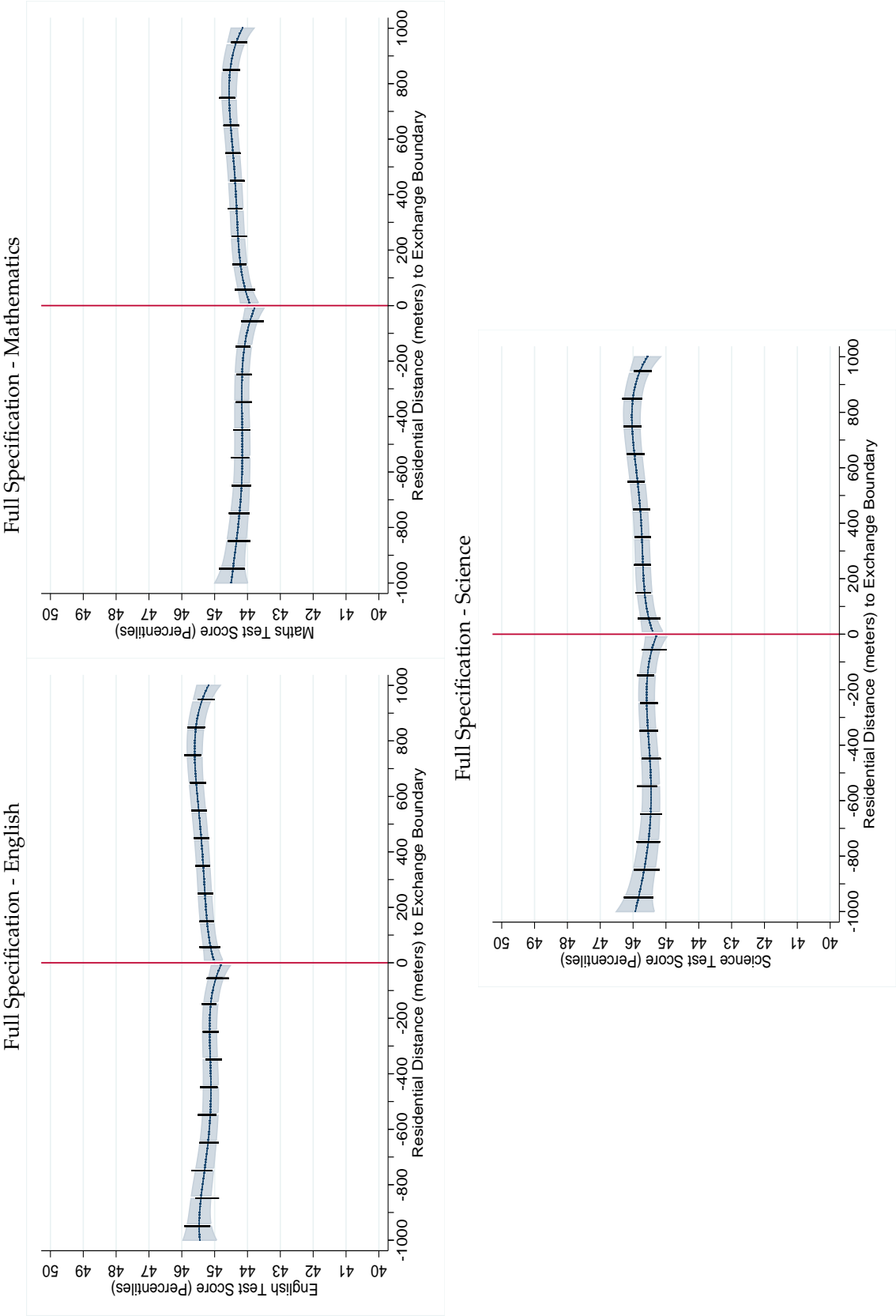


Full Specification - Year 2008



Notes: The left hand side in all graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 5. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Descriptive statistics on the estimation samples are reported in Table 1.

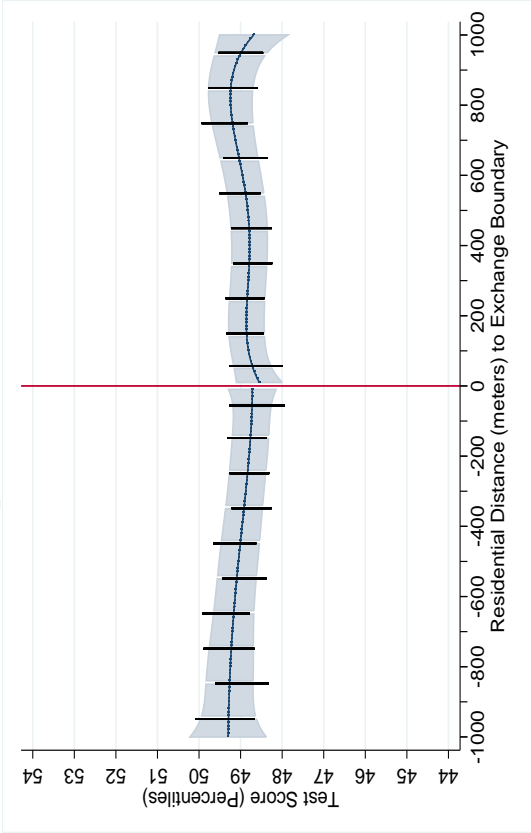
Figure A.5: Heterogeneity across Subjects



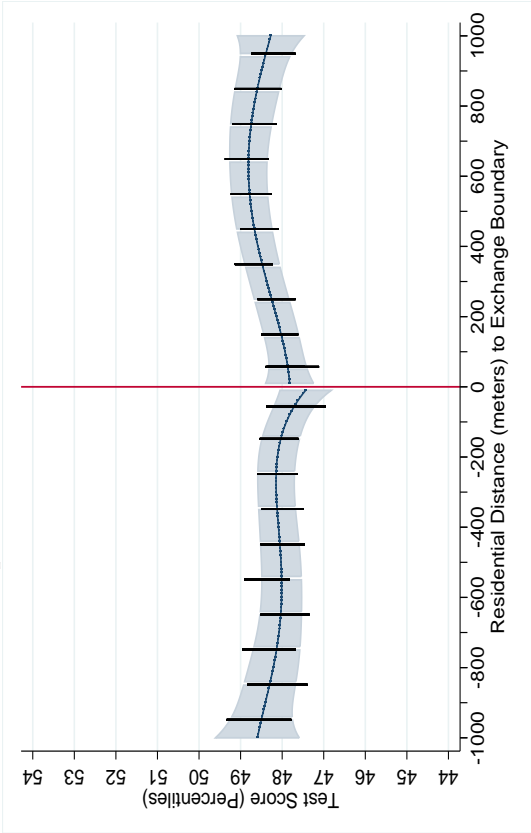
Notes: The left hand side in all graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 5. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments.

Figure A.6: Heterogeneity with Respect to Intensity of Jump

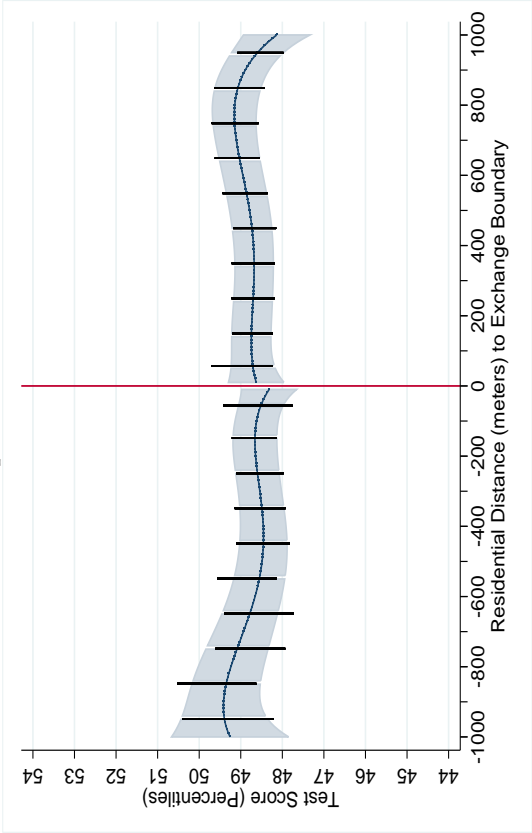
Full Specification - First (Smallest) Quartile



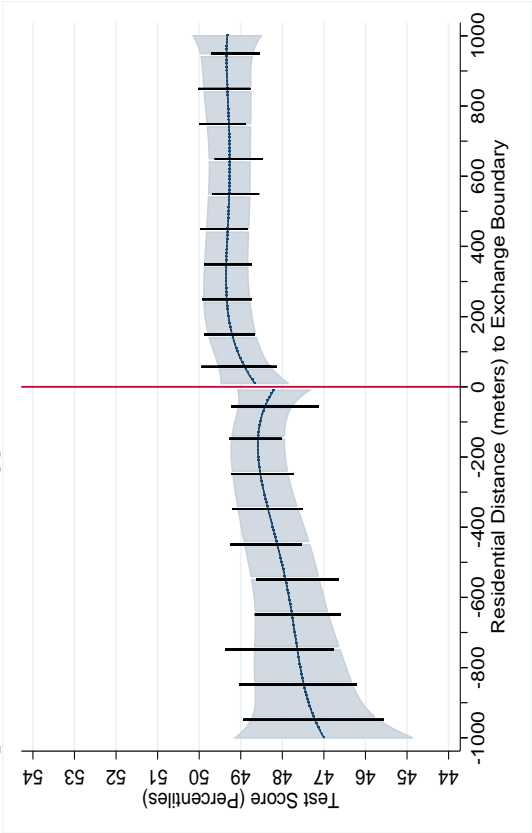
Full Specification - Second Quartile



Full Specification - Third Quartile

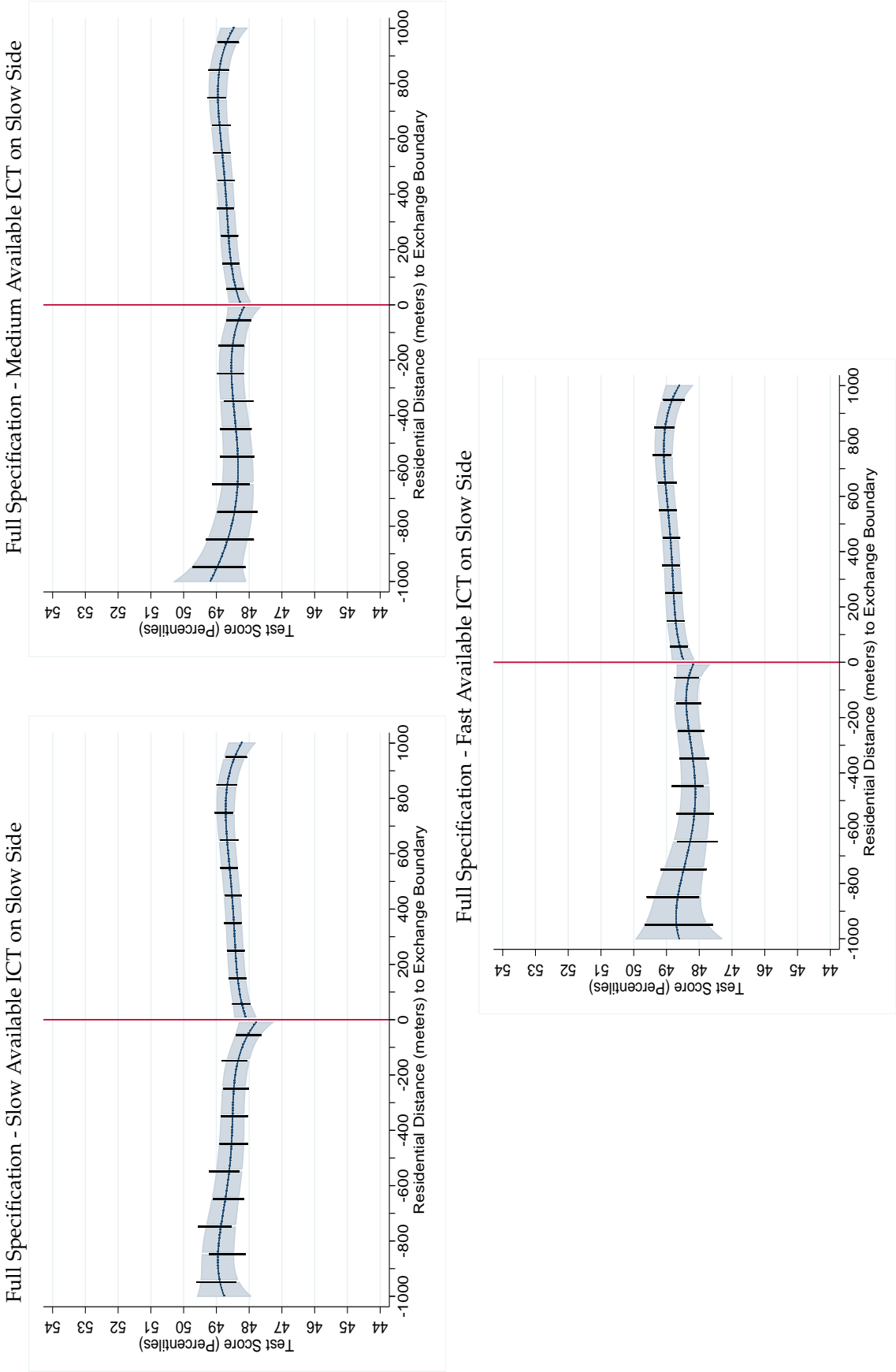


Full Specification - Fourth (Biggest) Quartile



Notes: The left hand side in all graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 5. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Quartiles refer to the distribution of the intensity of the jump across boundaries.

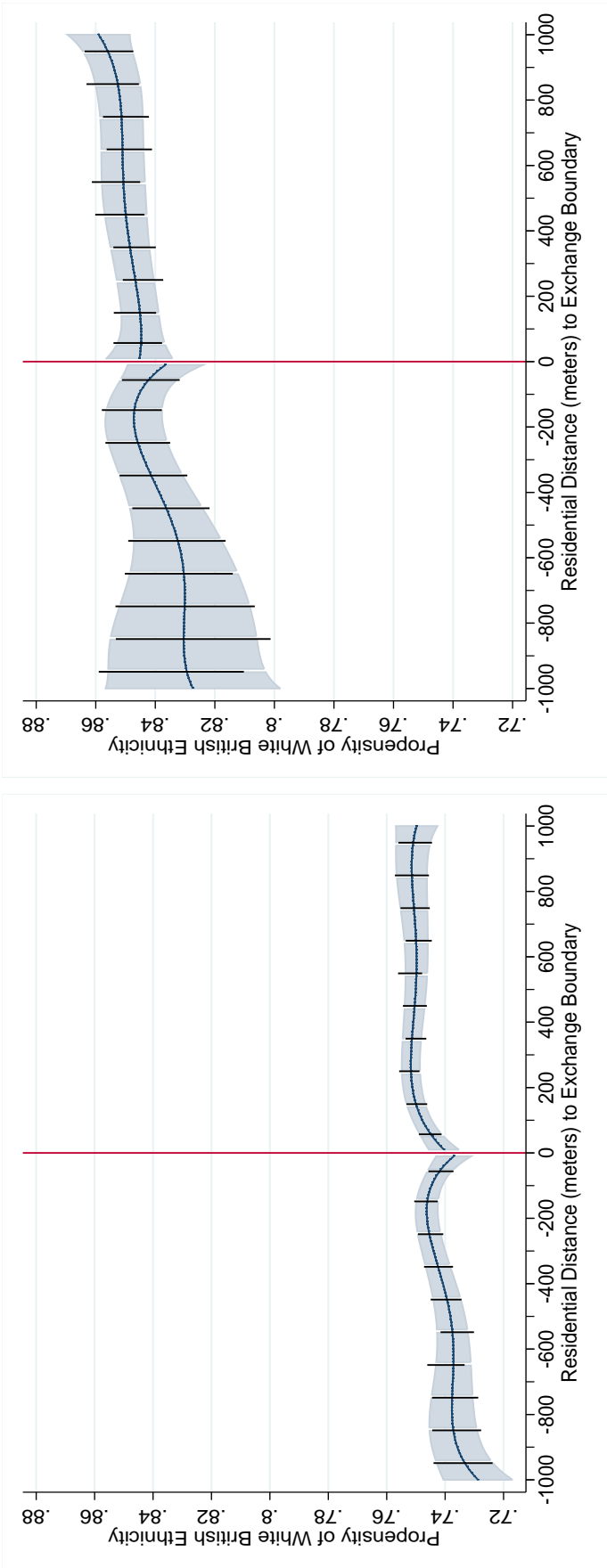
Figure A.7: Heterogeneity across Initial Levels of Available ICT



Notes: The left hand side in all graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 5. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Descriptive statistics on the estimation samples are reported in Table 1.

Additional Robustness Results

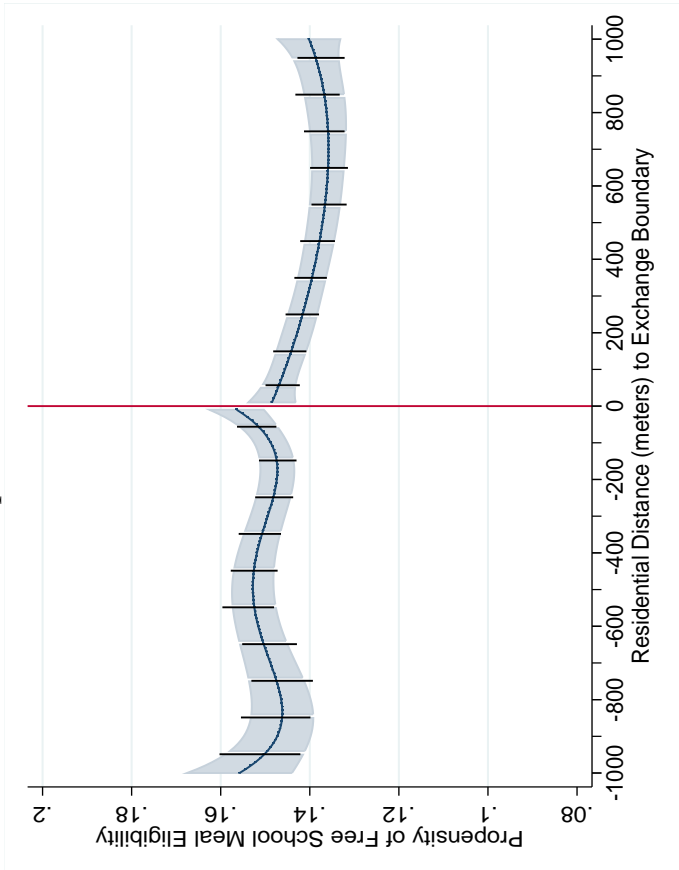
Figure A.8: The Boundary Effect on the Propensity of White British Ethnicity
Full Specification - All Boundaries



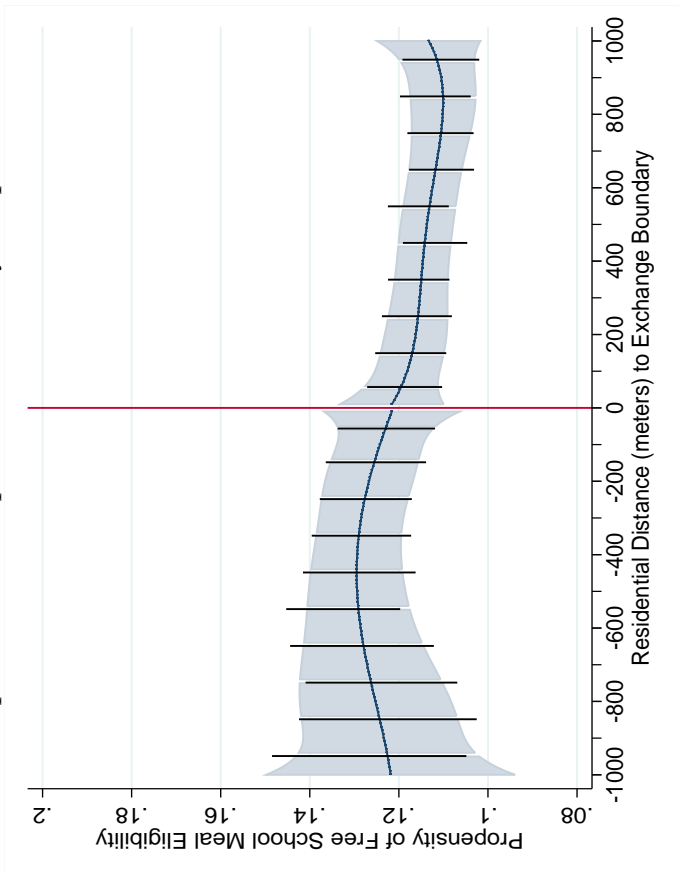
Notes: The left hand side in both graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 5. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Descriptive statistics on the estimation samples are reported in Table 1.

Figure A.9: The Boundary Effect on the Propensity of Free School Meal Eligibility

Full Specification - All Boundaries



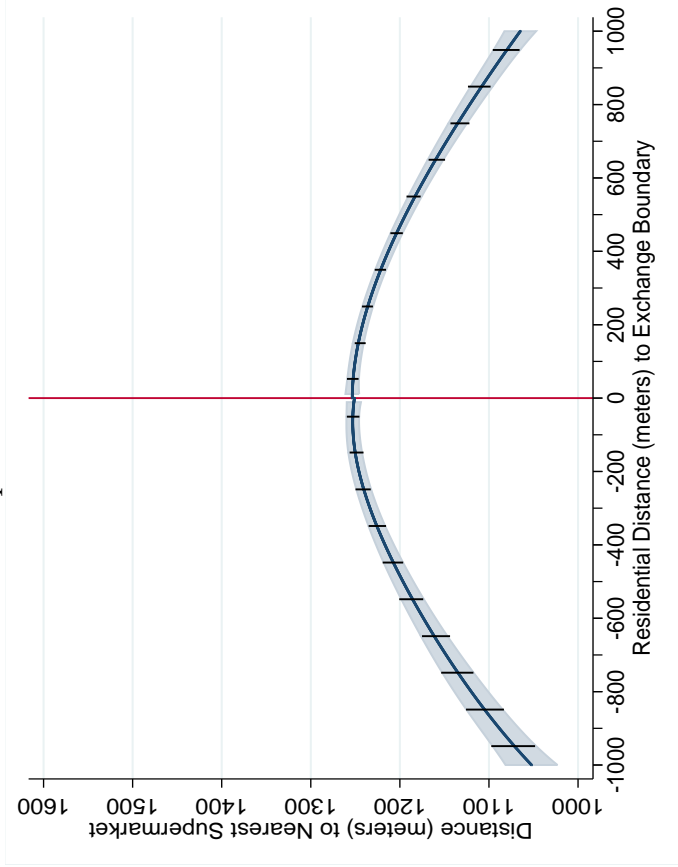
Full Specification - Top 30% of Boundary Jumps



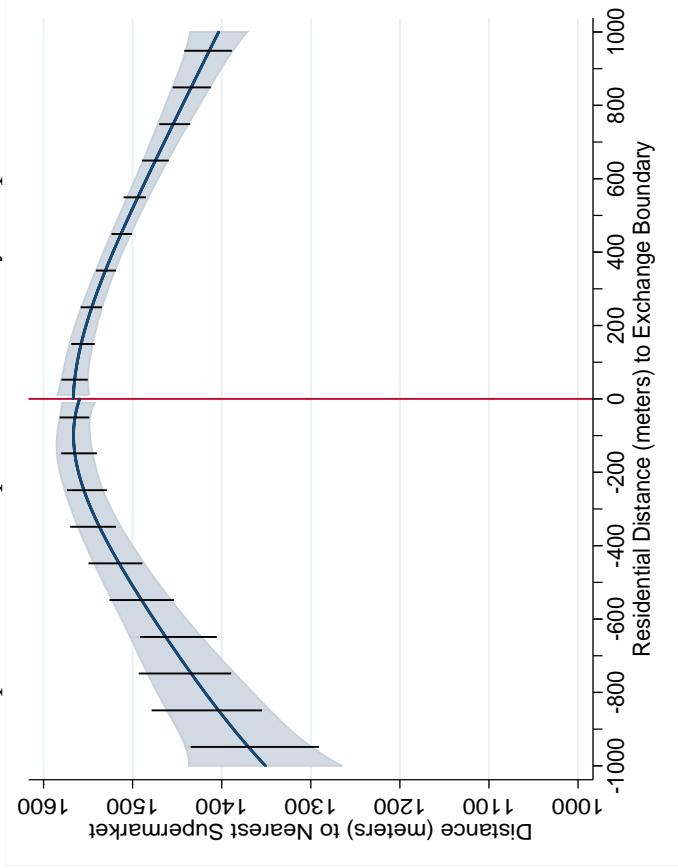
Notes: The left hand side in both graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 5. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Descriptive statistics on the estimation samples are reported in Table 1.

Figure A.10: The Boundary Effect on Local Amenities - Distance to Nearest Supermarket

Full Specification - All Boundaries



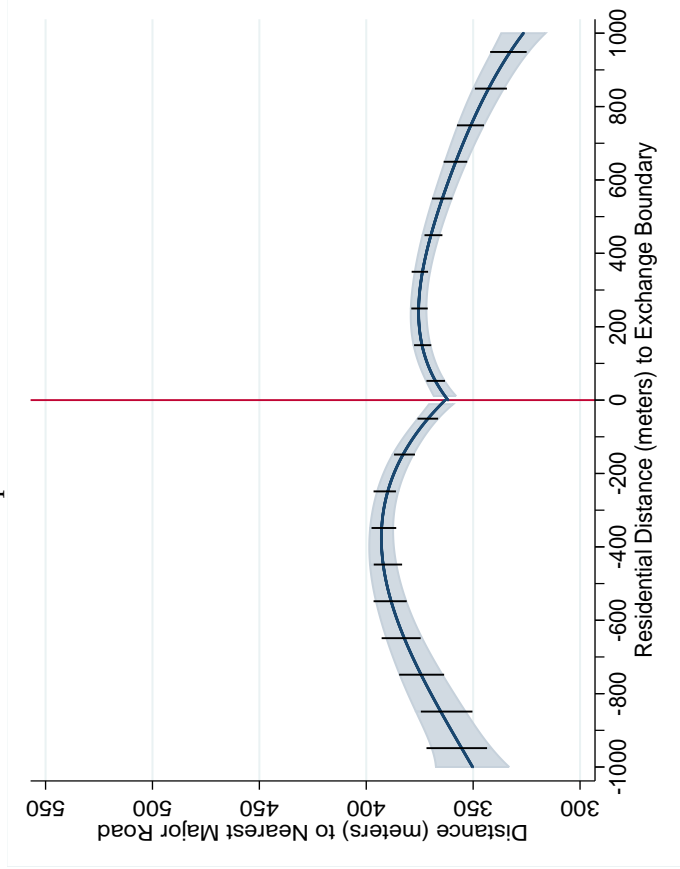
Full Specification - Top 30% of Boundary Jumps



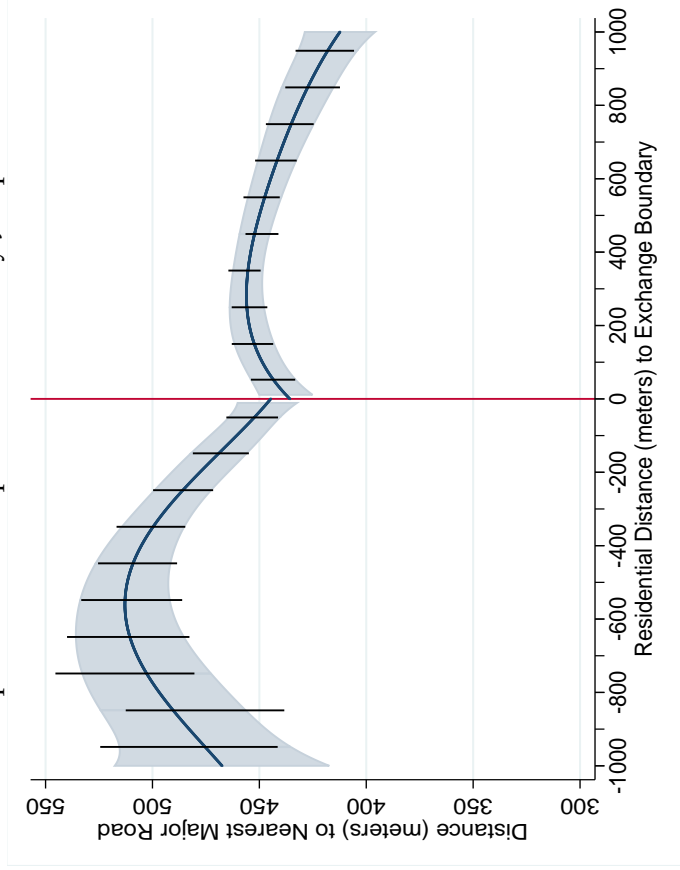
Notes: The left hand side in both graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 5. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Descriptive statistics on the estimation samples are reported in Table 1.

Figure A.11: The Boundary Effect on Local Amenities - Distance to Nearest Major Road

Full Specification - All Boundaries

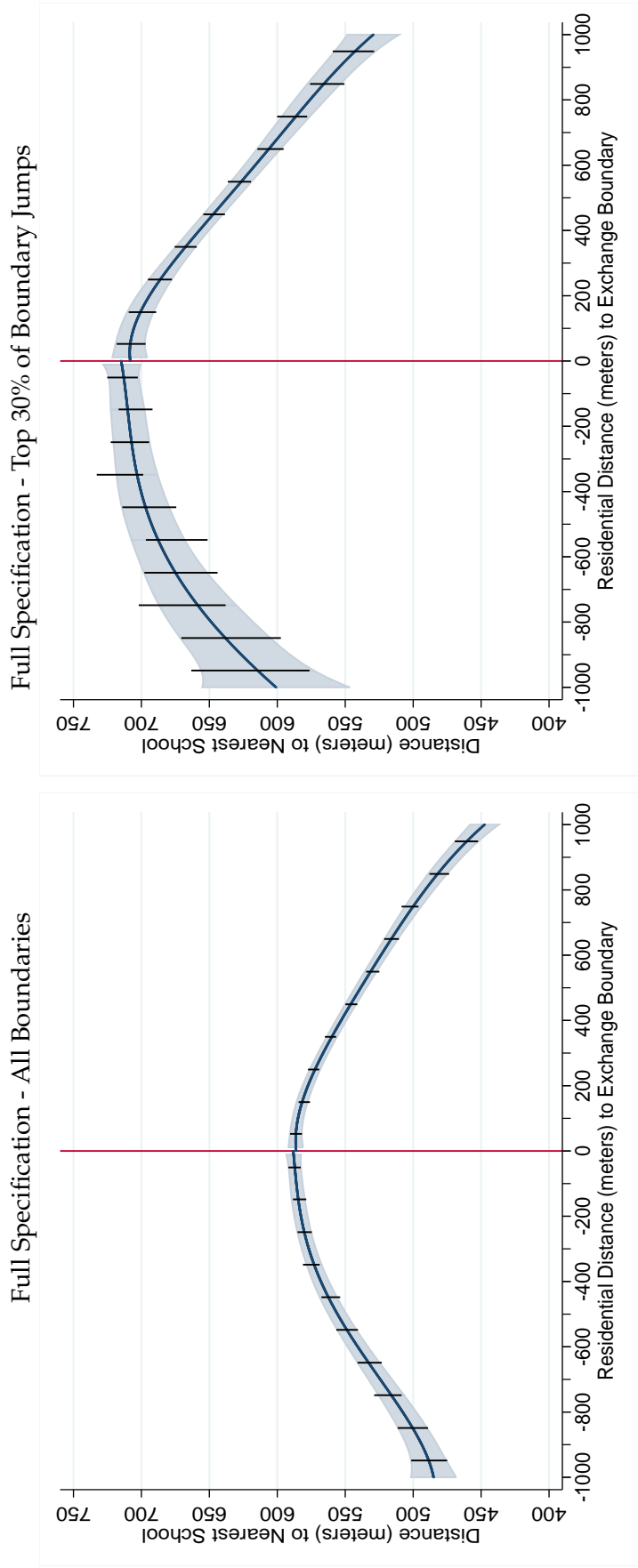


Full Specification - Top 30% of Boundary Jumps



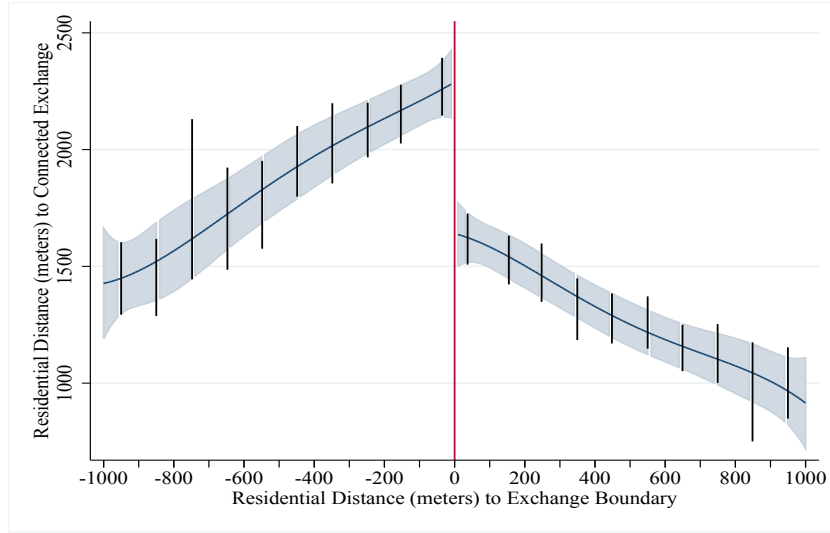
Notes: The left hand side in both graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 5. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Descriptive statistics on the estimation samples are reported in Table 1.

Figure A.12: The Boundary Effect on Local Amenities - Distance to Nearest School



Notes: The left hand side in both graphs refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. Each figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 5. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Descriptive statistics on the estimation samples are reported in Table 1.

Figure A.13: The Boundary Effect in the Survey Microdata Samples



Notes: The left hand side in the graph refers to the slower side of a given exchange boundary segment, i.e. the side with higher mean residential distances to the connected exchange station. The figure depicts estimation results from two separate specifications: The plotted functional forms present results after estimating fourth-order polynomials conditional on boundary segment fixed effects and additional covariates as discussed in Section 5. The shaded area depicts 95% confidence intervals. The black bars in both graphs depict the 95% confidence intervals for the less parametric specification (replacing the polynomials with 100m distance bins). For display in the graphs, these bin coefficients have been centered at the mean distance to the boundary within each 100m bin on the x-axis. Both sets of estimates are depicted as partial predictions while holding covariates constant at their sample means. Confidence intervals are based on standard errors that are clustered at the level of boundary segments. Descriptive statistics on the estimation samples are reported in Table 1.