Essays in Empirical Economics of Digitization

Inaugural-Dissertation
zur Erlangung des Grades

Doctor oeconomiae publicae (Dr. oec. publ.)

an der Ludwig-Maximilians-Universität München

2022

vorgelegt von

Valentin Johannes Lindlacher

Essays in Empirical Economics of Digitization

Inaugural-Dissertation
zur Erlangung des Grades

Doctor oeconomiae publicae (Dr. oec. publ.)

an der Ludwig-Maximilians-Universität München

2022

vorgelegt von

Valentin Johannes Lindlacher

Referent:	Prof. Dr. Oliver Falck
Korreferent:	Prof. Dr. Niklas Potrafke
Promotionsabschlussberatung:	20. Juli 2022

Datum der Promotionsabschlussberatung: 20. Juli 2022

Namen der Berichterstatter: Oliver Falck, Niklas Potrafke, Claudia Steinwender

Acknowledgements

Over the course of the past years, I have enjoyed the tremendous support of many people. All of them deserve being acknowledged, but I can only thank a selection of them here.

I am very grateful to my supervisor Oliver Falck for the confidence he placed in me, although I was coming from a different field. He made it possible for me to write my doctoral thesis in economics and I cannot stress enough how much I appreciate his invaluable advice. I would also like to thank Niklas Potrafke for his academic support from very early on. I further thank Claudia Steinwender for her job market advises and encouragement and for completing my dissertation committee.

I am very grateful to Jean-Victor Alipour and Moritz Goldbeck, with whom I had the great pleasure to share an office. Their contributions as my co-authors and colleagues were instrumental for many parts of this dissertation. Very early advise came also from Andreas Mazat and Tobias Lohse. They helped me at lot at the beginning both professionally and personally. I am thankful for the time we spend together. I would also like to thank Thomas Fackler. Our discussions and collaboration have shaped my academic thinking as an economist. I have greatly enjoyed discussing many aspects in the field of economics, and I am especially thankful for collaborating in a further paper besides this dissertation. I am grateful to have met Lukasz Grzybowski, who invited me to ParisTech for a research stay and who also trusted me early on to become his co-author. To Felix Ehrenfried and Onkokame Mothobi, I am thankful for their collaboration as co-authors. I further thank Sebastian Wichert, Anne Kerkhof, Valentin Reich, Christina Langer, Nicolas Apfel, and many other colleagues and friends, both present and past, for their great support throughout the years.

Finally, yet foremost, I thank my family and my girlfriend Lea. Their unconditional support, encouragement, and patience went far beyond what I could have asked for but were the privilege I needed.

Thank you. Sincerely.

Valentin Lindlacher, March 2022

Contents

Preface

In digital economics, digital technologies shift different costs and these shifts in costs result in an exploration of how standard economic models change [Goldfarb and Tucker, 2019]. Five types of costs are thereby identified: (i) search costs, (ii) replication costs, (iii) transportation costs, (iv) tracking costs, and (v) verification costs. Search costs imply information gathering and reflect the difficulty of the search but also the quality of found information. They are especially present in the economy of platforms but also shape economic thinking in various other aspects. Thus, many parts of this dissertation will investigate indeed these very costs as mechanisms. While lower replication costs are often associated with copyright policies, transportation costs affect the distribution of digital goods and are often paraphrased as a *death of distance* [Cairncross, 1997]. However, they also affect physical goods, which are shopped online and then shipped to the consumer. Tracking costs refer to tracking individuals and relate, for instance, to price discrimination, and relatedly, verification costs allow for building individual reputations but also for discrimination.

This dissertation contains essays starting with classical questions of economic growth and consumer utility and goes further to questions connected to political sciences. It follows digital economics from the effects of broadband Internet at basic speeds in developing countries, through the question of how elastic high-speed Internet demand is in Austria, to the question of further possibilities digital technology allows for in the voting process. Connected to the described costs by Goldfarb and Tucker [2019], the first chapter investigates the effects of lower search costs induced by Internet availability at basic speeds in Sub-Saharan Africa. The second chapter explores whether Internet users in Austria gain a higher utility from increased bandwidth. As many users are very price sensitive at high bandwidths, this means that the reduced information costs at basic speeds are not lowered a lot further with higher bandwidths. Finally, the third chapter analyzes the mechanisms of altered search and transportation costs in voting behavior. The setting is in Munich between 2009 and 2020. Hence, it can be assumed that finding the polling place (or the mailbox for postal voting) and requesting postal voting are already facilitated by Internet search. So far, transportation costs are lower by traditional postal voting. The digital transformation, however, would in theory also allow to reduce transportation costs in the future to zero by voting over the Internet (*I-voting*) or a mobile phone (*mobile voting* or *SMS voting*).

Throughout all three chapters of this dissertation, I exploit regional differences. In the first chapter, these regional differences lie in the Internet availability at basic speeds induced by the location of access points to the national backbone and the timing of their construction. In the second chapter, regional differences in active Internet providers lead to different cost structures of available Internet contracts. In the last chapter, regional differences within a city, Munich, are measured as distances between addresses of eligible voters to their assigned polling place. In such regional studies, an essential complication with regard to the identification of causal effects is the fact that regional differences are often determined endogenously. Therefore, the first chapter exploits a national shock of the exogenous timing of Internet connections through sub-marine cables. The main effect is then identified by an interaction term of country-wide Internet availability and within-country variation of Internet access to the national backbone. The second chapter uses a structural discrete choice model for identification of the effect of Internet contract prices on their subscriptions. In the last chapter, the exogenous variation comes from a natural experiment of changes in the assigned polling places. Here and in the first chapter, the effect is estimated by a difference-in-differences approach. I shortly introduce on a chapter-by-chapter basis, the major questions asked, the key features of the identification strategies pursued, and the results derived throughout this dissertation.

In the first chapter, *Digital Infrastructure and Local Economic Growth: Early Internet in Sub-Saharan Africa*, Moritz Goldbeck and I analyze the economic development of towns and changes in regional industry structure when Internet becomes available at basic speeds. In a difference-in-differences setting, we exploit quasi-random variation in Internet availability induced by the arrival of the 'first generation' of sub-marine cables as the treatment year and the timing of the rollout of the national infrastructure to define treatment and control group towns.[1] Thus, we explore effects of the extensive margin of Internet: the first ever Internet connections. Focusing on incidentally connected towns between nodal cities ((regional) capitals and economic centers), we study how nighttime light emission changes for towns which were connected early through the national backbone network versus towns which were connected later. We analyze 220 towns in 10 Sub-Saharan African countries in the early 2000s and measure short-term local economic growth on a balanced panel.

We find a 7 percent higher growth in light intensity, which can then be translated into a 2 percentage point increase in GDP growth. This result seems to be driven mainly by per capita productivity growth and, if at all, only to a small extent by migration into connected

[1] The exogenous variation coming from sub-marine cables was established by Hjort and Poulsen [2019], who investigate labor market effects of an Internet speed upgrade by the 'second generation' of sub-marine cables.

towns. We can back up this suggestion by a strong increase in the intensive margin of towns' light emissions: Connected towns become brighter on average, which is an indication of an increase in productivity. Additionally, we do not find any effects on population growth. Moreover, we find that in connected regions, the share of manufacturing jobs increases faster, while the share of agricultural jobs declines faster. These changes in the industry structure should be the main channel through which the Internet effect manifests. Here, we enrich our analysis by exploiting survey data from IPUMS-International.

While in the first chapter broadband Internet at very basic speeds and its growth effects are of interest, in the second chapter, *Is High-Speed Internet an Infrastructure of General Interest?*, I investigate differences in the elasticity of demand of broadband Internet at basic speeds and at higher speeds. Moreover, as I move forward in time and to Austria, in 2016 not only higher speeds are present, I also consider mobile Internet as a substitute. The question of the demand elasticity of Internet contracts allowing for a high bandwidth arises automatically when trying to bring together the stark increase in the availability of high-speed Internet and its remaining low adoption rates. Austria has very high adoption rates of Internet at basic speeds.[2] This means that for most users high-speed Internet does not increase their utility. Using a mixed logit discrete choice model, I define choice alternatives along bandwidth and technology (fixed-line and mobile). While the existing literature, e.g., Cardona et al. [2009], Srinuan et al. [2012], and Grzybowski et al. [2014], has mainly investigated substitution between technologies regardless of the bandwidth, primarily for regulation purposes, I add to this literature substitution patterns between technologies at different bandwidths and substitution patterns for different bandwidths within a technology.

I find that high-speed and basic Internet are not substitutes as cross-price elasticities show a different pattern for basic and higher speeds. Users who are satisfied with basic speeds refrain from adopting high-speed Internet even if it is available. I conclude that there is a certain type of users who needs higher speeds. Thus, increasing the availability of high-speed Internet will only increase the use of high-speed Internet if these users live in the area that gets an upgrade. By contrast, users who do not require higher speeds will always choose the relatively cheap basic-speed alternative, regardless of the availability of high-speed Internet. I find that for lower speeds, fixed-line and mobile technologies are substitutes. At higher speeds, these technologies do not function as substitutes. Furthermore, the derived own-price elasticities indicate that the demand is generally very elastic. As users who do not need higher speeds, choose basic speeds regardless of high-speed availability, policy-makers cannot increase the usage of high-speed Internet by solely fostering its rollout.

[2] Over time, the definition of Internet at basic speeds moved on. While in the first chapter Internet at basic speeds means a bandwidth of less than 1 Mbps, in this chapter, higher speeds start at a bandwidth of at least 50 Mbps.

The last chapter, *No surprises, please: Voting Costs and Electoral Turnout*, is only indirectly affected by the digital transformation as I-voting is not possible in Munich. With Jean-Victor Alipour, I investigate how exogenous shocks to voting costs affect electoral turnout in Munich. Moreover, we focus on the habit-forming effects of the changed voting costs. In our setting, voting costs change due to institutional processes. We exploit changes in polling places through newly defined precinct borders prior to each election in Munich and the unavailability of polling places on Election Day, e.g., due to renovation works. Individuals whose polling place is relocated experience changes to their voting costs due to altered distance, i.e., the transportation effect, and unfamiliarity with the new polling location, i.e., the search effect [Brady and McNulty, 2011; McNulty et al., 2009]. Our setting has the unique feature that changes in distance are almost equally distributed between an increase and a decrease. We use precinct-level data on eight elections. The precincts are harmonized over time, such that we have 618 precincts in all elections.

We find that polling place relocations reduce turnout by .46 percentage points on average: in-person voting declines by .75 percentage points and is only partly compensated by a .29 percentage points increase in mail-in voting. Moreover, we estimate that a polling place would have to move on average approximately 38 percent or .35 km closer to the voter to offset the search costs completely. However, the turnout drop appears transitory as mail-in votes balance the decline in in-person votes in subsequent elections. In the following elections, the turnout returns to the original level, such that we can reject the necessary condition for habit formation. Nonetheless, we find a persistent substitution towards mail-in voting. The transitory drop in turnout can be explained by inattentive voters, who only notice the polling place reassignment after the deadline for requesting mail-in has passed. Inattentive voters who would have switched from in-person to mail-in voting will either turn out at the new polling place anyway or abstain from voting. But aware about the change, these voters return to mail-in voting in the subsequent elections, recovering the temporary drop in overall turnout. Some inattentive voters switch to nonvoting today but revert to mail-in voting in ensuing elections. The pattern is consistent with rational choice models of voting. The digital transformation could make voting from home possible and would therefore lower voting costs very close to zero. Moreover, it would allow for absentee voting on Election Day without a prior request and therefore erases inattentiveness.

All chapters of this dissertation are self-contained and can be read independently. They are all influenced by the digital transformation and contribute to the field of regional economics.

Chapter 1

Digital Infrastructure and Local Economic Growth: Early Internet in Sub-Saharan Africa *

1.1 Introduction

In the last decades, the provision of digital infrastructure in many countries enabled widespread access and adoption of modern information and communication technologies (ICT), most prominently the Internet. Evidence shows positive effects of broadband Internet availability on individual-level economic performance [Akerman et al., 2015] and country-level economic growth [Czernich et al., 2011] for developed countries. Hopes are high that Internet access fosters regional economic growth in the developing world as well [World Bank, 2016]. For example, in Sub-Saharan Africa (SSA), where impulses for economic growth are required to fight poverty and deprivation, governments, public-private partnerships, and companies alike invest large amounts of money to bring the Internet to everyone. However, the provision there is more complex and costly due to lacking legacy infrastructure, i.e., fixed-line telephony networks. Hence, SSA countries invested more than 28 billion US-Dollar into their national Internet backbone to date [Hamilton Research, 2020].[1] Despite these enormous investments in Africa's digital infrastructure, a growth effect of Internet in SSA is less obvious than it seems. Low population density apart from a few mega-cities, missing hardware, financial constraints, and a lower willingness to pay lead to low adoption rates [World Bank, 2016]. On the other hand, the potential of Internet seems particularly high in SSA since alternative ICT like fixed-line telephony is largely absent. It is thus crucial to investigate how Internet availability affects regional economic development in developing countries.

* This chapter is based on joint work with Moritz Goldbeck.

[1] Facebook recently announced an effort to build a new Internet-enabled sub-marine cable (SMC) to Africa for one billion US-Dollar [Bloomberg, 2020]. And China plans to invest more than 60 billion US-Dollar in Africa's digital infrastructure as part of its Belt-and-Road initiative [Invesco, 2019].

In this chapter, we examine if there is a causal effect of Internet availability on local economic growth in SSA even at basic speeds. We focus on the initial introduction of Internet in SSA through the 'first generation' of Internet-enabled SMCs starting in the early 2000s. We investigate this effect at the town level to analyze whether potential individual-level effects, found by Hjort and Poulsen [2019] for later high-speed Internet availability, materialize on a more aggregate level as well. Moreover, this allows us to explore whether Internet availability has an effect beyond political and economic centers and thus whether it can affect countries' regional development. We capture the evolution of 220 towns in 10 SSA countries which get international Internet connection before the Internet speed upgrade and which rolled out a national backbone. We measure growth of towns by spatial expansion (extensive margin) and density of economic activity (intensive margin) and interpret these components as pointing more towards population or productivity growth, respectively. Furthermore, we investigate changes in the industry composition as a potential mechanism.

We tap two main data sources. First, we measure local economic growth, the key outcome of interest, using nighttime light (NTL) intensity captured by satellites, a well-established proxy introduced by Henderson et al. [2011] at the country level and validated by Storeygard [2016] on the city level for SSA. To get the local town-level measure, we assign NTLs to individual agglomerations by linking lit pixels to built-up areas of SSA cities and towns from *Africapolis*. Second, we use data on the rollout of national backbones to measure Internet availability of individual cities and towns. The data comes from Hamilton Research [2020] and comprises the geo-location of all Internet access points in SSA. Because data on the establishment year of access points only starts in 2009, we backdate the establishment year of access points to their actual construction year via an extensive review of national backbone deployment projects for each SSA country. This enables us to study the 'first generation' of SMC arrivals, which introduced Internet in SSA for the very first time on a noticeable scale.

To identify the causal effect of Internet availability on local economic growth, we exploit quasi-random variation in the timing of countrywide Internet connections induced by the arrival of the 'first generation' of SMCs in SSA in the early 2000s. This approach was established by Hjort and Poulsen [2019], who exploit an Internet speed upgrade induced by SMCs with higher capacities between 2009 and 2012. We focus on towns that are located between nodal cities (political and economic centers) and are therefore incidentally connected. These towns are relatively small and are primarily connected due to their fortunate location. In a difference-in-differences setting, we define treatment and control group towns using the rollout of the national backbone, which makes Internet available through access points. We assign the treatment status to towns that were connected to the national backbone when the Internet became available countrywide, while the control group consists of towns which get an Internet connection through an access point only some years later. In a two-way fixed effects (TWFE) model with town and country-year fixed effects, we then compare the growth of towns with Internet access at the time when broadband Internet at basic speeds becomes available countrywide for the first time to a control group of similar towns getting access only later. Our key identifying assumption is that treatment and control group towns would have evolved similarly in the absence of treatment. Although this assumption cannot be tested directly, we perform a dynamic event-study specification of our model to show that there are no differences in pre-treatment trends of economic activity between treatment and control group towns. The event-study results are robust to accounting for heterogeneous effects in the staggered timing of the treatment using recently proposed estimators by Roth and Sant'Anna [2021*a*], Callaway and Sant'Anna [2021], and Sun and Abraham [2020] and using less restrictive calendar-year fixed effects instead of country-year fixed effects.

We find that connection to the Internet through an access point leads on average to a 7 percent increase in NTL intensity of SSA towns in the first four years since countrywide connection compared to a control group of similar towns not connected through an access point at that time. Applying the established light-to-GDP elasticity from Henderson et al. [2012], this translates into about 2 percentage points higher economic growth. We then differentiate between growth in the average brightness of lit pixels, which is associated with a higher productivity or density in the towns (intensive margin), and growth in the number of lit pixels, indicating a spatial expansion of towns (extensive margin). We find that towns with Internet access are becoming both brighter and larger. This provides suggestive evidence that towns with Internet access become more productive. We do not find effects on population growth and therefore reject the mechanism of migration. This strengthens our hypnotises of an effect of economic development. Moreover, we find a shift in regional industry shares. In connected regions, manufacturing employment shares increase by around 2 percentage points in comparison to regions getting connected later. These shares are mainly gained from decreasing agricultural employment shares in these regions, though this coefficient lacks statistical significance. This suggests that the increase in economic activity is at least partly a result of a changing industry structure induced by Internet availability.

To ensure that our results are driven by Internet availability, we control for the rollout of mobile coverage (GSM). At the time we are investigating, all countries only had basic mobile coverage enabling calls and SMS but not surfing the Internet. Specifically, 3G coverage, and therefore mobile Internet, was not existing. Additionally, we perform placebo tests with access to other potentially confounding infrastructure, such as roads, railroads, and the electricity grid. Moreover, we perform placebo exercises on the timing of countrywide

Internet connections. We use 1,000 simulations with placebo country-connection years prior to the countries' actual connection year to show that the effect is only present for the actual connection years. For the robustness of our results, we test alternative assumptions about the variance-covariance matrix, including changing the level of fixed effects and standard errors, adding linear time trends at the town level, and applying novel event-study estimators. Furthermore, we can show that the results cannot be explained by ethnic favoritism. Finally, we extend the sample by relaxing some assumptions to assure the external validity of our results and investigate heterogeneous effects by focusing on coastal countries only.

With the notable exception of Hjort and Poulsen [2019], who find sizable positive individual-level effects of an Internet speed upgrade on employment in SSA between 2009 and 2012, causal evidence on the economic impact of Internet availability in developing countries is surprisingly rare. This is the first study investigating the causal effect of the introduction of (fixed-line) Internet availability at basic speeds on overall local economic growth in developing countries, which is measured by NTL satellite data. Furthermore, we study early Internet effects in a rural developing country setting with no pre-existing fixed-line telephony network, low penetration rates, and labor-intensive local economies.

We contribute to two main strands of the literature. First, we add to the broad literature assessing the impact of infrastructure on economic outcomes. Our study is the only one investigating the overall impact of (fixed-line) Internet availability on local economic growth for developing countries when Internet becomes available for the first time.[2] Most closely related to our work is Hjort and Poulsen [2019], who study the employment effects of broadband Internet on an individual level when broadband capacity increases. They find a skill-biased and net positive employment effect for an Internet speed upgrade in SSA around 2010. Our analysis contributes to these findings by showing that the benefits of digital infrastructure are present not only at the individual level but at the more aggregate town level and for an overall measure of economic activity as well and even at basic speeds.[3]

For developed countries, the effect of digital infrastructure and especially (broadband) Internet has been assessed widely. Czernich et al. [2011] identify an effect of broadband infrastructure on annual per capita growth for OECD countries. For the US, Kolko [2012] finds a positive relationship between broadband expansion and local economic growth, i.e.,

[2] Hjort and Tian [2021] give a comprehensive overview of the effects of Internet connectivity in developing countries, dividing this literature into supply-side and demand-side mechanisms and overall impact of connectivity.

[3] There is also a large body of related literature on the effect of nondigital infrastructure on economic outcomes in developing countries. Assessed infrastructure includes transportation infrastructure [see e.g., Storeygard, 2016; Ghani et al., 2016; Banerjee et al., 2020; Faber, 2014; Donaldson, 2018], electrification [see e.g., Dinkelman, 2011; Grogan and Sadanand, 2013; Rud, 2012].

growth in population, employment, average wage, and employment rate.[4] While Internet speeds and the timing are very comparable to our setting, adoption rates were a lot higher in developed countries, mostly because pre-existing fixed-line telephony infrastructure made household DSL connections a lot easier. We add to this literature by showing that Internet availability benefits regional economic development also with low adoption rates. This implies that, if a few adopters generate such an effect that it is measurable at the aggregate, the Internet must have great spillover effects.

Related to Internet are mobile phones. Jensen [2007] shows that the adoption of mobile phones led to a reduction in price dispersion and an increased consumer and producer welfare. In a related paper, Aker and Mbiti [2010] study how the introduction of mobile phones between 2001 and 2006 affected grain prices in Niger. These papers emphasize the importance of rolling out mobile network infrastructure for improving economic efficiency of markets. More generally, mobile communication offers a major opportunity to advance economic growth in developing countries, for example by providing information about prices, improving the management of supplies, increasing the productive efficiency of firms, reducing transportation costs, and other means [Aker and Mbiti, 2010]. Fixed-line Internet, as we analyse, might work through the same channels but accesses international information sources.

Second, our work contributes to the literature on urban and regional development. Starting with Nunn and Puga [2012] who showed that in Africa less fortunate geography has a positive impact on today's economy and Henderson et al. [2012] who indicated that the hinterland grows faster than coastal areas and that primate cities do not grow faster than their hinterland, a strand of literature focuses on the catch-up from secondary to primate cities, with no conclusive results. While many papers show that secondary cities are meaningful to reduce poverty [see e.g., Christiaensen and Todo, 2014; Christiaensen and Kanbur, 2017; Fetzer et al., 2016], Bluhm and Krause [2018] show with an adjustment for top coding in NTLs that primate cities remain the economic centers. We contribute by focusing on even smaller towns and showing that even there economic development is happening.

In Section 1.2, we provide a brief overview of early Internet in SSA. Section 1.3 lays out the empirical strategy and in Section 1.4 the data is described. Results are presented in Section 1.5. Section 1.6 discusses our results in comparison with related research. Section 1.7 concludes.

[4] Moreover, labor market effects [see e.g., Atasoy, 2013; Czernich, 2014; Akerman et al., 2015] and effects on firm productivity [see e.g., Akerman et al., 2015; Grimes et al., 2012; Colombo et al., 2013] both with mixed results were investigated.

1.2 Background

There are three major components of Internet infrastructure determining the availability and bandwidth of Internet in a given location. First, international fiber-optic sub-marine cables (SMCs) connect SSA countries to the global Internet backbone.[5] Second, within-country inter-regional fiber-optic cables form the national backbone. Precondition for Internet availability in a location is an access point to the national backbone close by. Finally, individual users in a location are reached via the ‘last mile’ infrastructure.

1.2.1 International Backbone: Sub-Marine Cables

Since the vast majority of web pages and applications is hosted on servers located in North America or Europe, almost all African Internet traffic is routed inter-continentally [Kende and Rose, 2015; Chavula et al., 2015]. Before the first SMCs landed on SSA shores, the only way to connect to the Internet on the continent was via satellite.[6] While being largely unconstrained by geography and local infrastructure, satellite connection is costly and allows only for very narrow bandwidths. With SMCs, a joint effort of governments, private investors, and/or multinational organizations, an Internet connection was first brought to SSA at a noticeable scale.

As shown in Figure 1.1, the first wave of Internet-enabled SMCs arrived in SSA countries only in 1999 and the early 2000s. These ‘first-generation’ cables had the capacity to provide Internet at basic speeds.[7] The biggest of them was SAT-3 and started operating in 2001. It featured landing points on the shores of nine SSA countries on the western coast of Africa.[8] These landing points, typically one per country, constitute the starting point for the respective national backbones (cf. Section 1.2.2). Until the late 2000s, most SSA countries were connected to the Internet via these ‘first-generation’ SMCs.[9]

Landlocked countries are only indirectly connected through SMCs. They rely on their neighboring countries which connect them through a national backbone. The rollout of these inter-regional fiber-optic cables is explained next.

---

5 We define Sub-Saharan Africa as the mainland of the African continent without the Northern African countries, Algeria, Egypt, Libya, Morocco, Tunisia, and Western Sahara. Moreover, we exclude South Africa as it is economically more developed and therefore less comparable to the other SSA countries.

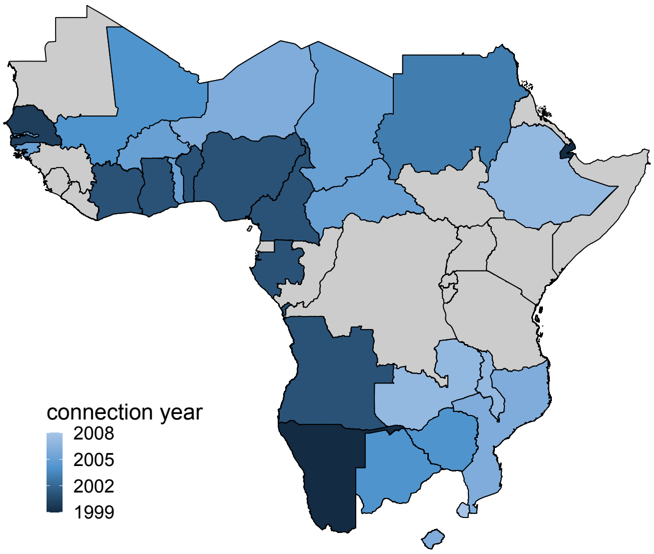
6 Single-channel and co-axial SMCs for telegraphy and telephony already existed before. The first telegraphy cable (‘East coast’ cable) started operating as early as 1879.

7 Hjort and Poulsen [2019] state that SSA users had on average 430 Kbps before the ‘second generation’ of SMCs arrived. In Benin, for instance, ADSL connections with up to 2 Mbps were possible before the upgrade SMC arrived [Agyeman, 2007]

8 These countries are: Angola, Benin, Cameroon, Côte d’Ivoire, Gabon, Ghana, Nigeria, Senegal, and South Africa. It started in Sesimbra, Portugal, and Chipiona, Spain, and also passed the Canary Islands in Alta Vista.

9 The ‘second-generation’ of SMC landed very similarly between 2009 and 2012.

Figure 1.1: Internet connection years



Notes: The figure shows SSA with all countries getting an Internet connection before 2008. The color gradient depicts the connection year (darker blue colors indicate earlier initial SMC connection years). Gray indicates countries not connected to the Internet until 2008.

1.2.2 National Backbone: Inter-Regional Cables

After being routed through an SMC, Internet traffic travels through the national backbone. The national backbone infrastructure consists of inter-regional fiber-optic cables. Therefore, as soon as a new SMC arrives at a landing point of a SSA country, Internet becomes available countrywide in every location with access to the national backbone. As Internet capacity, i.e., speed, of the national backbone does not depend substantially on distance to the landing point, this upward shift occurs uniformly across the country's connected locations. In the last two decades, national backbones were continuously improved and expanded in parallel with the installation of SMCs.[10] This backbone expansion focused heavily on connecting economically and/or politically important locations since they feature the largest market potential (high population density and GDP per capita).[11] This often led to a backbone evolution where the national capital (often a coastal city and located closely to the landing point) was connected first. Then, the backbone spread out to the next largest or politically important cities. Due to their role as nodes in the national backbone networks, we call these cities 'nodal cities'.

[10] Many of these cables were constructed decades ago as part of the telegraph and telephone infrastructure and were only later used for the transmission of early Internet traffic. They typically have been installed by the national telecom. Each country typically has an own, self-contained backbone. There are no network operators owning backbones in more than one country.

[11] Routes establishing connections to (landlocked) neighboring countries are a focus of backbone expansion as well.

Inter-regional cables are almost always constructed along pre-existing infrastructure, e.g., roads, but also railroads, the electric grid, and pipelines, to minimize construction costs. Even though the goal was to connect nodal cities, in many cases, towns on the route of inter-regional cables got Internet access as well due to their fortunate location between two nodal cities. Our empirical strategy (cf. Section 1.3) focuses on these incidentally connected towns which get an Internet connection because of their location next to an inter-regional cable.

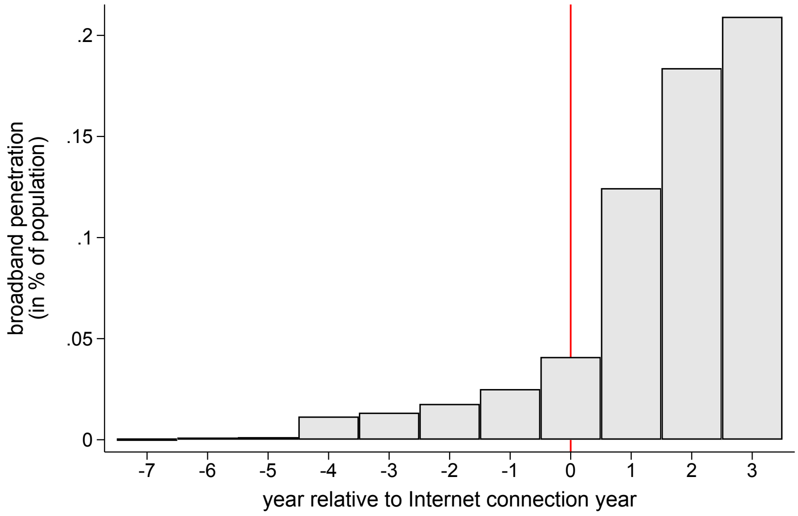
1.2.3 Local Transmission: 'Last Mile' Infrastructure

Internet traffic transported by inter-regional cables is accessed at access points. There are several technologies transmitting Internet traffic from these access points to the user. These 'last mile' transmission technologies include fiber cables (FTTH/B), copper cables, and wireless transmission using cellular towers (e.g., mobile or WiMax). Unlike in many developed countries which rely heavily on transmission to the end user via pre-existing telephony cable infrastructure, in SSA countries households are seldomly connected through copper or even fiber cables. Instead, traffic data is exchanged wirelessly. For this technology, no local cable network connecting each user's exact location (firm, household) is needed. Relative to the costs to construct an inter-regional cable, it is thus cheap to establish Internet access along the cable, making it profitable for the network operator to establish access points even in on-route towns, which are typically much smaller than nodal cities.

Figure 1.2 shows how the usage increases in countries that were served by a 'first generation' SMC. Though, the change in absolute numbers is rather low (.2 percentage points), one should notice that the increase starts when the Internet becomes available. Compared to the year before Internet was accessed through an SMC, but only via a satellite connection, broadband penetration increased by eight times in only four years. Although broadband penetration is low among the population, anecdotally most users access the Internet through cybercafés. Thus, the share of Internet users might be a lot higher than the penetration in Figure 1.2 suggests. Moreover, the broadband penetration in firms might be a lot higher. Although data on broadband adoption of SSA firms does not exist on a wide scale for that time, the *World Bank Enterprise Survey* shows even before the 'second generation' of SMCs landed on SSA shores that 52 percent of all firms used email for communication and 23 percent had an own website.[12]

[12] https://www.enterprisesurveys.org

Figure 1.2: Internet connection and adoption



Notes: Adoption rates are calculated relative to the establishing year of the Internet connection in each country and then aggregated taking the weighted mean. Weights are population size in 2000.

1.3 Empirical Strategy

We are interested in the relationship between Internet availability and local economic growth. However, their correlation is not informative about the causal effect of Internet availability on local economic growth due to endogeneity concerns. In particular, towns with and without Internet access might be very different as Internet access is not randomly assigned and likely driven by commercial interest and/or political and administrative planning.

To address these endogeneity concerns, we leverage a distinct feature of Internet infrastructure evolution in SSA countries. First, we use plausibly exogenous time variation in connections to sub-marine cables (SMCs), which determine Internet availability countrywide for coastal countries, to investigate the effect of Internet availability. Following Hjort and Poulsen [2019], we argue that the exact timing of SMC arrival is essentially random.[13] The arrival is exogenous, first, because each SMC typically connects many countries. Therefore, coordination difficulties among consortium members might delay the construction.[14] Second, the connection years are highly uncertain due to unforeseen delays in construction. For example, the cable EASSy was delayed by five years due to coordination difficulties among consortium members [Poppe, 2009]. Moreover, a country's geographical location within SSA can influence the connection year. First, Eastern and Western SSA

[13] This exogeneity was also exploited by Cariolle [2021].

[14] Consortium investors usually are public and private telecom operators and neighboring and foreign investors [Jensen, 2006].

countries get independently their respective SMCs. Second, landlocked countries get their connection through the national backbone of their neighboring countries and rely therefore on the construction speed of another country's national backbone. This construction speed again is exogenous for the respective landlocked country.

We estimate the effect of early Internet at basic speeds and exploit the arrival of the 'first generation' of SMCs. When the next generation with higher capacities arrives, starting in 2009, countries immediately get a speed upgrade. Therefore, we estimate on a sample containing only years for which countries did not receive a speed upgrade yet. Due to the staggered timing of the 'second generation' of SMC, this sample is unbalanced. To estimate on a balanced sample, we restrict the estimation to three post-treatment years.

In a difference-in-differences (DiD) design, we compare towns that already have access to the national backbone when Internet becomes available countrywide to a control group of similar towns getting an access point in later years (first difference) before and after the country's Internet connection (second difference). The definition of towns in the treatment and control group is depicted in Figure A.1. We exclude nodal cities, i.e., cities close to an access point that are endogenously connected (cf. Section 1.2.2): the landing point, the capital, regional capitals, and economic centers (cities with a population of more than 100,000 inhabitants). For robustness, we vary the threshold of 100,000 inhabitants as the definition of an economic center.

All towns in our analysis get connected eventually, mainly because of their favourable location between nodal cities. Hence, towns that are still waiting for an access point today will only serve for robustness. As an additional robustness check, we vary the last possible year of connection for the control group, such that treated towns are not compared with very late connected towns. Moreover, towns being connected after the Internet became available countrywide but in the estimation period (up to three years after the treatment year) are excluded as they would contaminate the control group. They do not get the full treatment and would thereby confound our analysis. In a robustness check, we define them as treated with the access point construction year as treatment year.

The basic model used to identify the average treatment effect on the treated (ATT) of Internet availability on local economic growth is given by

$$y_{c(i)t} = \beta_0 + \beta_1\big(connection_{ct} \times access_{c(i)}\big) + \beta_2 GSM_{c(i)t} + \alpha_{c(i)} + \delta_{ct} + \epsilon_{c(i)t}, \tag{1.1}$$

where $y_{c(i)t}$ is economic growth of town i in country c in calendar year t as proxied by nighttime light (NTL) intensity (cf. Section 1.4), where the logarithm is used to estimate changes in the growth rate instead of changes in levels. The dummy variable $connection_{ct}$ indicates if country c has a countrywide Internet connection in calendar year t. The variable

$access_{c(i)}$ is one if town i in country c is located within 10 kilometers distance to an access point that was established in the year when the Internet became available countrywide or before. Contrary, the indicator is zero if town i in country c is located within 10 kilometers to an access point that was established in the years afterwards. Thus, the interaction term $connection_{ct} \times access_{c(i)}$ indicates Internet availability in town i in country c in calendar year t. The coefficient of interest is β_1. It captures the effect of Internet availability on local economic growth of early versus later connected towns. For robustness, we vary the distance including smaller and higher values than 10 kilometers. The control variable $GSM_{c(i)t}$ contains time-varying mobile coverage as the share of a town's area covered by GSM technology. We include two types of fixed effects into the model. Time-constant differences across towns are captured by town fixed effects $\alpha_{c(i)}$. Differences across calendar years common to all towns within a country are absorbed by country-year fixed effects δ_{ct}. Note that this allows for country-specific time trends, especially country-specific growth rates, and variations in satellite sensor quality over years. Like in many other DiD applications, our panel data are serially correlated in the time dimension. Hence, we use cluster-robust standard errors whereby we cluster at the town level.

The key identifying assumption for this DiD model is that treatment and control group towns would have evolved similarly in the absence of the treatment (parallel-trends assumption). This assumption cannot be tested. Its plausibility can, however, be examined by testing for pre-treatment differences in time trends between the treatment and the control group. Therefore, we analyze the dynamic impact of Internet availability on local economic activity using an event-study design:

$$y_{c(i)t} = \beta_0 + \sum_{j=\underline{T}}^{\overline{T}} \beta_{1j}\big(t_j \times access_{c(i)}\big) + \beta_2 GSM_{c(i)t} + \alpha_{c(i)} + \delta_{ct} + \epsilon_{c(i)t}, \tag{1.2}$$

where t_j indicates the year relative to treatment year, i.e., the year when the Internet became available countrywide, starting in relative year $j = \underline{T}$ and ending with relative year $j = \overline{T}$. The treatment year is normalized to $j = 0$. We exclude $j = -1$ as the reference point. Thus, the interaction $t_j \times access_{c(i)}$ indicates if town i in country c is part of the treatment group and restricts the coefficient to one particular relative year j. The coefficients β_{1j} inform about the dynamic effect of Internet availability. Thereby, each coefficient captures relative-year-specific treatment effects. We expect to see no effect before the treatment. Thus, if we cannot distinguish the estimates of the coefficients of the pre-treatment relative-year dummies from zero, the treatment and control group follow similar trends before the treatment, supporting the parallel-trends assumption.

As a number of recent contributions have pointed out, two-way fixed effect (TWFE) event-study (or DiD) approaches, similar to the specification in Equation (1.2), may still yield biased estimates when treatment effects vary over time [see e.g., Athey and Imbens, 2021; de Chaisemartin and D'Haultfœuille, 2020; Borusyak et al., 2021; Goodman-Bacon, 2021; Sun and Abraham, 2020]. The main reason is that the TWFE estimator uses already-treated towns as control group for newly-treated towns, causing a violation of the parallel-trends assumption in the presence of treatment effect dynamics. However, this is precluded when applying more rigorous country-year fixed effects as we do. If country-year fixed effects are applied, there is only one treatment (the year the international Internet connection was established). Nonetheless, we can relax the fixed effects and use the classical TWFE model with town and calendar-year fixed effects. To account for the resulting threat to identification, as described above, we perform alternative approaches proposed by Callaway and Sant'Anna [2021], Roth and Sant'Anna [2021*a*], and Sun and Abraham [2020] for robustness. For instance, Callaway and Sant'Anna [2021] suggest a two-step estimation strategy by first estimating 'group-time average treatment effects', where groups are defined by when towns are first treated, before aggregating the treatment effects by relative time using a propensity-score weighting method.

1.4 Data

We analyze the effect of Internet availability on local economic growth in SSA. To this end, we tap two main data sources. First, local economic activity is measured by nighttime light (NTL) satellite data. Second, locations connected to the Internet are identified via the geo-location and construction year of access points to the national fiber-cable backbone. Moreover, we use data on towns' built-up area, merged with characteristics, such as administrative status and population, and infrastructure, such as (rail)roads, mobile coverage, and the electricity grid. Finally, we make use of the countries' connection dates to the sub-marine cables (SMCs) or via neighboring countries.

1.4.1 Local Economic Activity: Nighttime Lights and Built-up Areas

We measure economic activity at the town level. To identify town locations and extent, we use the established data from *Africapolis* on built-up areas.[15] This database contains the geographical delineation of 5,811 SSA agglomerations with more than 10,000 inhabitants in 2015. The median size is around 20,000 inhabitants and about 90 percent have less than 100,000 inhabitants. The population is also been made available for earlier years by *Gridded Population of the World (GPW)*.[16]

[15] https://africapolis.org
[16] https://sedac.ciesin.columbia.edu/data/collection/gpw-v4

Since geographically and chronologically granular data on economic activity in SSA is lacking, especially for the period we investigate, we deploy NTL satellite data. This data measures human-caused NTL emissions in a geographically high resolution and on a yearly basis. The data was collected in the *Defense Meteorological Satellite Program* (DMSP) *Operational Linescan System* (OLS) between the early 1990s and 2013. The instruments of DMPS-OLS satellites measure light intensity on an integer scale from 0 to 63 with pixels covering 30 arc-second grid cells (an area of .86 square kilometers at the equator). The data is then combined to yearly composite images. We use the harmonization by Li et al. [2020]. This procedure excludes noise from aurora, fires, boats, and other temporal lights and inter-calibrates the data globally for each year as well, making it temporally consistent.

On the country level, NTL data is well established as a measure of economic activity and widely used by economists (Henderson et al. [2012] and Chen and Nordhaus [2011] among the first ones). Closely related to our work, Storeygard [2016] established this data on the city level. At larger geographic resolutions, Bruederle and Hodler [2018] added the relation to household wealth, education, and health for *Demographic and Health Surveys* cluster locations as well as for grid cells of roughly 50×50 kilometers.

1.4.2 Internet Infrastructure: Backbone Access Points and Sub-Marine Cables

For the treatment year, we use information on SMCs' landing dates on the shores of SSA countries for coastal countries from *Submarine Cable Map*.[17] We geo-coded the landing point to merge it to the respective built-up area. If the connection was established through a neighboring country, we assign the establishment year of a country border access point to the national fiber-cable backbone as the treatment year. The geo-locations of the access points and their respective establishment years come from *Africa Bandwidth Maps*.[18] Figure A.2 shows a map of all access points and their construction year. Table A.1 shows the country-specific connection years for all SSA countries that were connected before 2009. In the last column, the year of the speed upgrade through the next SMC is shown. These SMCs had a lot higher capacities and landed in SSA between 2009 and 2012.

Africa Bandwidth Maps contains the most comprehensive set of access points for Africa. It covers the period starting from 2009 and is updated on a yearly basis. The data is directly sourced from the network operators.[19] As access points existing in 2009 were largely

[17] https://www.submarinecablemap.com

[18] http://www.africabandwidthmaps.com

[19] To date, there are 2,708 access points in SSA countries. About half of them were constructed since 2013. Especially in bigger cities, more than one access point is usually established to account for the limited capacity of each access point. In 2019, for example, although 189 new access points were constructed, only 27 new cities and towns were connected. In total, around 900 cities and towns have an access point close by.

established earlier, we conducted an extensive review of backbone deployment projects for each country. Thereby, we determined the construction years of the access points from 2009 going back to the late 1990s for all SSA countries. Note that it was not always possible to determine the exact year of construction. However, in these cases, it was still possible to determine which access points were constructed until in the year the countrywide Internet connection was established, which is still sufficient for our analysis. This makes it possible to identify which towns already had access to the national fiber-cable backbone when the Internet became available for the first time. We match access points to towns via their geo-location: First, we calculate the distance between the towns' border and the closest access point. Then, we assign a national fiber-cable backbone connection to towns within a distance of less than 10 kilometers.[20]

1.4.3 Further Data Sources

We use the share of the area a town has mobile coverage as control variable for the rollout of an alternative digital infrastructure.[21] The data is sourced from *Collins Bartholomew*.[22] Though, since the early 2000s the new mobile-phone standard became 3G, none of the countries in our analysis has rolled out 3G. Therefore, mobile coverage in our data refers to GSM (2G) which allows for basic applications (calls and SMS) but not for mobile Internet.

From *OpenStreetMap (OSM)*, we take the definition of nodal cities. Capital cities and region capitals are marked there. For the definition of economic centers we take the population in the year 2000 from *Africapolis*. As time constant measures of infrastructure, we take shapefiles for roads and railroads from *Natural Earth (NE)*. *Africa Infrastructure Country Diagnostic (AICD)* provides a shapefile for the electric grid in the year 2007.

We examine changes in industry shares as a mechanism for the Internet growth effect. We aggregate census microdata from *IPUMS-International* to a regional level of second order.[23] For the industry shares, the data contains whether the employment is in agriculture, manufacturing, or services. The data comes usually every ten years. Therefore, we estimate a long difference with one pre-treatment and one post-treatment period.

[20] We conducted interviews with industry experts to verify this decision. In addition, in a robustness check we vary this distance.
[21] The share is usually either 0 or 1.
[22] https://www.collinsbartholomew.com/
[23] The admin-2 level is below the state level.

1.4.4 Combining the Data

Our analysis is focused on rather small towns. These towns might not be precisely measured by the satellites' instruments. In fact, for very small towns we observe that they are not bright enough to reach the instruments' sensitivity threshold in each year. We therefore remove towns which do not have positive light intensity in all years. Thus, we reduce measurement error and additionally the sample loses very small towns. If towns are visible in all years, we can additionally be sure that they have stable electricity available. So, we can rule out a potential source which might confound our results.

As light blurs out to adjacent pixels, cities appear bigger in the data than they actually are. By taking the extent of the towns in 2015, we capture some of the blurring as the towns might have been growing after our observation period. However, for some towns, the NTLs still might blur over the extent of the built-up areas. Therefore, we account for blurring by adding a radius of 2 kilometers to the built-up area, such that the growth of light emissions in the extensive margin is properly captured.[24] Unlike in the developed world, very high light intensities, i.e., top-coded pixels, are less a concern in the context of SSA [Bluhm and Krause, 2018]. In our sample, less than 2 percent of pixels are assigned light intensities of 60 or more.

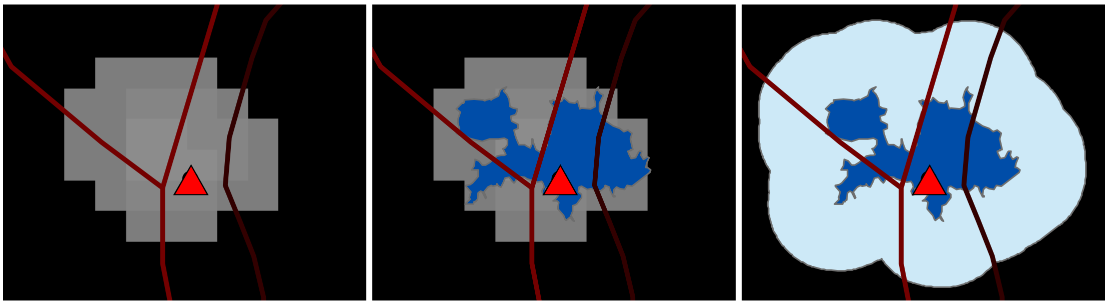
Figure 1.3 shows for Dassa-Zoumè, Benin, its NTL emissions, built-up area, and infrastructure. A road and a railroad connecting Dassa-Zoumè with its neighbouring cities (red and darker red line) and the access points (red triangles) constructed in 2001 are shown in all panels. Panel (1) shows moreover the NTLs for the year 2004 (three years after the countrywide Internet connection and at the end of the analysis period), where a brighter gray reflects higher NTL intensity. Panel (2) adds Dassa-Zoumè's built-up area from *Africapolis* in a dark blue. It shows that through blurring, the NTLs exceed the built-up boundaries. Therefore, we draw a buffer of 2 kilometers around the built-up area in a lighter blue (shown in Panel (3)). This allows us to take all NTL emissions into account.

Within each town, we define several outcome measures.[25] Local economic activity is measured by summing the light intensity of all pixels within a town (and the 2 kilometer buffer) in each year. This measure was established by Storeygard [2016] and accounts for both increased light intensity and geographical extension. As alternative measures, we calculate the average light intensity of pixels and the sum of all lit pixels, ignoring light intensity. We interpret the average light intensity as a proxy for density in terms of population or per capita economic activity (intensive margin) and the sum of lit pixels as a proxy for spatial extension of a town (extensive margin). For an example treatment town and for an

[24] For robustness, we also show the results for a specification without a buffer as well.

[25] As specified in Section 1.3, we apply the logarithm of each outcome measure in the regression analysis.

Figure 1.3: Data example: Dassa-Zoumè, Benin (2004)



(1) (2) (3)

Notes: Panels (1) through (3) show our data for Dassa-Zoumè, Benin, in 2004. Dassa-Zoumè is in the treatment group as one of the incidentally connected towns. Panel (1) shows the access point existing in 2001 (red triangles) and NTLs for the year 2004 (three years after the connection year of Benin). The access point lies within the towns boundaries. The red line represents a major road connecting Dassa-Zoumè with its neighbouring cities and the darker red line the railway connection. The black-to-white scale indicates light intensity, with brighter colors reflecting higher light intensities. Panel (2) adds its built-up area from *Africapolis* (shown in darker blue). Finally, Panel (3) shows in blue a 2 kilometer buffer around that built-up area.

example control group town, Figure A.3 shows how their respective NTLs have changed from the year of the Internet connection to three years later.

The final estimation sample consists of ten countries, which (i) were connected at basic speeds, (ii) have at least one town in the treatment and in the control group, which is not a nodal city, and (iii) have at least three post-treatment years before a high-capacity SMC connects them.[26] While the first restriction is due to the capacity of SMCs, the second restriction depends on the access point rollout within each country. The last restriction is necessary to estimate the effects of Internet availability at basic speeds on a balanced panel. A longer post-treatment period would shrink the sample further. Therefore, we estimate the main specification with three post-treatment years. For robustness, we will relax these restrictions. The first connected country in our sample is Senegal, which was connected in 2000. Therefore, we are less restricted in the pre-treatment period and do not lose any country there. For the estimation sample, Figure A.4 shows the geographical distribution and the location of the treatment and control group towns without the nodal cities. There are four countries in West Africa and Southern Africa, respectively, and two countries in East Africa in our sample. Of the ten countries, five are coastal and five are landlocked.[27]

[26] These countries are Angola, Benin, Botswana, Ethiopia, Mali, Sudan, Senegal, Togo, Zambia, and Zimbabwe. An overview of the procedure how the estimation sample of countries emerges can be found in Section A.4.

[27] Sudan is a special case as ten towns are in the control group but only one town is in the treatment group. Angola has very similar issues. We account for that by grouping fixed effects for East, Southern, and West African countries for robustness.

1.4.5 Descriptive Statistics

We focus our analysis on mid-sized towns. From 510 agglomerations, for which NTL data is detected in each year, in the ten countries of the estimation sample, 143 were connected to the Internet via an access point before the country was connected via SMC or a neighboring country. Therefore, they are part of our treatment group. Of these agglomerations, 70 are nodal cities. Another 147 towns got an access point in the subsequent years and are therefore in the control group. The remaining 118 agglomerations are still not connected. Further 102 towns were connected in the three years after the countrywide Internet connection and are therefore not considered as they would confound our control group.

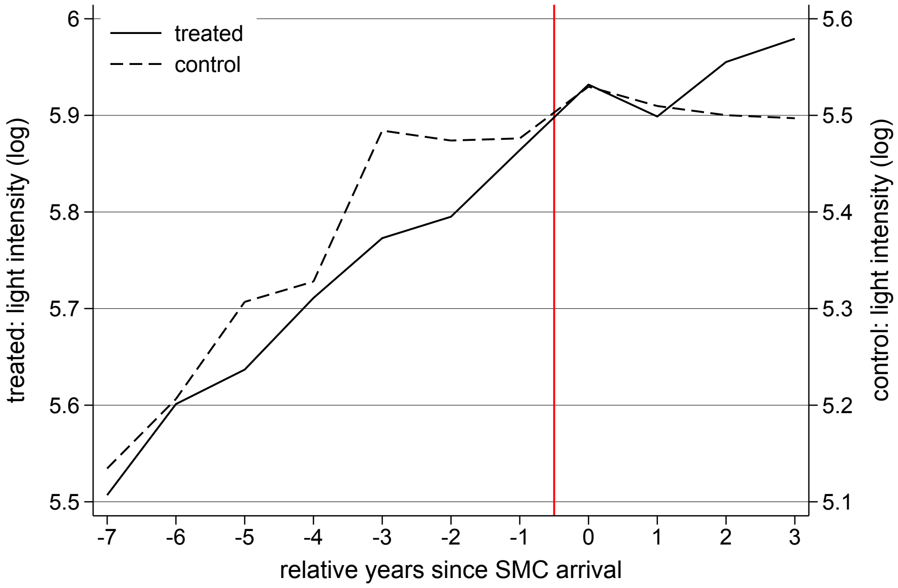
Figure A.5 compares the average size of cities and towns by the year they get an access point, relative to the treatment year. In the early years, until the Internet becomes available countrywide, many nodal cities are connected besides the towns in the treatment group. While connected nodal cities are bigger on average in the early years and decline in their size in subsequent years, towns in the treatment group only have a population of around 25,000 inhabitants on average. Control towns, which are connected in the subsequent years after the observation period of three years, vary for all subsequent years between a population of 10,000 and 35,000 inhabitants as well. Especially, when only examining treated and control towns, i.e., excluding nodal cities, there is no clear (decreasing) pattern with respect to population size over time anymore in the first ten years. Such a pattern is clearly identified for nodal cities. Only towns which get access rather later are smaller on average. We account for this timing in a robustness check. This finding suggests that treated towns are not selected into treatment because of their population size. Moreover, nodal cities are still connected in further years after the arrival of the first Internet connection, showing that the rollout continues to other parts of the country. Their size decreases after the first two years as capital cities are usually connected early and are usually a lot bigger than other nodal cities. Nonetheless, the size of later connected nodal cities is still bigger on average than the size of the control towns.

Table A.2 gives a broad overview of the towns. The statistics of the outcome measure of the light intensity show a value of 463.04 on average one year before the treatment (161.50 at the 25th percentile, 285.00 at the median, and 530.50 at the 75th percentile). For the size of the towns, measured with the NTL data, values are as followed: 43.35 on average one year before the treatment (24 at the 25th percentile, 35 at the median, and 53 at the 75th percentile). On average, including no-lit pixels with a value of zero, towns have values of 7.50 on average one year before the treatment (3.22 at the 25th percentile, 5.25 at the median, and 10.07 at the 75th percentile). Given that the instruments pickup light usually at a threshold of 4, the average values are rather modest. The rather high number of lit pixels corresponds to the condition that towns have to show up in each year in the

NTL data. Coming to the other variables, mid-sized towns have a population of around 20,500 inhabitants on average in 2000 (8,500 at the 25th percentile, 16,000 at the median, and 30,000 at the 75th percentile). Mobile coverage is available in about 62 percent of the towns one year before treatment, given that usually the percentage covered is either zero or one. By construction, the maximal distance to the closest access point is with 9.43 kilometers smaller than 10 kilometers. On average, this distance is a lot smaller with 1.26 kilometers. More than half of the towns have an access point even within the built-up area and most cities have it within 2 kilometers (1.21 kilometers at the 75th percentile). The distances to further infrastructure, such as the road network, railroad network, or electricity grid, are usually small with median distances of 0 kilometers (3.8 kilometers for the railroad network). Further distances are given for the next port, for coastal countries, as well as to the capital city, to the next regional capital, and geographical measures, such as the coastline or the next river.

Before presenting the estimation results, we show the development of cities and towns over time. We use the main outcome measure, log light intensity, averaged over the treatment and control group separately but do not include any fixed effects or controls. Figure 1.4 shows that in the early pre-treatment years both groups grew on average with similar rates. While the gap between treatment and control group towns is almost equal in t-7 and t-1, the control group grew slightly faster in the earlier year and the treatment group in the later years. At the end of the observation period, the gap between these towns grew by about .1 on the logarithmic scale. While for the control group there is a stagnation, for the treatment group the growth is also slowed down. This pattern holds for almost all types of cities. Figure A.6 shows the growth rates of nodal cities and towns. For all groups but the economic centers, annual-growth rates declined after the Internet connection, while the country's GDP growth accelerated in the same period (Figure A.7). Thus, the slowdown in NTL growth is probably an inconsistency in the satellites. As Figure A.6 shows, before the Internet connection all town types had a similar annual growth rate. However, when Internet became available, treated towns showed an a lot larger growth rate than control group towns.

Figure 1.4: Time trends of treatment and control group



Notes: The figure depicts the average growth of the towns in the treatment and control group over a period of eleven years (seven before and three after the treatment year). The measurement is the logarithm of light intensity.

1.5 Results

1.5.1 Main Effects and Mechanism

We estimate the effect of Internet availability on local economic growth. Particularly, we are interested in the effect of early Internet availability brought by the 'first generation' of SMCs. Nodal cities are excluded. We estimate a linear model on a balanced panel by difference-in-differences, where town and country-year fixed effects are included and standard errors are clustered at the town level. We measure economic activity by the logarithm of the sum of NTL intensities as the main outcome. Table 1.1 shows the main results. Columns (1) and (2) show the effect of Internet availability on light intensity. This effect is then separated into growth on the intensive and extensive margin (Columns (3) and (4)). Column (5) investigates population growth.

In line with our expectations, we find an economically and statistically positive effect of the availability of Internet at basic speeds on local economic growth. In our preferred specification (Column (2)), towns which were connected to the Internet in the year of an SMC arrival become 7 percent brighter than towns without Internet access. This finding supports our initial claim that towns which get incidentally connected to the Internet grow faster in comparison to otherwise comparable towns. The mobile coverage control does not turn out to be statistically significant and is smaller in size in comparison to the main effect. It makes the estimation more precise as it controls for differences in another ICT. As it increases slightly the point estimates of the main effect, we are not worried that the main effect is transported through mobile coverage. We will discuss the role of mobile coverage in more detail in Section A.6.

Table 1.1: The effect of Internet availability on the economic growth of towns

	(1) light intensity	(2) light intensity	(3) intensive margin	(4) extensive margin	(5) population
post x treated	0.0633*	0.0703**	0.0513**	0.0516*	0.0116
	(0.0344)	(0.0349)	(0.0231)	(0.0282)	(0.0183)
GSM coverage		0.0486	0.0477**	0.0281	0.00699
		(0.0342)	(0.0240)	(0.0263)	(0.0119)
observations	2,420	2,420	2,420	2,420	2,420
R-squared	0.943	0.943	0.947	0.924	0.999
#countries	10	10	10	10	10
#towns	220	220	220	220	220
share treated	.445	.445	.445	.445	.445
town FE	✓	✓	✓	✓	✓
country x year FE	✓	✓	✓	✓	✓
w/o nodal cities	✓	✓	✓	✓	✓

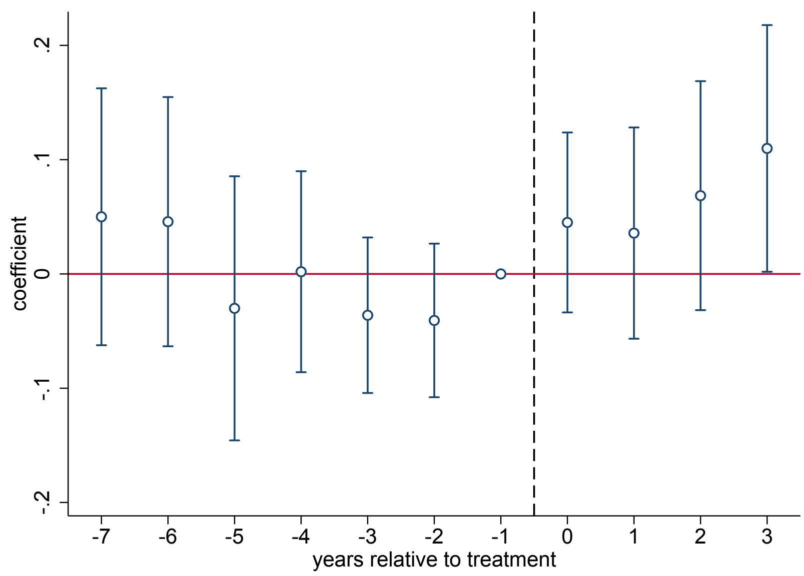
Notes: Light intensity is measured as the logarithmic sum of light intensities of DMPS-OLS pixels within town area and a 2 kilometer buffer. The measure for the intensive margin is logarithmic mean light intensity and for the extensive margin logarithmic sum of lit pixels (coded as 1 if a pixel is lit). Population is measured as inhabitants per square kilometer within town area (including the 2 kilometer buffer) from *Gridded Population of the World* (GPW). GSM mobile coverage is calculated as the percentage share of town area covered with signal. Nodal cities include landing point and capital cities, regional capitals, and cities with more than 100,000 inhabitants. All specifications include town and country-year fixed effects. Robust standard errors clustered by town reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We translate the effect in terms of light intensity to an approximation of the implied economic effect, by a back-of-the-envelope calculation using the GDP-luminosity elasticity from Henderson et al. [2012]. Henderson et al. [2012] show that growth in light intensity serves as a good approximation of economic development at the country level. The elasticity remains robust for a global sample as well as a sample of low and middle income countries. Storeygard [2016] further shows that the elasticity at the country level is if anything slightly higher for SSA countries and for coastal primates. Moreover, he shows that the elasticity holds at the sub-national level as well. We follow his argumentation that the elasticity in SSA could be lower as countries have lower light intensities on average or that the elasticity could be higher as SSA finds fewer top-coded cities. Both issues are specifically the case in our sample. We therefore use the elasticity of $\epsilon_{GPD,light} = 0.284$ from Henderson et al. [2012], for which the calculation translates the increase in light intensity of 7 percent into about 2 percentage points higher GDP growth.

Figure 1.5 presents event-study coefficients for our preferred specification (Column (2) of Table 1.1). Before SMC connection, the point estimates are close to zero and statistically insignificant. This supports the assumption that treatment and control group towns are not of different growth parts preceding the countrywide Internet connection, conditional on fixed effects. From the dynamic perspective, there is no evidence for a potential fading out of the effect. In contrast, in the year of the countrywide Internet connection, the point estimate turns positive but remains statistically insignificant until year t+3. In the years after the countrywide Internet connection, the point estimates are between .05 and .1 and have a slight tendency to increase to nearly .15 in year t+3, when the effect gets statistically significant at the 5% level.

The increase of the effect over time is in line with the expectation that Internet adoption takes time and that growth effects develop only some time after Internet adoption. Moreover, it indicates that the effect might not be completely induced by adopting firms but be partly induced by spillovers of the local economy. Finally, it is a strong sign that the effect is not coming from an electricity demand that was satisfied after giving Internet access to the town. If that was the case, the increase in light emission should be found earlier and would not be increasing over time after the treatment year.

Figure 1.5: Event-study coefficients



Notes: Coefficients for event-study specification of Column (2) of Table 1.1. Robust standard errors clustered by town. Confidence intervals reported at the 95% level.

We further investigate the intensive and extensive margin of regional development. Therefore, we take the logged mean luminosity and the logged sum of lit pixels (Table 1.1, Columns (3) and (4)) as outcome measures. The observed increase in light intensity of towns having Internet available could be explained by growth in productivity, population density, or populated area, among other explanations. While we investigate the effects on population later, the intensive and extensive margin can show whether the town is growing only in size at its border (extensive margin) or whether existing pixels are getting brighter (intensive margin). While towns' light intensity increases by 7 percent, their brightness increases by 5 percent and their size increases by 5 percent as well. The observed increase in light intensity can thus be explained by both an increase in brightness (Column (3)) and in size (Column (4)). However, while the effect on the intensive margin is statistically significant at the 5% level, the effect on the extensive margin is only statistically significant at the 10% level. As brighter lights glow further, the increase in the extensive margin is at least partly an indirect effect of the growth in the intensive margin. This finding is important as it is not clear a priori that the effect manifests beside the towns' border.

Investigating population as an outcome, the estimate is close to zero and lacks statistical significance. Therefore, the effect of Internet availability is mainly inducing economic development and not population growth or migration. This suggests that the towns' increase in light intensity reflects primarily an increase in their productivity. As we use an interpolation for population from *GPW*, Table A.3 shows different specifications regarding the included periods and years for robustness. The effect of Internet connection on population lacks statistical significance and the point estimate is close to zero in all specifications.

In a further robustness test, we add population as a (bad) control instead of using it as an outcome as before. In Table A.4, the regressions from Table 1.1 are repeated in odd columns and population is added in even columns. By doing so, we control for which part of the effect is driven by economic growth and which by growth in population. The main effect on light intensity remains robust and statistically significant (Column (2)), strengthening our claim that the effect of Internet availability is mainly inducing economic development and not population growth. Population has an elasticity of around .4 on light intensity. The coefficient of the population control is statistically significant at the 10% level (Column (2)). The main effect on the intensive and extensive margin remains robust as well when additionally controlling for population. However, the population estimate differs in both specifications. For the intensive margin, the population estimate is smaller and statistically insignificant (Column (4)). For the extensive margin, the population estimate is positive and statistically significant (Column (6)). As population is not correlated with brightness, towns are not getting populated more densely. Again, strengthening our claim of an effect on towns' productivity. We take from the correlation between population and size that if there is any small increase in population, it is induced by a few settlers at the towns' borders.

Next, we show how the industry composition changes in regions with Internet access as these changes might be a channel through which Internet availability affects local economic growth. We use survey data from IPUMS-International, a collection of census microdata, to calculate the share of jobs in each industry (agriculture, manufacturing, and services).[28] We estimate the effects on growth rates and changes in the industry shares.

The data contains 21 SSA countries. However, only eight of them have more than one year, as for many countries early data is not available. From the remaining countries, seven countries were connected to the Internet in 2008 or earlier and six of them had constructed at least some access points when the countrywide Internet connection was established. Finally, for five countries, both a treatment and control group can be defined. These countries are: Benin, Mali, Malawi, Mozambique, and Zimbabwe. Allowing for surveys close to the connection year, we can estimate also on countries that were connected late (in 2006 and 2007), for which the upgrade induced by 'second generation' SMCs came shortly after the first connection. Malawi and Mozambique only have two post-treatment years and were not included in the estimations so far. The treatment is defined as before: We remove regions that contain a nodal city and define a region as treated if at least one town in that region has early Internet access. For most of the remaining countries, data is available with a frequency of ten years. Only Mali (and Benin) has a difference of eleven years (once). For Benin, there are three survey rounds available. However, as one round was in 2002, only one year after the connection year, we drop this year. Thus, we estimate a long difference with one survey year before the arrival of the countrywide Internet connection and one afterwards.

Figure A.8 shows the changes descriptively. Both the treatment and the control group have relatively high shares of agricultural employment. Unexpectedly, the shares in services are higher than in manufacturing. After the treatment, agricultural shares decline in both groups and manufacturing and services shares increase in both groups. Especially for manufacturing, but also in both other industries, the changes are larger in the treatment group.

The results in Table 1.2 indicate that Internet availability shifts jobs from agriculture to manufacturing (and slightly to services). In Columns (1) through (4), we estimate growth rates, while we estimate changes in industry shares in Columns (5) through (7). In the final sample 144 towns are contained, of which 26 are treated. The share of jobs in agriculture declines by more than 3 percentage points. In contrast, job shares in manufacturing increase by more than 2 percentage points and job shares in services increase by around 1 percentage point. However, only the effect on manufacturing is statistically significant at the 5% level. Comparing these results with the growth rates in Columns (2) through (4), the signs for the agricultural and manufacturing sector are equal. However, all estimates lack statistical significance. Column (1), again underlines that the effect is not driven by migration as the total number of individuals is not statistically greater in the treatment group than in the control group with a point estimate very close to zero. The results are in line with Hjort

[28] https://international.ipums.org/international/

and Poulsen [2019], who find an increase in employment in Ethiopian manufacturing firms and an increase in net firm entry in services. As manufacturing might emit NTLs differently than agriculture, the estimated higher light intensity might reflect both: economic growth through more manufacturing jobs and a change in the industry structure (independent from the growth). We lack evidence whether the effect stems from newly created firms or growing already existing ones.

Table 1.2: Employment growth rates and shares by industry

	growth rate				industry share		
	(1) total	(2) agriculture	(3) manufacturing	(4) services	(5) agriculture	(6) manufacturing	(7) services
post x treated	-0.0215	-0.105	0.131	-0.0672	-0.0328	0.0225**	0.0103
	(0.0569)	(0.0908)	(0.104)	(0.0970)	(0.0239)	(0.0102)	(0.0184)
GSM coverage	-0.0570	-0.0824	-0.180*	-0.0877	-0.00696	-0.000510	0.00747
	(0.0570)	(0.0739)	(0.0968)	(0.0878)	(0.0172)	(0.00583)	(0.0130)
observations	288	288	288	288	288	288	288
R-squared	0.967	0.964	0.934	0.951	0.954	0.893	0.958
#countries	5	5	5	5	5	5	5
#regions	144	144	144	144	144	144	144
share treated	.181	.181	.181	.181	.181	.181	.181
region FE	✓	✓	✓	✓	✓	✓	✓
country x year FE	✓	✓	✓	✓	✓	✓	✓
w/o nodal cities	✓	✓	✓	✓	✓	✓	✓

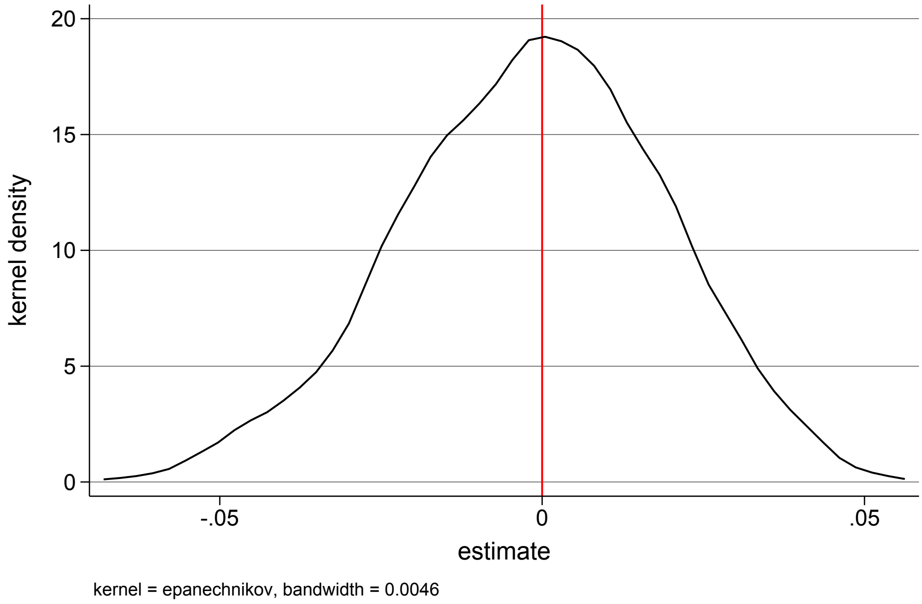
Notes: Regional industry composition comes from *IPMUS International*. All specifications are estimated on a sample restricted by excluding landing point cities and capitals, regional capitals, and cities with more than 100,000 inhabitants and include town and country-year fixed effects. Robust standard errors clustered by town reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

1.5.2 Placebos

Randomly Assigned Connection Years We identify a causal effect under the assumption that the connection year of each country is exogenous. In the event-study plot, it was already shown that the effect only starts after the connection year. As a placebo test, we now assign randomly connection years to each country individually (prior to the actual connection year). We also shift the rollout of the access points accordingly by the same number of years as the actual connection year was shifted. We re-estimate Equation (1.1) with these randomly assigned connection years 1000 times and plot the distribution of the main effect. Figure 1.6 shows a very symmetric distribution with the peak being very close to zero. The main effect only remains with the actual connection years and diminishes with randomly assigned ones. This indicates that the connection years are exogenous.

Placebo Tests with Nondigital Infrastructure In Table 1.3, we include additional controls. To save costs, fiber cables are rolled out along the existing (transportation) infrastructure (cf. Section 1.2.2). We control for this other infrastructure to rule out that towns closer to this nondigital infrastructure grow faster when the countrywide Internet connection is

Figure 1.6: Placebo (randomly assigned country-connection years)



Notes: The figure depicts the distribution of the main effect, estimated with 1000 random country-connection year combinations.

established, irrespective of whether they are in the treatment or the control group. Similarly, we control for the electricity grid.[29] Unlike for mobile coverage, we do not have time-varying data on other infrastructures. We therefore generate indicators for whether the distance is below 10 kilometers (as we defined treated towns with access points) and then intersect these indicators with the post dummy for the time after the countrywide Internet connection is established and construct placebo treatments. Column (1) repeats our preferred specification (Table 1.1, Column (2)). In Column (2), we control for different effects for towns next to a greater (paved) road. In Column (3), we control for different effects for towns next to the railroad network. In Column (4), we control for different effects for towns next to the electricity grid. Finally, we include all infrastructure controls jointly (Column (5)). In each case, the estimate of the main effect remains close to .07 and is statistically significant. The estimates of all placebos lack statistical significance and are rather small in their economic significance. In the case of roads, the main estimate looses slightly precision and decreases by half a percent to .065. Nevertheless, controlling for access to the railroad network, the electricity grid, or all infrastructures jointly, the estimate is

[29] For the infrastructure, we cannot be sure that roads and railroads were existing prior to the rollout of the national backbone. However, as we know that the rollout followed this infrastructure we take the existing data for the placebo exercise. Moreover, the electricity grid data we have is from 2007 and thus might assign an electricity grid to towns which only were connected to the electricity grid when they became access to the Internet after the countrywide Internet connection was established. We therefore discuss the role of the electricity grind in more detail later.

again statistically significant at the 5% level and slightly higher than in the baseline (7.3 to 9.4 percent).

Table 1.3: Placebo (competing infrastructure)

light intensity (log)	(1)	(2)	(3)	(4)	(5)
post x treated	0.0703**	0.0647*	0.0942**	0.0733**	0.0903**
	(0.0349)	(0.0349)	(0.0376)	(0.0356)	(0.0375)
GSM coverage	0.0486	0.0477	0.0463	0.0480	0.0450
	(0.0342)	(0.0341)	(0.0340)	(0.0343)	(0.0340)
post x road network (dummy)		0.0928			0.104
		(0.0768)			(0.0757)
post x railroad network (dummy)			-0.0624		-0.0657
			(0.0385)		(0.0406)
post x electricity grid (dummy)				-0.0227	-0.00718
				(0.0391)	(0.0405)
observations	2,420	2,420	2,420	2,420	2,420
R-squared	0.943	0.943	0.943	0.943	0.943
#countries	10	10	10	10	10
#towns	220	220	220	220	220
share treated	.445	.445	.445	.445	.445
town FE	✓	✓	✓	✓	✓
country x year FE	✓	✓	✓	✓	✓
w/o nodal cities	✓	✓	✓	✓	✓

Notes: Light intensity is measured as the logarithmic sum of light intensities of DMPS-OLS pixels within town area and a 2 kilometer buffer as in Table 1.1. GSM mobile coverage is calculated as the percentage share of town area covered with signal. All specifications are estimated on a sample restricted by excluding landing point cities and capitals, regional capitals, and cities with more than 100,000 inhabitants and include town and country-year fixed effects. Railroad and road networks are source from *Natural Earth*. The electricity grid is sourced from *Africa Infrastructure Country Diagnostic*. Robust standard errors clustered by town reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Electricity Grid For already named reasons, the rollout of the electricity grid might be a thread to isolate the effect of Internet availability. We therefore analyze whether households get connected to the electricity grid with survey data from *Afrobarometer* [BenYishay et al., 2017].[30] We take the first four rounds, if they are before the SMC upgrade, from 1999 to 2009 to maximize the number of countries. We take averages for each town and generate weights for the number of surveyed individuals per town. In some countries, only very few towns are visited besides of nodal cities. In Table A.5, we first include all cities (Columns (1) and (2)) and restrict the sample stepwise, dropping first all capitals (Columns (3) and (4)), and in a second step, all nodal cities (Columns (5) and (6)). In odd columns, no weights are used, while in even columns, the number of surveyed individuals per town is used as a weight. All specifications lack statistical significance. In the overall sample (Columns (1) and (2)), the point estimate is very close to zero. It increases slightly for the sample without capital cities (Columns (3) and (4)) and becomes negative for the most restrictive sample

[30] https://afrobarometer.org/

(Columns (5) and (6)). We take from these estimations that for some nodal cities (without capital cities) the rollout of the electricity grid advanced at the same time as the Internet was rolled out. However, for smaller towns this cannot be confirmed. We can therefore rule out the concern that the electricity grid confounds our results.

1.5.3 Robustness

Different Fixed Effects As explained above, we apply country-year fixed effects to account for country-specific growth paths in the countries' economies. For robustness, we re-estimate Equation (1.1) with the classical two-way fixed effects: towns and calendar years. This specification is less demanding in the set of fixed effects. A concern with these fixed effects might be that countries on a higher growth path might construct more access points faster. Therefore, this specification serves as a robustness check and not as the main specification. Nevertheless, the estimate presented in Column (2) of Table 1.4 remains robust. However, the estimate of the control variable GSM mobile coverage turns statistically highly significant. Figure A.9 shows the event-study plot, which is very similar in comparison to the preferred specification. Though, the pre-trends have higher (statistically insignificant) point estimates. In Subsection 1.5.4, External Validity, we estimate further models with the classical two-way fixed effects, which allow for a bigger sample, containing more countries as the treatment and control group are compared across countries. Moreover, we apply novel DiD/event-study estimators to account for the staggered timing of the treatment [Roth and Sant'Anna, 2021*a*; Callaway and Sant'Anna, 2021; Sun and Abraham, 2020].

Error Correlation within Regions Another potential concern is that model errors are correlated within regions. This might be the case when the effects of the access point might generate further spillover effects in the towns' surrounding area. In our preferred specification, we cluster the standard errors at the town level as the treatment, the access point construction, is occurring there. If more than one town is located within 10 kilometers to the access point, an access point would serve more than one town. To take this into account, we apply a higher level of clustered standard errors for robustness. We re-estimate Equation (1.1) correcting standard errors for clusters at the level of states. Column (3) of Table 1.4 presents the estimates with a higher level of clustered standard errors. The standard error of our variable of interest increases only very slightly (from .0349 to .0358).

Linear Time Trends In Column (4) of Table 1.4, we also test the robustness of our results against the inclusion of linear time trends on the town level. To account for possible differential trends among towns, we re-estimate Equation (1.1) including a linear town-specific yearly trend. This is the most demanding specification. The estimate increases

Table 1.4: Robustness

light intensity (log)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
post x treated	0.0703**	0.0580*	0.0703*	0.102**	0.0625*	0.0600*	0.0797***	0.121***	0.0720*
	(0.0349)	(0.0338)	(0.0358)	(0.0449)	(0.0336)	(0.0323)	(0.0265)	(0.0392)	(0.0374)
GSM coverage	0.0486	0.0908***	0.0486	0.0193	0.0233	-0.00762	0.0193	0.0611	0.0482
	(0.0342)	(0.0320)	(0.0330)	(0.0332)	(0.0310)	(0.0308)	(0.0249)	(0.0380)	(0.0408)
observations	2,420	2,420	2,420	2,420	2,827	3,839	2,343	2,860	1,804
R-squared	0.943	0.927	0.943	0.962	0.951	0.927	0.978	0.937	0.939
#countries	10	10	10	10	10	10	10	10	9
#ethnic group-countries									13
#towns	220	220	220	220	257	349	213	220	164
share treated	.445	.445	.445	.445	.525	.281	.502	.445	.445
town FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
country x year FE	✓		✓	✓	✓	✓	✓	✓	
year FE		✓							
ethnic group-country x year FE									✓
w/o capital+landing point	✓	✓	✓	✓	✓	✓	✓	✓	✓
w/o regional capitals	✓	✓	✓	✓	✓	✓	✓	✓	✓
w/o population >100k	✓	✓	✓	✓	✓	✓	✓	✓	✓
cluster			state-level						
linear time trends				town-level					
late APs					✓				
no not treated towns						✓			
no buffer							✓		

Notes: Light intensity is measured as the logarithmic sum of light intensities of DMPS-OLS pixels within town area. GSM mobile coverage is calculated as the percentage share of town area covered with signal. All specifications are estimated on a sample restricted by excluding landing point cities and capitals, regional capitals, and cities with more than 100,000 inhabitants. Robust standard errors clustered by town reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

slightly to .10 and is also statistically significant at the 5% level. Thus, we can show that after the countrywide Internet connection is established, the connected towns grow in a nonlinear manner.

Access Points During the Post-Observation Period Thus far, we tested the robustness of our results to alternative assumptions about the variance-covariance matrix. In Column (5) of Table 1.4, we add these towns to the estimation sample that were connected only within three years after the treatment year. In our main specification, these towns are excluded as they neither belong clearly to the treatment nor the control group and would thus confound our analysis. Including them, we allow the treatment status to switch from 0 to 1 in the year the access point is constructed (the post dummy switches in the same year as well). With this approach, we can add them without confounding our analysis. Within our ten countries in the analysis, we can add further 37 towns to increase the estimation sample. We estimate Equation (1.1) as in our preferred specification with town and country-year fixed effects and cluster standard error at the town level. As the estimate remains, including slightly later connected towns does not affect the results.

Extending the Control Group In Column (6) of Table 1.4, we add to the control group also all towns without an access point. They are comparable to the control group as they cannot access the Internet when it becomes countrywide available. At that time, it might not yet be known which towns will get an access point in the future. While this sample increases to

349 towns, the share of treated towns declines to 28.1 percent. The estimation results remain unaffected.

No Buffer In Column (7) of Table 1.4, we remove the 2 kilometer buffer and estimate on the original *Africapolis* built-up areas. We adjusted the built-up areas because of the blurring of the NTL data. When examining the smaller built-up area, we might lose some pixels at the towns' border. However, these pixels might be of low intensity. With this robustness check, we can thus not only show that our results do not depend on the adjustment of the built-up area but also that local growth does not predominantly happen at the towns' border. The estimate not only remains, but turns statistically significant at the 1% level.

Extending the Post-Treatment Period Originally, we limited the sample to the years before the 'second generation' of SMCs arrived, such that we can estimate the effect of broadband Internet at basic speeds. The event-study estimates in Figure 1.5 show that the effect on local economic growth increased on a yearly basis. Therefore, we show for robustness how this effect evolves in the two subsequent years (Column (8) of Table 1.4). One should note, however, that the effects in the last periods might be driven by fast Internet induced by the new SMCs. Six countries still have Internet at basic speeds available at the end of the new sample. The event study estimates are shown in Figure A.10 and indicate that the growth rate increases further to more than 20 percent.

Ethnic Favoritism A further concern could be that certain ethnic groups were favored during the rollout. Though, the exogenous shock comes from the countrywide connection year and the parallel trends in the event study do not underpin this concern, a remaining threat could be that certain ethnic groups are also favored in other dimensions, which cause the observed difference in growth over time. Our strategy to overcome this threat is two-fold. First, many countries construct access points for more than one ethnic group before the treatment period (Figure A.11). This indicates that not a specific ethnic group is favored by giving them access to the Internet. For the countries in our analysis, all countries but Angola provided at least two different ethnic groups with access points. And Angola only established very few access points in total. Therefore, the low number of equipped ethnic groups is not surprising. On the other hand, Ethiopia and Togo provided Internet access for even six different ethnic groups very early. Second, we perform our analysis constructing country-ethnic group entities instead of countries. By estimating Equation (1.1) including town fixed effects and country-ethnicity-year fixed effects, treatment and control group towns are compared within an ethnic group. If ethnic favoritism were at play and would drive the found effects, our estimate should vanish as towns with certain ethnic groups should grow, and less importantly get an access point, while towns with other ethnic groups remain on a worse growth path. The results are shown in Column (7) of Table 1.4. A hardly smaller sample size shows that for most ethnic groups for which access points were constructed in the treatment period, access points were also constructed afterwards. Only in Botswana this is not the case. In the remaining nine countries, there are eleven ethnic groups and thirteen country-ethnic group entities in the estimation. The result remains robust, showing that even comparing treatment and control group towns of the same ethnic group, Internet availability has a positive effect on local economic activity.

1.5.4 External Validity

Applying the classical two-way fixed effect model of Column (2) in Table 1.4, we re-estimate Equation (1.1) on a broader sample. Comparing treatment and control group towns in the whole sample through the calendar year fixed effect, we allow also for countries containing either only control or treated towns. Table 1.5 shows the results with a stepwise reduction of nodal cities. In Column (1), the most basic specification, all nodal cities are included. Then, we eliminate nodal cities stepwise until we reach our preferred specification, where the remaining towns are comparable: In Column (2), we remove the city of the landing point and the national capital for each country. In Column (3), we additionally remove all regional capitals. In Column (4), we further remove all cities of more than 100,000 inhabitants. Finally, we add GSM mobile coverage as a control variable to account for other telecommunication technologies (Column (5)). With this approach, we estimate on 491 (including nodal cities) to 352 towns in 19 to 17 countries. In one country, only the capital (and/or the landing point) is connected, in another country only economic centers with more than 100,000 inhabitants are connected. It can easily be seen that this sample contains more control towns as most countries added to this sample started the rollout rather late. A concern might be a correlation between countries' development and timing of its construction of access points. However, estimating on a sample containing more countries shows that our estimates also reach a high external validity.

The estimate of the main effect increases slightly from Columns (1) through (5). This was already the case in Table 1.1 in Column (2) when mobile coverage was added as a control. The estimate is slightly higher, reaching an effect size of 9 percent and reaches significance at the 1% level in our preferred specification in Column (5). In comparison to Column (2) in Table 1.4, where we applied the same fixed effects on the main results sample, the estimate of the main effect increases by nearly .04 and is statistically significant at a higher level. Hence, in the other SSA countries, towns might have developed worse than the control group towns in the original sample.

The event-study graph shows parallel trends before the treatment (Figure A.12). Especially in the four years prior to the treatment, the point estimates are close to zero.

Table 1.5: External validity

light intensity (log)	(1)	(2)	(3)	(4)	(5)
post x treated	0.0620**	0.0712**	0.0833**	0.0888**	0.0938***
	(0.0263)	(0.0282)	(0.0325)	(0.0345)	(0.0343)
GSM coverage					0.0415*
					(0.0237)
observations	5,401	5,170	4,048	3,872	3,872
R-squared	0.963	0.947	0.936	0.916	0.916
#countries	19	18	18	17	17
#cities	491	470	368	352	352
share treated	.334	.309	.307	.287	.287
town FE	✓	✓	✓	✓	✓
year FE	✓	✓	✓	✓	✓
w/o capital+landing point		✓	✓	✓	✓
w/o regional capitals			✓	✓	✓
w/o population >100k				✓	✓

Notes: Light intensity is measured as the logarithmic sum of light intensities of DMPS-OLS pixels within town area and a 2 kilometer buffer as in Table 1.1. GSM mobile coverage is calculated as the percentage share of town area covered with signal. Sample successively restricted by excluding landing point cities and capitals, regional capitals, and cities with more than 100,000 inhabitants and all specifications include town and year fixed effects. Robust standard errors clustered by town reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Only the estimate in year t-5 is statistically significant at the 5% level and has a negative sign. This is not the case in the main specification in Figure 1.5, where all pre-treatment coefficients are statistically insignificant. After the treatment, both event-study estimates are very comparable. In both cases, the point estimate increases especially in year t+2 and even more in year t+3 and becomes statistically significant at the 5% level only in the last observation period.

One concern in this setting is that the classical two-way fixed effect estimator does not account for the staggered timing of the treatment and thus the potential heterogeneous effects of the individual treatment. Recent literature developed estimators for this setting [Roth and Sant'Anna, 2021*a*; Callaway and Sant'Anna, 2021; Sun and Abraham, 2020]. Figure A.13 shows event-study estimates for the respective estimators.[31] All three approaches show very similar results. Again, it is shown that the results hold when accounting for heterogeneous outcomes in staggered treatment timing by the propensity score weighting method and potential comparing of treated and not-yet-treated observations. In contrast to Figure A.12, only the point estimate of year t+2 is statistically significant and the point estimate in year t+3 also declines slightly in comparison to the year before. However, the point estimate in year t+2 is bigger than in Figure A.12. Another differences to Figure A.12 is that in only one estimation the point estimate in year t-5 is

[31] We used the `staggered` R-package by Roth and Sant'Anna [2021*b*].

negative and statistically significant at the 5% level (in the approach by Sun and Abraham [2020]). In the other two cases, this point estimate lacks statistical significance.

1.5.5 Heterogeneity: Coastal Countries

Storeygard [2016] investigates coastal countries due to the design of estimating on the distance to a primate city with a harbor. Hjort and Poulsen [2019] analyze coastal countries as they exploit the landing of SMCs and do not consider Internet connections of landlocked countries. So far, we have used the additional information about the connection year through a neighboring country we have on landlocked countries for estimating on a bigger sample. Nonetheless, one might have concerns about the validity of the exogeneity assumption of the timing of the connection year. Therefore, we reduce the sample one more time to estimate only on coastal countries for which the ground work for the identification is already established by Hjort and Poulsen [2019].

A priori, it is not clear whether the Internet has different effects for coastal and landlocked countries and if it does, which geographic location profits more from an Internet connection.[32] On the one hand, one could argue that more developed countries might have higher growth rates as some development has to be existing for the Internet to have an economic effect. On the other hand, less developed countries might have higher growth potential and the Internet could work as a substitute for worse nondigital infrastructure. In this case, countries could be leapfrogging and overtake more developed countries. However, coastal countries might profit more from international trade as the Internet lowers information costs, e.g., for international prices.

Table 1.6 shows results for coastal countries only. Column (2) presents our preferred specification, while Column (1) shows results without the mobile coverage control. We estimate on the five coastal countries (Angola, Benin, Sudan, Senegal, and Togo), which contain 75 towns in either the treatment or control group of which slightly more than half are in the treatment group. The estimate increases slightly and the standard errors decrease only very slightly from Column (1) through (2). The effect size is higher than in the whole sample (Table 1.1), indicating that coastal countries profit more from Internet access than landlocked countries. An advantage of investigating coastal countries separately is that they were connected earlier. Therefore, it is possible to analyze them with a longer post-treatment period without allowing the upgrade SMCs to confound the results. In Column (3), we re-estimate Equation (1.1) on a sample with five post-treatment years. The main effect increases to .27 and is statistically significant at the 1% level. The event-study plot, Figure A.14, shows that the effect size increases further in the fourth and especially

[32] Though, coastal countries are not necessarily more developed, in terms of their GPD per capita, than landlocked countries as Botswana has by far the highest GDP per capita of the ten countries in our analysis.

Table 1.6: Heterogeneity (coastal countries)

light intensity (log)	(1)	(2)	(3)	(4)	(5)	(6)
post x treated	0.161*	0.165**	0.271***	0.218**	0.214**	0.214**
	(0.0820)	(0.0815)	(0.0884)	(0.0979)	(0.0925)	(0.0918)
GSM coverage		0.0797	0.102	0.0878	0.0873	0.0887
		(0.0622)	(0.0687)	(0.0642)	(0.0639)	(0.0625)
post x distance to next port				0.309		
				(0.227)		
post x treated x distance to next port				-0.406		
				(0.497)		
post x distance to coastline					0.352	
					(0.241)	
post x treated x distance to coastline					-0.482	
					(0.504)	
post x distance to landing point						0.136
						(0.216)
post x treated x distance to landing point						-0.358
						(0.398)
observations	825	825	975	825	825	825
R-squared	0.902	0.903	0.898	0.903	0.904	0.903
#countries	5	5	5	5	5	5
#towns	75	75	75	75	75	75
share treated	.507	.507	.507	.507	.507	.507
town FE	✓	✓	✓	✓	✓	✓
country x year FE	✓	✓	✓	✓	✓	✓
w/o nodal cities	✓	✓	✓	✓	✓	✓

Notes: Light intensity is measured as the logarithmic sum of light intensities of DMPS-OLS pixels within town area and a 2 kilometer buffer as in Table 1.1. GSM mobile coverage is calculated as the percentage share of town area covered with signal. All specifications are estimated on a sample restricted by excluding landing point cities and capitals, regional capitals, and cities with more than 100,000 inhabitants and include town and country-year fixed effects. Ports are source from *OSM*. Robust standard errors clustered by town reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

the fifth year after the treatment. Parallel trends in the pre-treatment periods are again present. In contrast to the main specification with all countries (Figure 1.5), there is a significant increase in the treatment year, before the estimate declines slightly and becomes insignificant again.

A further advantage of coastal countries is that international trade, approximated through the towns' proximity to ports, can be investigated as a potential mechanism. We therefore include an additional interaction term, distance to the next port, in Column (4). The main effect increases slightly and remains statistical significance at the 5% level, while the triple interaction lacks statistical significance and has a negative point estimate. Later means that proximity to a port does not have a significant positive effect on local economic growth and that the mechanism of trade does not play an important role. The additional interaction of the post dummy and the distance to the next port lacks statistical significance as well. Hence, towns closer to a port do not grow faster after the countrywide Internet connection than towns further away from a port. In Columns (5) and (6), we repeat the triple interaction with the distance to the coastline and to the landing point. Results are similar to the triple interaction with the distance to the next port. The distance to the landing point accounts for both the distance the fiber-cable rollout has to cover and for the distance to the country's most important port. The distance to the coastline might reflect general growth potential and economic development of the coastal regions in comparison to the hinterland [Henderson et al., 2012] and also account for smaller ports, which might not be included in *OSM*.

1.6 Discussion

Previous estimates of economic growth induced by broadband Internet serve for comparison with our results. We find for cities with Internet access an increase in economic growth of 2 percentage points.

Czernich et al. [2011] investigate GDP growth induced by broadband Internet in OECD countries. They find that the broadband Internet increased GDP per capita by 2.7 to 3.9 percent, implying a .9 to 1.5 annual per capita growth when Internet penetration is increased by 10 percentage points (with penetration ranging between 13.5 percent in Greece and 37.2 percent in Denmark in 2008). Regarding Internet speed and timing, their study is very comparable to ours. They define broadband if a user can surf with at least 256 Kbps. In comparison, Hjort and Poulsen [2019] state that SSA users had on average 430 Kbps before the 'second generation' of SMCs arrived. Most OECD countries introduced broadband Internet between 1999 and 2000 with some late adopters like Greece (2003) and Ireland (2002). In our study, the first countries were connected in 1999 to 2001. However, many countries were connected in the mid-2000s or even later. Two major differences are that we (i) cannot investigate broadband penetration and (ii) compare cities within countries and not across countries. Though, broadband penetration is very low in SSA, it is likely that the very first adopters, mainly firms, have the biggest impact on economic growth. Kolko [2012] investigates broadband Internet expansion in the US and finds, especially in areas with low population density, a positive effect on local economic growth

For SSA, Hjort and Poulsen [2019] estimate a 3.3 percent increase in light activity for the later arrival of fast Internet. First of all, their work differs by the Internet speed available. But most importantly, while we use variation between towns, they use variation within local cells and not across towns. Hence, though in both cases local economic activity is measured, the comparison is different. Finally, the selection of cities and towns differs slightly as we focus on mid-size towns. All together, it is hard to compare whether the estimates tell something about different speeds or whether they are affected by the named differences. Finally, it cannot be rejected that the effects of the extensive margin, the first connectivity, are still in play when the next generation of SMCs landed. Nevertheless, both studies show that SMCs that brought Internet to SSA at different speeds had both a similar positive effect on local economic growth.

Finally, we want to compare our results to Storeygard [2016] who also estimates local economic growth across cities. Though, not estimating the effects of a digital infrastructure, he is most closely related to our work regarding the outcome measure. Therefore, our estimated effect of Internet availability on a town that is 200 kilometers away from the primate city is equivalent to an oil price shock of 70 US-Dollar.

1.7 Conclusion

Locations can benefit from the Internet to change to a manufacturing industry if digital infrastructure is in place. We investigate if the availability of Internet at basic speeds fosters economic development in developing countries. In particular, we study the arrival of the first sub-marine Internet cables in ten Sub-Saharan African countries in the 2000s. To learn about the causal effect of Internet availability on local economic growth, we compare in a difference-in-differences setting economic activity, measured by nighttime light satellite data, of towns connected to the national Internet backbone at the time of countrywide Internet arrival to a control group of similar towns not (yet) connected to the national digital infrastructure but that get an access point later.

We find that the connection of towns to the *World Wide Web*, on average, leads to an increase in light intensity of about 7 percent, relative to similar towns not (yet) connected. This translates into 2 percentage points higher growth in terms of GDP. Moreover, we differentiate the growth in more pixels, where towns increase in their area (extensive margin), and in a higher average of the light intensity, which is associated with a higher productivity (intensive margin). We find that towns with Internet availability due to access to digital infrastructure typically grow on both margins, i.e., become brighter and increase their size. Furthermore, our results suggest that this growth is only partly driven by growing populations in connected towns. So, the effect is mainly of an economic development and not a migration effect. Finally, we can show that one mechanism that leads to the growth effects is the change of the industry structure. In regions where Internet access exists, manufacturing has higher growth rates. While the industry shares in employment of manufacturing increase, shares of agriculture decrease.

The rollout of new infrastructure is always expensive. Therefore, policy makers might think of saving money and only rolling out this infrastructure where the effects pay off the costs of the rollout. Our study comes in at this point: We show that even smaller towns that were connected incidentally are growing faster than comparable towns without access points to the Internet. Therefore, first, it is important to account for these smaller towns when evaluating the benefits of an infrastructure. Second, one can derive from our results that the Internet has growth potential not only for economic and political centers but also for smaller towns. Moreover, the effects of the Internet are not bound to a high uptake, but the few adopters generate spillovers. Hence, we recommend to rollout this infrastructure further even when only a low, but positive, uptake is expected. An uptake by some firms might generate external effects for the whole town. Moreover, of course, the Internet might have further effects on educational or political outcomes. Hence, there might be other reasons to connect the whole country which are not targeted in this study, but that could be an interesting direction for further research.

Chapter 2

Low Demand Despite Broad Supply: Is High-Speed Internet an Infrastructure of General Interest? *

2.1 Introduction

The Internet is widely regarded as a general-purpose technology as it is used on a large scale and has had significant social and economic effects. Czernich et al. [2011] and Kolko [2012], among others, have shown that the introduction of DSL at basic speeds accelerated economic growth. It is almost considered a truism that higher speeds will again go hand in hand with economic growth. It is thus unsurprising that policy-makers throughout Europe are concerned about raising broadband quality. The Digital Agenda for Europe from 2016 aims to connect all (50 percent) households to broadband Internet with at least 30 Mbps (100 Mbps) by 2020.[1] Coverage rates for these bandwidths already reached 79.0 percent (55.1 percent) in 2017.[2] In contrast to the adoption of Internet at basic speeds, however, the adoption of high-speed Internet remains low: only one-third (one-fifth) of households with higher available speeds have also subscribed to at least 30 Mbps (100 Mbps).[3] This apparent subscription gap raises the question whether high-speed Internet is of general interest or only demanded by a very specific subset of users. Without a broad subscription, it is different from broadband Internet at basic speeds and cannot be described as a general-purpose technology.

There is still little research analyzing Internet at higher speeds. When investigating the effects of broadband Internet, the focus remains on basic speeds. More recent work on higher speeds relates to availability and therefore misses the role of adoption. This chapter establishes a link between availability and adoption of high-speed Internet and provides one

* This chapter is based on work published in *Information Economics and Policy*.
1 https://ec.europa.eu/digital-single-market/en/broadband-strategy-policy
2 https://ec.europa.eu/digital-single-market/en/connectivity
3 https://ec.europa.eu/digital-single-market/en/european-digital-progress-report

explanation for the lack of economic effects of high-speed Internet availability. It analyzes why high-speed Internet is not adopted widely in Austria, despite high coverage rates. I consider specifically contracts of more than 100 Mbps. The adoption of high-speed Internet depends on the substitutability over bandwidths and between stationary fixed-line and mobile technologies. Only if basic and higher speeds are considered substitutes, high-speed Internet will be adopted widely. If basic and higher speeds are not considered substitutes, there is only one specific type of Internet user with high benefits from high-speed Internet. Other users do not have such high benefits and will always choose a cheaper basic contract. In this case, the adoption rate will remain low because the high prices do not match the low utility that most users receive from higher bandwidths. Then, high-speed Internet would not be an infrastructure of general interest.

I apply a mixed logit discrete choice model, from which own-price and cross-price elasticities of demand are derived. While own-price elasticities show how price-sensitive consumers are, cross-price elasticities define substitutes. I apply a cross-sectional setting for Austria with regional differences in terms of active providers at the municipality level. The methodology applied in this chapter is widely used in the market definition literature, where different technologies or bandwidths are analyzed in terms of their substitutability when deciding whether they form a single market or different markets.

I find that high-speed and basic Internet are not substitutes as cross-price elasticities show a different pattern for basic and higher speeds. Users who are satisfied with basic speeds refrain from adopting high-speed Internet even if it is available. I conclude that there is a certain type of users who needs higher speeds. So, increasing the availability of high-speed Internet will only increase the use of high-speed Internet if these users live in the area that gets an upgrade. By contrast, users who do not require higher speeds will always choose the relatively cheap basic-speed alternative, regardless of the availability of high-speed Internet. I find that for lower speeds, fixed-line and mobile technologies are substitutes. At higher speeds, these technologies do not function as substitutes. Furthermore, the derived own-price elasticities indicate that the demand is generally very elastic.

This chapter uses novel geo-referenced data. I use extensive information from around 150,000 Austrian stationary broadband Internet tests conducted on the regulatory authority's web site (RTR-NetTest) between April and November 2016. I observe the selected provider and the measured download and upload speed and can use this information to infer the selected contract. The number of Internet users testing their speed has increased sharply in recent years (Figure B.1.1), making new methods such as the one presented in this chapter feasible.[4] The increase in speed tests is not due to the higher Internet penetration in

[4] The annual increase shows a certain seasonality. In the colder half of the year, the number of tests increases, such that the highest number of tests is always found in December. In the warmer months, households spend more time outdoors and perform fewer tests of their Internet quality.

Austria.[5] Rather, the increase in tests shows that users are no longer satisfied with just being connected to the Internet, but are concerned with the speed at which they can use it. This data also makes it possible to track the increased average download speed (Figure B.1.2). In addition, I observe the price of the selected contract as well as the competitive environment at the household's place of residence, which enables me to define a choice set for each user in the sample.

Austria is examined as a representative country for Europe in terms of (bandwidth) coverage rates and Next Generation Access (NGA) coverage. As in many other countries, there is an incumbent provider (A1 Telekom Austria)[6] that owns the copper network and is under regulation. The incumbent mainly offers high-speed contracts by copper-based VDSL technology. In addition, there are regional cable providers (kabelplus, Tele2, and UPC) that have historically been active in the TV market and therefore own a coaxial cable network. Their NGA technology is DOCSIS 3.0. In recent years, fiber network providers (Salzburg AG and LIWEST) have entered the market with very high-speed contracts at a more local level. Austria's NGA mix contains only to a lesser extent fiber technologies (FTTP). Finally, Austria has one of the strongest mobile networks in Europe. Three providers (Hutchinson Drei, A1 Telekom Mobile, and T-Mobile) offer contracts for mobile broadband Internet at home. Importantly, A1 Telekom Mobile is one of them. So, the incumbent also offers mobile contracts.

This chapter proceeds as follows. Section 2.2 gives an overview of the related literature. Section 2.3 defines the discrete choice mixed logit regression model and how the elasticities are derived. In Section 2.4, broadband coverage in Europe and the case of Austria are explained. Then, I introduce the data sets I use and merge. Section 2.5 presents the estimation results and the price elasticities of demand. Section 2.6 concludes.

2.2 Related Literature

The literature so far has concentrated on the effects of the introduction of DSL, i.e., on Internet at basic speeds (among others Czernich et al. [2011] and Kolko [2012]). Very few more recent studies focus on higher speed (among others Ford [2018] and Briglauer et al. [2019]). Although, these analyses are based on the availability of higher speeds. Additionally, higher speeds refer to higher speeds compared to the introduction of DSL, but not to the currently available speeds of more than 100 Mbps. To the best of my knowledge, this is the first work that analyzes the adoption of high-speed Internet where it is available on

[5] In 2013, already 98 percent of households had a broadband coverage of at least 2 Mbps, and four out of five Austrians could access the Internet from home, meaning that the households did have a valid Internet subscription.

[6] In the following, I will always refer to A1 Telekom Austria as A1 Telekom or A1 TA.

a broad scale. The estimated utilities and price elasticities of demand from this analysis show limitations in the use of high-speed Internet that are necessary before economic effects of higher speeds can be found. Finally, it applies methodology from the literature on market definition in the broadband market and adds here by providing substitution patterns between fixed-line and mobile technologies across different bandwidths.

Starting with the literature on the impact of broadband Internet on economic outcomes, Czernich et al. [2011] identify an effect of broadband infrastructure on annual per capita growth for OECD countries. For the US, Kolko [2012] finds a positive relationship between broadband expansion and local economic growth. He applies the slope of the terrain as an instrument for broadband expansion and detects growth in population, employment, the average wage, and the employment rate. Focusing on labor market effects, Atasoy [2013] discovers that gaining access to broadband services in a county is associated with approximately a 1.8 percentage points increase in the employment rate. This contrasts with Czernich [2014], who finds no effect on the unemployment rate for Germany. On the firm level, Akerman et al. [2015] identify different broadband Internet effects depending on the skill level of workers on labor market outcomes and productivity for Norway. For Germany, Bertschek et al. [2013] detect broadband Internet effects on the firms' innovation activity, but not on their labor productivity. Colombo et al. [2013] find that the productivity performance of small and medium enterprises in Italy is not influenced by basic broadband applications. However, depending on the sector, advanced broadband applications do influence productivity. On broadband adoption, Grimes et al. [2012] discover an increase in firm productivity. All of these studies use basic DSL at lower speeds than are used in this chapter.

More recent studies analyze various broadband speeds. Ford [2018] compares 10 Mbps and 25 Mbps in the US and finds no economic payoff at the county level in terms of jobs and income. Similar results are shown in a correction of Bai [2017] by Whitacre et al. [2018]. For Germany, Briglauer et al. [2019] evaluate the Bavarian state aid program for speed upgrades in broadband Internet availability. They detect no effect on the employment rate. However, 16 Mbps and above is the highest category. Therefore, I investigate whether there might not exist enough users with a high utility from high-speed Internet.

Methodologically, this chapter builds on the literature on market definition in the broadband market. So far, the literature has concentrated on the investigation of potential technology-level boundaries. Early papers by Crandall et al. [2002], Rappoport et al. [2003], and Flamm and Chaudhuri [2007] analyze the demand for dial-up, DSL, and cable modem in the US and focus on the substitution between narrowband and broadband. Flamm and Chaudhuri [2007] analyze cross-price elasticities of demand and find that dial-up and broadband are substitutes. Similar results are found by Pereira and Ribeiro [2011] for Portugal.

In a more recent study, Cardona et al. [2009] also analyze the Austrian broadband market applying a nested logit model. They find a strong substitution between DSL and cable where cable exists and suggest adding cable to the DSL market. Additionally, they exploit the fact that about half of Austria is connected to the cable network. They detect higher elasticities in areas connected to cable network than in areas without the possibility to access cable network. Stationary mobile Internet was not yet very important in the market. Therefore, the authors, expecting it to gain importance, leave it to further research to discuss the substitution pattern between fixed-line and mobile Internet. Other more recent studies, which increasingly focus on whether fixed-line and mobile technologies are substitutes, were presented by Srinuan et al. [2012] and Grzybowski et al. [2014] for Sweden and Slovakia, respectively. The derived price elasticities point towards a high price-sensitivity for all technologies, indicating them to be close substitutes. While Cardona et al. [2009] and Srinuan et al. [2012] apply a nested logit model, Grzybowski et al. [2014] apply a mixed logit model. They all use survey data. Here, this work contributes by providing substitution patterns across speed and technology as first indications of future market definition for the broadband Internet market in Austria.

2.3 Discrete Choice Model

I apply a discrete choice model following McFadden and Train [2000], which is the standard methodology in the literature on market definition. Internet subscribers choose a contract with a specific provider, bandwidth, and price. All available contracts are defined as the choice set's alternatives. The alternatives form an exhaustive and mutually exclusive discrete choice set.

The most basic regression method in this context would be a multinomial logit model. However, it imposes the independence of irrelevant alternatives (IIA) property. Since the IIA might not be guaranteed, I apply a more general approach, a mixed logit regression, which relaxes the IIA by allowing correlation of choices between the alternatives. This model allows for unobserved heterogeneity among individuals, which might apply in this case as I observe only very few individual characteristics of the Internet subscribers. Moreover, Internet subscribers do not only consider speed when making their choice, but also other characteristics, in particular the price. Therefore, to account for the unobserved heterogeneity among Internet subscribers, a random coefficient for the price is applied.

2.3.1 Choice Set

The choice is defined by the contract selected by the Internet subscriber. The alternatives are represented according to speed categories for fixed-line and mobile providers, respectively. There are seven alternatives in total: surfing at a maximum of (i) 20 Mbps, (ii) 40 Mbps, (iii) 80 Mbps, or (iv) at more than 80 Mbps with a fixed-line provider and (v) surfing at a maximum of 20 Mbps, (vi) 80 Mbps, or (vii) at more than 80 Mbps with a mobile provider. Note that I do not differentiate among providers nor technologies (DSL, cable, and fiber) within each alternative.

2.3.2 Utility

The utility V_{ij} is defined for each Internet subscriber i and each alternative j. It depends on subscriber- and alternative-specific valuations (β_j and γ) and the Internet subscriber's price sensitivity $\widetilde{\alpha_i}$:

$$V_{ij} = \widetilde{\alpha_i} p_{ij} + \beta_j x_i + \gamma z_j + u_{ij},$$

where x_i is a vector of subscriber-specific variables and z_j is a vector of alternatives-specific variables. Finally, u_{ij} is the logit error term, which is, as suggested by theory, identically and independently distributed across contracts according to the Type I extreme value distribution. The random price coefficient $\widetilde{\alpha_i}$ accounts for unobserved heterogeneity among the individual Internet subscribers. Assuming a normal distribution ($\widetilde{\alpha} \sim N(\alpha, \Sigma)$), $\widetilde{\alpha_i}$ is calculated as follows:

$$\widetilde{\alpha_i} = \alpha + \sigma_\alpha v_i, \tag{2.1}$$

where α is the price coefficient's mean valuation, σ_α refers to its standard deviation and v_i is a random variable with a standard normal distribution ($v \sim N(0,1)$).

2.3.3 Choice Probabilities

I assume that Internet subscribers maximize their utility. With the previously defined utilities, the individual choice probabilities for each category can be calculated as:

$$l_{ij}(\widetilde{\alpha_i}) = P(V_{ij} = \max_{k \in C_i} V_{ik}) = \frac{\exp(\widetilde{\alpha_i} p_{ij} + \beta_j x_i + \gamma z_{ij})}{\sum_{k \in C_i} \exp(\widetilde{\alpha_i} p_{ij} + \beta_j x_i + \gamma z_{ij})},$$

where C_i is the choice set for each Internet subscriber i as defined above. The last equation follows from the distributional assumptions of the logit error term u_{ij}. The mixed logit model with unobserved heterogeneity requires integration over the distribution over $\widetilde{\alpha_i}$, the random coefficient, which has to be simulated:

$$s_{ij} = \int_{\widetilde{\alpha_i}} l_{ij}(\widetilde{\alpha}) f(\widetilde{\alpha}) d\widetilde{\alpha}.$$

The estimate $\widehat{s_{ij}}$ is then applied to the maximum likelihood estimation:

$$\mathscr{L}(\theta) = y_{ij} \sum_i \sum_j \log(\widehat{s_{ij}}),$$

where y_{ij} equals 1 if individual i has chosen alternative j and 0 otherwise.

2.3.4 Price Elasticities of Demand

I am interested in own- and cross-price elasticities for the choice set averaged on the country-level. These elasticities are based on the regression's outcome, the choice probabilities, which were derived before, and the prices. From the mixed logit regression, the random coefficient for the price $\widetilde{\alpha_i}$ is simulated according to Equation (2.1). The regression output is applied as well for the choice probabilities $\widehat{s_{ij}}$. The individual own-price elasticities of demand are then defined as follows:

$$\epsilon_{ij}^{indv} = \frac{\delta s_{ij}}{\delta p_{ij}} p_{ij} = \widetilde{\alpha_i} s_{ij}(1 - s_{ij}) p_{ij}. \tag{2.2}$$

As in Grzybowski et al. [2014], I calculate elasticities of demand instead of semi-elasticities of demand, i.e., I do not simply average over individual's price elasticities but weight these by the individual's choice probabilities. Hence, following the individual's own-price elasticity (Equation (2.2)), the country level elasticity is:

$$\epsilon_{ij}^{country} = \frac{\sum_i \widetilde{\alpha_i} s_{ij}(1 - s_{ij}) p_{ij}}{\sum_i s_{ij}}. \tag{2.3}$$

The cross-price elasticities are calculated accordingly from the individual cross-price elasticities:

$$\epsilon_{ik}^{indv} = \frac{\delta s_{ij}}{\delta p_{ik}} p_{ik} = \widetilde{\alpha_i} s_{ij} s_{ik} p_{ik},\ for\ k \neq j.$$

2.4 Data on Broadband Alternatives

2.4.1 Broadband Coverage in Europe and the Case of Austria

The broadband coverage data for Europe shows that across countries the mix of technologies for their Next Generation Access (NGA) to provide high bandwidths to users varies. These technologies include copper-based (VDSL), fiber (FTTP), and coaxial cable (DOCSIS 3.0) technologies. Table B.2.1 shows that NGA coverage is 80.1 percent in 2017 (46.9 percent in rural areas).[7]

Austria is examined because it is a representative country in terms of coverage rates. The coverage of households connected with at least 30 Mbps (100 Mbps) is 81.1 percent (57.2 percent), which is marginally above the European average. NGA coverage has generally increased from 69.5 percent in 2012 to 90.0 percent in 2017 and has more than tripled for rural areas over this period. This increase was mainly due to the expansion of VDSL and DOCSIS 3.0.[8] Although the availability of FTTP in Austria is below the EU average, above-average coverage was achieved for connection speeds of at least 30 Mbps and at least 100 Mbps. Moreover, Austria has a strong mobile network and thus a high demand for stationary mobile contracts. This makes Austria an even more interesting case. I, therefore, consider Austria to be a good representative for this study.

2.4.2 Spatial Broadband Data

I exploit usage data from the national regulators from speed tests, network providers, and coverage data from the *Broadband Atlas* as well as broadband Internet contracts collected by *AK-Tarifwegweiser*. Although prices do not vary a lot over time, there are changes in the contracts offered with regard to download (and upload) speed. Taking the period from April to November 2016, there are no new contractual bandwidths and very few changes in prices. Therefore, tests can easily be assigned to the contracts. In addition, almost the entire NGA roll-out was privately financed during this period. Therefore, I can assume that the network providers did not expect much demand for high-speed Internet in the locations that did not have full coverage and are therefore not included in my analysis. Hence, my sample, which is limited to tests performed between April and November 2016, is ideal for my research question.

[7] Europe in particular has a coverage with VDSL of 53.4 percent (32.5 percent in rural areas), with FTTP of 26.8 percent (11.3 percent) and with DOCSIS 3.0 of 44.7 percent (10.8 percent).

[8] VDSL coverage increased from 50.5 percent to 82.2 percent and from no availability to 22.1 percent for its rural areas. The expansion of DOCSIS 3.0 was moderate and similar in size for both rural and urban areas. FTTP continues to play a minor role in providing fast broadband access to Austrian households. The coverage rate of FTTP rose from 6.3 percent to 13.5 percent and from 1.2 percent to 5.4 percent for rural areas, which is a small fraction compared to the EU 28 average of 23.7 percent.

2.4.2.1 Usage Data

From the RTR NetTest, I derive information on broadband usage throughout Austria from around 150,000 speed tests, which were performed by Internet subscribers.[9] Besides the exact timing of the individual test (date and time), importantly, the data provides the exact (geo-coordinates of the) test location and the network provider, establishing the Internet connection. Information about the measured upload and download speed is also included. Finally, it provides a rich set of additional information, such as the type of Internet connection (wireless LAN/LAN).[10]

I limit the data to tests of stationary Internet connections.[11] This includes tests of both fixed-line and mobile networks. Yet, tests of stationary Internet connections that are established via mobile networks are restricted to offers designed for the use *at home* rather than standard smartphone contracts. While the former usually include unrestricted broadband use with a certain bandwidth, smartphone contracts are limited in terms of consumable volume and bandwidth.

I address the issue of sample selection. This problem might occur if only users with an incorrect working Internet connection perform the test. However, this case is handled by reducing the sample to tests close to an existing contract. Although I have information on the date of the tests, this might not be the date on which the testers signed their contract. I assume that if the test was not performed due to a faulty Internet connection, most tests were carried out shortly after the contract was signed to ensure that the provider would deliver the contractually agreed speed. If there is a large number of tests with newly signed contracts, my results are not representative of all contracts but explain the behavior of a subgroup of Internet users who have recently signed a new contract. However, for policy-makers, this subgroup could be the relevant group to be interested in their behavior.

2.4.2.2 Broadband Coverage and Network Providers

From the Austrian *Broadband Atlas*, I use fine-grained (100 × 100 meters) data on the maximum available download speed in mid-2015 via fixed-line networks, which is provided

[9] The data is publicly available for download on the website of the Austrian regulatory agency (RTR - Rundfunk und Telekom Regulierungs-GmbH): https://www.netztest.at/en/. By providing a web browser tool and mobile applications (for Android and iOS), RTR enables users to test the quality of their current mobile and stationary Internet connections.

[10] If the test was performed via LAN, the browser in use is transmitted, and if the test was performed via wireless LAN, the model of the device in use is transmitted.

[11] Additionally, if a tester has performed several tests in one session, only the last test is taken and the others are removed from the data. I only observe the network provider and not the actual service provider, i.e., providers without their own infrastructure are not shown in the data. For example, A1 Telekom, the Austrian incumbent, is obliged to grant service providers access to its network, e.g., by unbundling the local loop.

by the Ministry for Transport, Innovation and Technology (bmvit).[12] Similarly, data on mobile coverage is provided as geo-data, indicating which Austrian regions are covered by a certain bandwidth.[13] While high-speed Internet access via fixed-line concentrates in rather urban areas, mobile broadband is distributed throughout Austria.

Apart from data on available bandwidths across the country, the *Broadband Atlas* lists all network providers at the municipality level. Among them, I consider the fixed-line providers A1 Telekom, kabelplus, Tele2, UPC, Salzburg AG, and LIWEST. It is important to note that these providers have different network technologies and most of them are only active in certain regions. While A1 Telekom, as the incumbent, owns the copper network throughout Austria, kabelplus, Tele2, and UPC are active with their coaxial networks in certain regions. The same applies to the fiber optic networks of Salzburg AG and LIWEST. Of the mobile providers, only Hutchinson Drei, A1 Telekom Mobile, and T-Mobile offer stationary contracts and are therefore included in the analysis. These fixed-line and mobile providers account for around 90 percent of all tests from the usage data.

2.4.2.3 Contract Data

From *AK-Tarifwegweiser*, I gather contract data. The data contains monthly information on all contracts offered by the major network providers mentioned above as well as by service providers without their own network.[14] Apart from the price, the data includes information on maximum download bandwidth, maximum capacity[15], and duration. I supplement this data with information on maximum upload bandwidth from the providers' websites and the Internet archive *Wayback Machine*.[16] For each contract, I apply the minimum price per month and average over the whole period. I only consider unlimited capacities or with stated *fair use* of at least 150 GB per month.

In Tables B.2.2 and B.2.3, contract information for the above-mentioned providers can be found. These include price (in Euro) and upload speed (in Mbps) for each offered download bandwidth (in Mbps). The contract data is visualized in Figure B.1.3 for the case of A1 Telekom, where each line represents one download speed.

The contractual maximum download speed varies between 8 and 250 Mbps, and largely depends on the provider. For example, only UPC and kabelplus offer contracts with

[12] The maximum attainable bandwidth is assigned to a specific network cell as soon as at least one provider can technically realize the bandwidth in at least one household in that cell.

[13] Yet, actual attainable bandwidths, especially via mobile networks, can deviate considerably from the theoretical availability, depending on the number of users and the intensity of use within a particular mobile cell.

[14] MMC, A.K.I.S., Comteam, CNet, DIC, Telematica, and TeleTronic, among others

[15] All contracts considered are unlimited, with the exception of the contracts of Salzburg AG, which contain a *fair use* policy.

[16] https://archive.org/web/

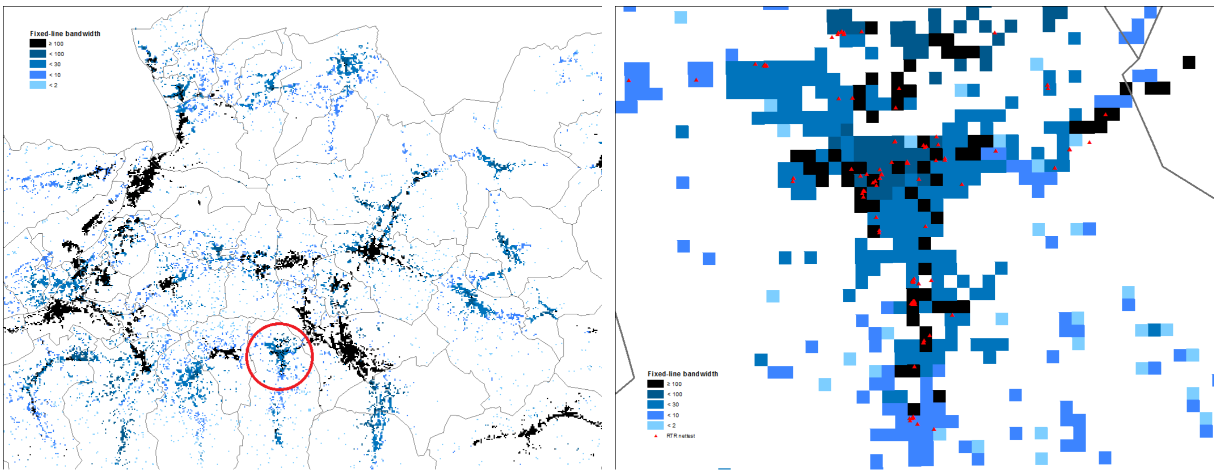
250 Mbps. This contract offered by kabelplus is relatively expensive, whereas the one offered by UPC is within the price range of contracts from other providers such as A1 Telekom or LIWEST, which have a contractually agreed download speed of 100 Mbps or 150 Mbps respectively and cost slightly less than 60 Euros. In general, prices vary between 17.8 and 99.9 Euros.

2.4.2.4 Estimation Data with the Choice Set

The linkage of the data sets is illustrated in Figure 2.1. First, the municipalities, with their information on active providers and the providers' contracts, are matched to the fixed-line coverage grid (left panel). Then, each test is related to its cell and the information assigned above (right panel), resulting in a data set at the household level. This data contains not only information on the selected contract, i.e., the provider and bandwidth, but also on the maximum available bandwidth. It also identifies possible alternative providers in addition to the one serving the household.[17]

Figure 2.1: Illustration of the spatial merge

(1) Municipalities and fixed-line coverage (2) Fixed-line coverage and tests in a municipality



Notes: The figure illustrates the municipality boundaries (gray lines), broadband coverage on grid level (blue cells), and the test locations of households (red dots). The circle in the left figure marks the municipality in the right one. Data provided by GfK GeoMarketing GmbH, Ministry for Transport, Innovation and Technology (bmvit), RTR

The combined data set allows me to compare the distribution of download speed by fixed-line coverage (Figure B.1.4). Up to a maximum fixed-line coverage of 50 Mbps, the median of the measured download speed is rather close to the maximum fixed-line coverage.

[17] Unfortunately, the degree of accuracy of the data deviates from an ideal setup. Preferably, I would want to know i) which providers a household could connect to, rather than the providers active in the household's municipality, and ii) which bandwidth each provider can offer in the grid, rather than the maximum available bandwidth of all providers collectively. A minor issue is iii) that coverage is not available for the location of the household, but rather for a small area around the household.

However, with a higher maximum fixed-line coverage (100 Mbps or >100 Mbps), the median of the measured download speed does not increase anymore.[18] These figures provide first descriptive evidence that up to 50 Mbps many consumers are restricted by the maximum fixed-line coverage, so that they would sign contracts with higher download speeds if they were available. By contrast, many consumers do not sign contracts with download speeds above 50 Mbps, even if they are available.

A comparable set of choices is needed for the regression. Therefore, tests conducted in a cell with less than the maximum fixed-line coverage (at least 100 Mbps) in 2015 are removed. Eventually, the data comprises to 64,182 tests.[19]

The mixed logit model requires a choice variable and a choice set containing all alternatives. The choice is assigned based on the information on download and upload speed as well as the chosen network provider, and the location where the test was performed.[20] The choice is defined as the closest contract with respect to the measured upload and download speeds from the displayed provider.[21]

The choice set is defined as four alternatives for fixed-line and three for mobile contracts.[22] The lowest alternative allows a maximum of 20 Mbps download bandwidth and the highest alternative includes contracts with a download bandwidth of more than 80 Mbps, which are in fact contracts with a three-digit download bandwidth (≥ 100 Mbps). Therefore, the alternative with more than 80 Mbps can be considered as the very high bandwidth alternative. Alternatives are formed from several contracts, as it is unlikely that a consumer would consider all available contracts as alternatives. These alternatives make it possible to answer questions on substitution patterns within fixed-line contracts and between them

[18] Up to a maximum coverage of 30 Mbps, the median of the measured download speed is at least 50 percent of the maximum coverage. At 10 Mbps, it is even more than 70 percent. At a maximum coverage of 100 Mbps, the median of the measured download speed is only 26 percent of the maximum coverage.

[19] Tables B.2.4 through B.2.6 shows in more detail how the sample shrinks when it is reduced for reasons of measurement or limited fixed-line coverage, particularly with regard to the number of tests from each provider. Summary statistics comparing the selected observations and the removed ones can be found in Table B.2.7.

[20] I update information on active providers and fixed-line coverage from the usage data. If a test transmitted a provider that is not listed in the *Broadband Atlas* in that municipality, the municipality data is updated to include that provider. Similarly, if a test is run at a higher download speed than it is covered in this specific cell, the information for all tests in this cell is updated.

[21] Contracts must allow at least the measured upload and download speeds. An exception is made if the measured speed is above the maximum contract offered by a provider. In these 2,105 cases, the test is assigned to this maximum contract. Closest is defined by the lowest Euclidean distance. Regarding the problem that for A1 Telekom tests I cannot distinguish between fixed-line and mobile tests, the contracts for mobile and fixed-line are so different that I can follow the described procedure.

[22] As explained in Section 2.3, the alternative borders start at 20 Mbps and double twice, resulting in alternative borders at 40 Mbps and 80 Mbps for the fixed-line contracts. It is important that all different fixed-line technologies are present in each fixed-line alternative. As there exist fewer mobile contracts, borders for the alternatives are drawn differently: The lowest alternative border is 20 Mbps (with contracts of 10 Mbps and 20 Mbps) and the middle alternative border is 80 Mbps (with contracts of 30 Mbps, 40 Mbps, and 50 Mbps).

and mobile contracts. For chosen alternatives, I apply the price from the contract. The average price and its standard deviation for the whole sample are 31.13 and 10.84 Euros, respectively.[23] For alternatives, I apply a weighted average price over the available contracts for each alternative. I take the weights from the number of times a choice has been made in the area starting with the same zip-code digit such that I can account for regional variations in active fixed-line providers and in the preference for mobile providers.

Figure B.1.5 shows the speed distribution across selected providers. Most A1 Telekom tests fall into the lowest bandwidth alternative, while most UPC tests fall into the two highest bandwidth alternatives. The distribution reflects the prices in Table B.2.2: UPC offers the cheapest high bandwidth contracts, while low bandwidth contracts are more expensive compared to other providers. A1 Telekom could still have market power within the lower bandwidths. However, the market share decreases sharply at higher bandwidths.

Determining the choice is complicated due to the fact that the test data contains noise (Figure B.1.6). Thus, a measured speed might deviate dramatically from the one assigned in the contract.[24] In pursuit of highest cleanliness possible, the data is restricted to tests with a realized download and upload speed within a certain corridor below a contract offered. The corridors are defined differently for fixed-line and mobile providers (Figure B.1.7). They also reflect the assumption that the measurement of upload speed is more reliable than the measurement of download speed.[25]

2.5 Different Types of Internet Users

The results of the alternative-specific mixed logit regression are briefly discussed before focusing on the elasticities of demand to analyze substitution patterns. I show that there are different types of broadband users: Consumers with a high (low) utility from high-speed Internet, those who have a contract in the second highest (lowest) fixed-line category, upgrade (downgrade) their contracts to the highest (lowest) category if the price of their contract is increased. Hence, an increase in broadband availability will only marginally

[23] For fixed-line choices, the average price and its standard deviation are 30.30 and 11.50 Euros, respectively. For mobile choices, the values are very similar at 32.01 and 9.90 Euros.

[24] While a particular tester is not allowed to surf the Internet at a speed higher than the contractually assigned maximum bandwidth, in many cases she will surf at a speed below the latter. For example, an unfavorable location of the broadband router can lead to actual download and upload speeds being lower than the contractually assigned ones if a wireless LAN connection is used instead of a LAN connection.

[25] For fixed-line providers, the corridors are either a deviation of 15 percent in the download speed and 30 percent in the upload speed or 15 percent in the upload speed and 50 percent in the download speed. This means that if the download speed was measured accurately, the upload speed may differ more and vice versa. For mobile providers, a deviation of 40 percent either in upload or in download from the contract is allowed. This broader restriction is necessary because mobile tests differ more in both upload and download speeds. However, it is still legitimate, as there are fewer different mobile contracts and therefore a more precise allocation of choices is possible.

increase high-speed usage. Moreover, mobile users are rather different from fixed-line users. However, this holds stronger for higher speeds than for basic Internet, where users might substitute between mobile and fixed-line.

2.5.1 Price Sensitivity and Substitution Patterns

2.5.1.1 Regression Results

Prior to the regression, I reduce the sample to tests with an available fixed-line coverage of at least 100 Mbps in 2015. Furthermore, all regressions contain weights with respect to the number of mobile and fixed-line contracts throughout Austria and robust standard errors are applied. As described in Section 2.3, the price of the alternatives is the weighted average of all contracts of all providers in an alternative. The weights are defined by the number of chosen providers within a region (first zip-code digit). Specifications with the unweighted average and without weights are provided as robustness checks.

Before presenting results of the mixed logit regression, I rule out alternative approaches for which the IIA does not hold. Starting with a multinomial logit model, I run the Hausman-McFadden test, excluding alternatives in different ways. Excluding single alternatives, the Hausman-McFadden test shows statistically very significant violations of the IIA, especially for the high-speed alternatives (both for fixed-line and mobile technologies). Between the broader categories, fixed-line and mobile, the Hausman-McFadden test shows a statistically very significant violation when all mobile choices are excluded. However, this could be handled applying a nested logit model. Still, within fixed-line alternatives, the Hausman-McFadden test shows statistically very significant violations of the IIA for the lowest and highest speed alternatives. Within mobile alternatives, the Hausman-McFadden test cannot unambiguously reject a violation of the IIA for all alternatives. Therefore, I reject the multinomial logit model and apply the mixed logit model as the main specification. For robustness, I show some estimates with the nested logit model.

In the main specification, I control for whether the device from which the test was performed was connected via wireless LAN or LAN. Furthermore, the first zip-code digit and the mobile upload coverage from the *Broadband Atlas* are applied as geographical covariates. In later regressions, I apply further control variables at the zip-code level for robustness.[26]

[26] These variables control for population size, level of education, gender and age distribution, proportion of newborns, and (un-)employment/retirement status.

Table 2.1: Mixed logit regression results

	(1)	(2)	(3)	(4)
price	-0.176***	-0.194***	-0.196***	-0.196***
	(0.000635)	(0.000621)	(0.000606)	(0.000606)
sd(price)	0.104***	0.127***	0.127***	0.128***
	(0.000354)	(0.000392)	(0.000399)	(0.000400)
basic controls	✓	✓	✓	✓
further zip-code level controls			✓	✓
all zip-code level controls				✓
sample	whole	reduced	reduced	reduced
observations	449,274	234,514	234,514	234,514
cases	64,182	33,502	33,502	33,502

Notes: The regression table shows the results for the whole (Column (1)) and the reduced sample (Columns (2) through (4)) with four fixed-line and three mobile alternatives. The sample is reduced to tests that had full fixed-line coverage (at least 100 Mbps) in 2015 to ensure that all alternatives are eligible and that the results are comparable. Weights with respect to the number of mobile and fixed-line contracts throughout Austria and robust standard errors are applied. It is controlled for whether the device from which the test was performed was connected via wireless LAN or LAN. Furthermore, the first zip-code digit and the mobile upload coverage from the *Broadband Atlas* are applied as geographical covariates. In Column (3), population size, gender, (un-)employment, retirees, and age structure are added at the zip-code level. In the last column, level of education and the proportion of newborns at the zip-code level are added. Standard errors are shown in parentheses. * p<0.05, ** p<0.01, *** p<0.001

In Table 2.1, the main specification for the whole (Column (1)) and the reduced sample (Column (2)) is presented.[27] The random coefficient has a mean valuation of $\hat{p} = -.176$ and $-.194$ and a standard deviation of $\hat{\sigma}_p = .104$ and $.127$ (all highly significant). A lower standard deviation (in absolute values) than the mean is revealed, which yields a negative $\widetilde{\alpha_i}$ for most of the observations, meaning that consumers are price-sensitive. In Columns (3) and (4), further control variables are added, which does not change the results. Therefore, I will concentrate on Column (2), as this model is more sparse. The coefficients for the choices are presented later with a focus on the socio-economic implications.

2.5.1.2 Price Elasticities of Demand

I present the own-price and cross-price elasticities calculated according to Equation (2.3) for the reduced sample, applying the estimates from the Column (2) of Table 2.1. Table 2.2 shows the mean values over 100 repetitions of an each time newly simulated random coefficient.

[27] For robustness, I show the same specifications estimated with a nested logit model. The price coefficient is very similar. However, the dissimilarity parameters show that the IIA is violated. Therefore, I follow showing my results only applying the mixed logit model. The results of the nested logit model are shown in Table B.2.8.

Table 2.2: Own- and cross-price elasticities

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-2.992	1.204	.931	1.187	.283	.698	.092
≤ 40 Mbps	1.323	-4.542	.994	1.202	.31	.808	.191
≤ 80 Mbps	1.039	.988	-5.562	1.744	.26	.66	.179
> 80 Mbps	.969	.883	1.308	-7.105	.25	.711	.227
mobile ≤ 20 Mbps	1.025	1.008	.841	1.079	-3.691	1.191	.759
mobile ≤ 80 Mbps	.666	.701	.577	.839	.314	-4.251	1.582
mobile > 80 Mbps	.109	.207	.196	.324	.256	2.011	-4.549

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients from the reduced sample (Table 2.1, Column (2)) and according to Equation (2.3).

Own-price elasticities show demand changes following a price increase for a particular choice, whereas cross-price elasticities detect movement patterns between the alternatives. The cross-price elasticity is defined as the demand increase of each contract if the price of the alternative in a specific row is increased by 1 percent. Hence, for each row, it can be seen how the demand for the contract in that specific row is affected by a price increase. Cross-price elasticities can also be interpreted column by column. For each alternative, it is possible to detect the alternative from which the highest demand increase comes from.

The diagonal in Table 2.2, the own-price elasticities, shows that all alternatives are elastic, with values starting from -2.99 for the lowest fixed-line alternative. Consequently, an increase in price by 1 percent reduces the demand for the lowest fixed-line alternative by 2.99 percent. For other alternatives, the demand decreases even stronger.[28] The own-price elasticities for the mobile alternatives are comparable with the respective fixed-line alternatives. Similar elasticities are found in the literature for substitution between technologies. For example, Grzybowski et al. [2014] find elasticities between -1.5 and -4.9, Cardona et al. [2009] find elasticities of -2.5 for all three technologies.

Cross-price elasticities are more important to detect substitution patterns.[29] Same technologies (fixed-line or mobile) are principally more likely to be substitutes. If the price of one of the fixed-line alternative is increased, the other fixed-line alternatives rather than the mobile alternatives are more demanded. For instance, the cross-price elasticities for fixed-line alternatives in the first row (Columns (2) through (4)) are higher than those of the mobile alternative (Columns (5) through (7)). The same applies in reverse for mobile alternatives. However, here it is more present for the two higher alternatives with cross-price elasticities of 1.58 and 2.01.

[28] The own-price elasticity for other fixed-line alternatives decreases with the bandwidth to a value of -7.11. It is defined by the choice probability and the price. Therefore, conceptually, choices with higher prices are less likely to be chosen. Nevertheless, the choices are also less demanded partly because of the higher price. Hence, the relatively high value of -7.11 reflects that as prices increase, the usage of higher speeds does not decrease uniformly but stronger.

[29] However, one has to take into account the correlation within columns. The correlation indicates a generally higher preference for certain alternatives, such as the fixed-line > 80 Mbps alternative, which always has higher values than the fixed-line ≤ 80 Mbps alternative.

The best substitute within the technology is the alternative with the closest bandwidth. For the alternative with the lowest (highest) bandwidth, the next higher (lower) alternative, again within the technology, has the highest cross-price elasticity.

The alternatives in between show the substitution patterns relevant for this analysis. Within the fixed-line technology, for the ≤ 80 Mbps alternative, the cross-price elasticity points upwards and for the ≤ 40 Mbps alternative downwards.[30] This shows that between these alternatives users are different in their utility, they receive from high-speed Internet. The lowest two fixed-line alternatives and the highest two fixed-line alternatives are the closest substitutes within each other. It is therefore appropriate to define the low-speed user for all fixed-line contracts up to 40 Mbps.

For mobile, three alternatives are defined. For the middle alternative (≤ 80 Mbps), the cross-price elasticities point upwards. For the lowest mobile alternative, substitution also exists between technologies. The two highest mobile alternatives are the closest substitutes. If the price of one of these alternatives is increased, the highest fixed-line alternative will receive a higher increase in demand than the lowest mobile alternative. Hence, the lowest mobile alternative is demanded by a different type of users than the higher mobile alternatives, but by the same users as the lower fixed-line alternatives.

In conclusion, as there is substitution from the lowest mobile alternative to basic fixed-line technologies, the low-speed type of users is defined for all technologies. For high-speed Internet, fixed-line and mobile contracts are signed by different users.

2.5.2 Robustness and Heterogeneity

2.5.2.1 Robustness

First robustness checks show that the results are not determined by the decisions of the researcher in defining the sample and the estimation specifications. For the whole sample, elasticities look very similar (Table B.2.9) compared to the reduced sample elasticities. Also, adding more control covariates at the zip-code level does not change the elasticities much (Tables B.2.10 and B.2.11).[31] I apply weights when calculating average prices to account for how often providers are chosen. Taking the unweighted average, less frequently chosen providers might become too important. Nevertheless, I present the elasticity results for the

[30] For the fixed-line ≤ 80 Mbps alternative (third row), the cross-price elasticity of the > 80 Mbps alternative is 1.744, whereas the respective value for the ≤ 40 Mbps alternative is with .988 smaller. For the fixed-line ≤ 40 Mbps alternative (second row), the cross-price elasticity of the ≤ 20 Mbps alternative is 1.323, whereas the respective value for the ≤ 80 Mbps alternative is with .994 smaller.

[31] At the zip-code level, I control for population size, gender, (un-)employment, retirees, and age distribution (as in Table 2.1, Column (3)). I also add the proportion of newborns and the level of education as controls (as in Table 2.1, Column (4)).

unweighted average prices with the reduced sample in the Appendix (Table B.2.12). Here, again, the elasticity results are very similar to the previous ones.

I also apply regression weights with respect to fixed-line and mobile contracts. The sample contains only tests and might not be representative for these technologies. One can imagine that mobile networks might react strongly to seasons, weather, and most importantly to the correct position of the router. Therefore, tests with stationary mobile contracts might occur more frequently in my data. Elasticity results without these weights suggest that elasticities are somewhat lower (Table B.2.13). However, the patterns remain the same. In addition, one might think that A1 Telekom could be signed more often due to its incumbent position. Adding a preference for the choice of the incumbent slightly increases the elasticities (Table B.2.14). Nevertheless, the results still look very similar to the initial results.

Since it has been shown that the results are not determined by the decisions of the researcher, I continue with further robustness checks before turning to the heterogeneous effects. I apply several random data splits: into two halves (Tables B.2.15 and B.2.16) and a randomly selected 5 percent subset (Table B.2.17). With the smaller samples, I can show that the results do not depend on my choice, which has already been shown above, but also hold in smaller randomly selected samples. Furthermore, since the sample is quite large, I can show that significant estimates do not come from a large number of observations, but also hold in smaller samples. In all subsets, the results look very similar.

Next, the number of tests per municipality is restricted. I present elasticities for municipalities with more than 10, 20, and 50 tests (Tables B.2.18 through B.2.20). In these municipalities, a few tests may already generate an outlier. The removed municipalities might also be rather small, corresponding more to rural and poorer locations. It can be seen that the elasticities become only slightly higher. However, the substitution patterns for the cross-price elasticities are still very robust. The same is true when the municipalities with very high number of tests are removed (Tables B.2.21 and B.2.22), which might be very dense and urban locations. In this case, the elasticities become somewhat lower. These small changes could be explained by increased competition in more urban areas. Nevertheless, the results regarding substitution patterns in cross-price elasticities are again very robust.

2.5.2.2 Heterogeneity

Heterogeneous effects are observed in a regional and a time dimension. With regard to regional differences, I analyze Vienna separately and compare it with the rest of Austria. Vienna is the capital of Austria and with 1.9 million inhabitants the largest city in Austria. Moreover, Vienna is significantly larger than the next largest city Graz (with less than 300,000 inhabitants). Due to a higher number of active providers, competition in Vienna is higher. It can, therefore, be seen that the own-price elasticities are also higher in absolute terms: People are more price-sensitive. However, the substitution patterns in the cross-price elasticities remain the same. Hence, I observe the same substitution patterns in Vienna (Table B.2.23) and the rest of Austria (Table B.2.24). For the rest of Austria, I compare the larger cities (Graz, Linz, Salzburg, and Innsbruck), each of which has more than 100,000 inhabitants, and the other regions. For the cities combined (Table B.2.25), the own-price elasticities are again higher than in the regions without larger cities (Table B.2.26), as there is less substitution where there is less competition. This might potentially be even more true for the not-included regions with less fixed-line coverage.

Next, the data is split along the test date. I divide the data into four groups, each containing data for two consecutive months. Tables B.2.27 through B.2.30 show a decrease in own-price elasticities. However, the own-price elasticities do not fall below a certain level of unimportance and can still be explained by random variation. Only the last table for October and November has higher values again. During these months, more tests were performed compared to the previous periods. A longer time horizon is necessary to obtain a clearer picture of these time trends.

Finally, I examine the utilities across the controls. Table B.2.31 shows the estimates for all applied covariates for all alternatives. The estimates refer to the base alternative (fixed-line ≤ 20 Mbps). Before I come to the socio-economic characteristics, the base controls (mobile upload coverage, first zip-code digit, and wireless LAN) are interpreted. The higher the mobile upload coverage is, the higher is the utility of a mobile alternative (first row, Columns (4) through (6)). Next, the utilities for the different regions are observed. The estimate for the first zip-code digit is interpreted with respect to the baseline, which is Vienna, for each alternative. It is therefore not surprising that most estimates are negative and that especially for the highest fixed-line alternative the estimates are highly negative. Internet users with a high utility from higher speeds are more present in Vienna. For mobile alternatives, this image is less clear, and for some regions also positive estimates can be found for some mobile alternatives. This shows that in certain regions, stationary mobile contracts are more popular than in Vienna. Even though a test was conducted via wireless LAN instead of LAN, the estimates for mobile and for the highest fixed-line alternative are negative, whereas the estimates for medium fixed-line speeds are positive with respect to the lowest fixed-line alternative.

In terms of the socio-economic characteristics, in smaller cities, higher speeds for both fixed-line and for mobile have highly negative estimates. Medium-sized cities show high utilities for medium speed fixed-line and for all mobile alternatives, especially for these with higher speeds. In contrast, larger cities show high utilities for high fixed-line speeds but less utility for mobile alternatives. Therefore, it appears that mobile alternatives are more demanded in rather rural areas, regardless of the region defined by the first zip-code digit. However, it should be borne in mind that the sample only includes areas where high fixed-line speeds are available.

Areas with a higher proportion of newborns also show higher utilities for high-speed fixed-line Internet. This could reflect the average household size and shows that with more people using the Internet, higher bandwidths are needed. A higher proportion of men also corresponds with higher utilities for higher speeds. This might reflect a generally higher interest in new technologies and more expensive devices among men compared to women. The level of education does not seem to play a major role. However, for lower education levels, higher speeds show much less utility. This group might be most budget-constrained and therefore does not demand higher bandwidths. The age distribution plays a large role. Younger people prefer higher speeds on both fixed-line and mobile alternatives. Older people prefer mobile alternatives. However, they do so less than younger people. Younger people use the Internet differently. They stream more movies and music and might use it also for work at home. Older people may appreciate the simplicity of using a mobile router that only needs to be plugged into the wall socket to be ready for use. It should be noted that all these estimates are not causal and are only indicative.

2.6 Conclusion

A novelty of this work is its data. I apply speed tests, with which Internet users measure their realized download and upload speed to compare it with their contractual bandwidth. This data is geo-referenced, such that I was able to merge it with coverage data and data on active providers. I showed that Internet in Austria is demanded by three different types of users. Only a very specific subset of Internet subscribers demands high-speed Internet. Thus, high-speed Internet is not of general interest. I add to the literature further findings regarding substitution patterns between technologies but within a bandwidth range. Whereas fixed-line and mobile are substitutes for Internet at basic speeds, at higher speeds, fixed-line and mobile are demanded by different types of users. Especially, high-speed fixed-line Internet (100 Mbps and more) is only demanded by a certain type of users. Hence, an increase in the availability of such speeds does not necessarily expand its usage as well. However, a high usage is necessary for a technology to have a broad (economic) impact and to be regarded as a general purpose technology. If the usage of high-speed Internet does not adjust to increased availability, there is no need for a government to foster its expansion, as demanded by the European Commission. It should be noted that only users with full fixed-line coverage were analyzed. However, in areas with less fixed-line coverage, the usage of high-speed Internet will not be high as even the users in this study show a high price-sensitivity.[32]

The results are robust in terms of the decisions of the researcher regarding data preparation and regression specifications. First, the whole sample and the reduced sample estimates are very similar. Adding more controls, changing the weights, reducing the sample with a random split, or by specific restrictions does not affect the result. The heterogeneity analysis shows that substitution varies between cities and between cities and smaller towns. Individual variation based on socio-economic characteristics is also shown.

The methodology applied in this chapter is also employed in the literature on market definition. Regulation in the telecommunications market is updated approximately every five years in response to innovation and changing market power. Since the TKG 2003, the regulation has followed a three-step approach. Regulation always begins with a thorough market definition before the market is analyzed in a second step, and if a company is identified as having significant market power, specific regulatory instruments are imposed on that company in order to ensure effective competition on this relevant market.[33]

The current regulatory status applies to the incumbent A1 Telekom as the owner of the copper network over which DSL is provided and defines the wholesale market for broadband Internet regardless of the bandwidth. The relevant market comprises DSL and fiber (FTTX[34]) as technologies. Other technologies such as cable (CATV), mobile telephony, or satellite, among others, are not part of the relevant market. Leased lines and unbundling are also not part of the market. There is no regional separation, i.e., the relevant market comprises the whole of Austria.[35]

As far as regulation or a thorough market definition is concerned, this chapter naturally provides only a first insight. Nevertheless, it has been shown that there are different types of users for higher speeds. This indicates that regulation should also consider basic speed as a different market from high-speed Internet. However, regulators might also consider including mobile Internet in the same market as basic speeds, as has been shown for the type of users who do not has high utilities for high-speed Internet. Finally, as only areas with full fixed-line coverage have been considered in this chapter, regulators could consider regional regulation.

In the coming years, 5G mobile Internet will become very important and will influence the substitution between fixed-line and mobile technologies. This is likely to have a particular impact on technological substitution between higher speeds. For this

[32] If usage were high, providers would have provided high speeds.
[33] Telecommunications law by the Austrian Regulatory Authority for Broadcasting and Telecommunications (RTR): https://www.rtr.at/en/tk/Marktanalyse
[34] FTTX stands for FFTH, FFTC and FTTB collectively
[35] https://www.rtr.at/en/tk/Breitband

substitution, it has been shown here that it is not yet very strong. Although this chapter analyzes data prior to the introduction of 5G, the results still serve as a baseline for future changes. Since 5G may not directly dominate the stationary mobile Internet, this chapter is also important for current technological substitution. Moreover, a lack of substitution between fixed-line and mobile technologies for high-speed Internet before the introduction of 5G does not necessarily mean that there is no substitution after the introduction of 5G. Lower quality of mobile high-speed Internet compared to fixed-line high-speed Internet might be the reason for the lack of substitution. With 5G, the quality of high-speed mobile Internet will increase, making substitution more likely.

Chapter 3

No Surprises, Please: Voting Costs and Electoral Turnout *

3.1 Introduction

Voting is the backbone of democracy. Yet, many democratic countries have experienced conspicuous declines in voter turnout in the past decades, prompting concerns about fading representativeness of electoral outcomes (Figure C.1.1). Early theories of electoral turnout have pointed out that the fact that people vote at all poses a paradox as the likelihood of a pivotal vote is dwarfed by any reasonable cost of casting a ballot [Downs, 1957]. To rationalize positive turnout rates, scholars have extended the trade-off to include factors such as a consumption value of voting, ethical considerations, and social rewards [Riker and Ordeshook, 1968; Feddersen, 2004; Ali and Lin, 2013; Funk, 2010]. Yet, the tension between voting as the essence of democracy and the insignificance of an individual vote remains unresolved, begging the question whether small increases in voting costs constitute a source of declining turnout rates.

We investigate the effect of shifting voting costs on electoral turnout by studying a seemingly innocuous shock: the relocation of polling places ("relocation shock"). We use a natural experiment in Munich, the third largest city in Germany, where voters may be reassigned to a new polling place for two reasons. First, for administrative reasons, the boundaries of some voting precincts are redrawn between election years so that a portion of the electorate is assigned to a different polling location. Second, polling venues, typically schools, must be newly recruited for every election. Although the electoral office seeks to retain previously operated polling venues, new requirements, construction work, and other circumstances might render some locations unavailable, producing variation in precincts' assigned polling place over time. We show that turnout is unrelated to reassignments

* This chapter is based on joint work with Jean-Victor Alipour.

in future elections and that sociodemographic differences between treated and untreated precincts are negligible once we partial out election-specific shocks and time-constant variation at the precinct level.

Our setting has the unique feature that the relocation shock is not the result of cost-cutting policies, but has purely administrative reasons. Moreover, employees of the Electoral Office are nonpartisan civil servants and have no direct incentives to manipulate the electoral process. Their primary aim in assigning voter addresses to polling places is to ensure high accessibility, primarily in terms of the number of eligible voters per precinct and the distance to the polling place. As a result, the extent to which reassignments in Munich lead to shorter or longer travel distances to the polling place is similar.

We expect polling place reassignments to impact the costs of voting in person, as opposed to voting by mail, via two distinct mechanisms: *i)* a "transportation effect" and *ii)* a "search effect" [Brady and McNulty, 2011; McNulty et al., 2009]. The transportation effect captures the increase or reduction in travel time resulting from the change in proximity to the polling location. The search effect refers to the cost of searching for the new polling place and going to an unfamiliar location (holding proximity constant). If the net increase in the costs of voting in person is sufficiently large, individuals will switch to mail-in voting or abstain from turning out. However, mail-in voting can only be requested until the Friday before the election Sunday, two days before the election. Hence, some voters, who would have switched from in-person to mail-in voting but who only notice the polling place reassignment after the deadline for requesting mail-in has passed, abstain from voting in the current election, but return to mail-in voting in the subsequent elections. These voters will be referred to as inattentive voters.

As a key novelty of our study, we evaluate the persistence of a relocation shock. Since reassignments typically produce lasting changes to voting costs (e.g., due to greater distance), we expect behavioral adjustments to carry over to subsequent elections. Persistence in voting patterns may also reflect habit formation, in the sense that today's act of voting increases the likelihood of voting in the future [Fujiwara et al., 2016]. Thus, to the extent that nonvoting is internalized into a new habit, this channel represents another driver of lasting changes in turnout. However, temporary drops in overall turnout can be explained by inattentive voters.

To empirically evaluate these predictions, we geo-locate the residential addresses of eligible voters and their designated polling place in the eight elections held between 2013 and 2020 in Munich. We identify changes in the assignment of addresses to polling places as well as the walking distance between each pair, before harmonizing precinct boundaries over our observation period. This leaves us with a panel of 618 precincts with time-constant boundaries for which we know the fraction of reassigned residential addresses, the average distance to the polling location, official election results (turnout at the polling place, turnout via mail, and overall participation), and time-varying sociodemographic characteristics.

We find that polling place reassignments engender a partial substitution of in-person for mail-in voting. On average, contemporaneous turnout declines by .46 (*SE* = .12) percentage points—or .74 percent, evaluated at the mean. Polling place voting declines by .75 (*SE* = .13) percentage points and is only partly compensated by an .29 (*SE* = .13) percentage points increase in mail-in voting. 80 percent of the overall decline is driven by the search effect and 20 percent by the transportation effect. Hence, to counterbalance the negative impact of the search effect on overall turnout, a polling place would have to move approximately 38 percent or .35 kilometers closer to the voter. Doubling the distance to a polling place reduces the turnout by 1.03 (*SE* = .20) percentage points, on average. The results are insensitive to including lag terms of reassignment and distance to the polling location, accounting for potential serial correlation in reassignments, and do not yield different results when distinguishing between relocations due to polling venue turnover and due to adjusted precinct boundaries.

To investigate the persistence of the relocation shock, we conduct an event study focusing on voting behavior around the first time a precinct is treated in our panel. We find no evidence of differential trends preceding the treatment, supporting the assumption that polling place reassignments occur randomly, conditional on precinct and election fixed effects. The event-study results further show that a relocation leads to a significant drop in overall turnout in the treatment year; however, mail-in votes completely offset the decline in polling place votes in the two subsequent elections. This pattern is consistent with the presence of inattentive voters and at odds with the hypothesis that voting is habit forming. Instead, the persistent substitution of in-person for mail-in voting is consistent with rational choice models of electoral turnout. The event-study results are robust to accounting for the staggered timing of the treatment using novel estimators by Roth and Sant'Anna [2021*a*], Callaway and Sant'Anna [2021], and Sun and Abraham [2020].

Although polling place relocations affect aggregate turnout only marginally and not persistently—therefore, unlikely to meaningfully influence the parliament's composition by changing the proportion of party votes—they can affect direct mandates, which are contested on a plurality rule in State and Federal Elections. In Munich, some direct mandates are at times highly contested. In the 2021 Federal Election, for instance, the constituency *Munich West/Center* was only won by 137 votes (.07 percentage points) by the conservative contender. In the 2018 State Election, a direct mandate (*Munich-Moosach*) was won by a margin of only 63 votes (.09 percentage points) by the *Green Party* candidate.

Heterogeneity analyses indicate the strongest effects for the youngest (aged 18 to 24 and 25 to 34) and the oldest (60 years and more) eligible voters. Moreover, the share of households with children reduces the effect (especially at the polling place). A strong substitution from polling place voting to mail-in voting is found with an increase in rents.

Our evaluation of the causal effects of polling place reassignments on turnout relates to two previous studies. Brady and McNulty [2011] exploit the consolidation of voting precincts in the 2003 Los Angeles gubernatorial recall election, which resulted in a reduction in the number of polling places. To account for non-random reassignment of individuals to polling locations, the authors employ statistical matching of registered voters in consolidated and unconsolidated precincts. They find a decrease in polling place turnout among reassigned voters, which was only partially offset by increased absentee voting. Using a similar strategy, McNulty et al. [2009] analyze the effect of reducing the number of polling places in the context of a 2006 school budget referendum in New York. The results show a lower turnout among voters who were reassigned to a new polling place. Both studies find that increased search costs and higher transportation costs jointly drive the decline in turnout. Causal identification in these settings rests on the assumption that matching on observables makes voters with new and unchanged polling locations comparable in all relevant characteristics. Our identification strategy instead hinges on the elimination of all residual variation that may confound our estimates by partialling out precinct and election fixed effects. Moreover, the administrative adjustment to precinct boundaries and turnover in polling venues induces as-good-as-random variation in the treatment. Our setting also allows us to examine the persistence of the treatment effects over subsequent elections and to shed light on habit formation in voting.

Several other studies also document the negative correlation between polling place reassignments or greater travel distance to polling places and electoral turnout. Amos et al. [2017] emphasize that reprecincting in the US is rarely a purely bureaucratic matter but prone to political influence. Against this backdrop, the authors find that the reduction of polling places for the 2014 General Election in Manatee County (FL) disproportionately affected minorities, younger voters and Democrats, and that turnout was significantly lower among reassigned voters. Exploiting individual-level variation for the 2001 mayoral election in the city of Atlanta, Haspel and Knotts [2005] show that citizens who have to travel longer distances are less likely to vote. The results are consistent with cross-sectional evidence from other contexts [Fauvelle-Aymar and François, 2018; Gibson et al., 2013; Bhatti, 2012; Dyck and Gimpel, 2005; Gimpel and Schuknecht, 2003]. However, these studies do not account for potential endogeneity, leaving room for biased estimates due to unobserved confounders or selection problems. One notable exception is Cantoni [2020], who studies the effect of distance to the polling location by exploiting geographic discontinuities at precinct borders in the US. Cantoni argues that citizens on opposite sides of precinct borders are identical on average, except for their assigned polling place. Comparing parcels of land and census blocks located near adjacent precincts, the author finds that a greater distance to the polling location significantly reduces the total number of votes. A key difference with our setting is that identification stems from cross-sectional variation. Instead, we estimate the effect of distance using *changes* in the distance to the polling location within voting precincts.

Our study also contributes to the empirical literature on habit formation in voting. Habitual voting implies that the act of voting itself increases the likelihood of voting in the future. Scholars have long been aware that differences in turnout tend to persist over time (see e.g. Plutzer, 2002; Green and Shachar, 2000; Brody and Sniderman, 1977) but causal evidence for habit formation remains ambiguous. Meredith [2009] demonstrates that voters who had just turned 18 at the time of the 2000 US general election (and thus had just become eligible to vote) are also more likely to cast their ballot in the subsequent election than their peers who fell just short of the age threshold. Gerber et al. [2003] provide evidence from a field experiment, suggesting that get-out-the-vote (GOTV) campaigns increase turnout in subsequent elections. By contrast, compulsory voting in Switzerland and Austria showed no persistent effects on turnout after its abolition [Bechtel et al., 2018; Gaebler et al., 2020]. Similarly, Potrafke and Roesel [2020] find that longer opening hours of polling places increased contemporaneous voter participation but did not affect turnout in subsequent elections when opening hours were no longer prolonged. Fujiwara et al. [2016] emphasize that, to appropriately identify habit formation, shocks that alter voting behavior in one election must not affect the costs or benefits of voting in the future. Specifically, the authors question whether experiencing a presidential campaign at a young age or receiving information and emotional messages from a GOTV campaign leaves a person's tastes, sense of civic duty, or cost of voting unaffected in a lasting way. Instead, they propose election-day rainfall as a transitory and unexpected shock to voting costs and show that the decrease in turnout induced by rainfall also reduces turnout in subsequent US presidential elections. In our setting, the relocation of a polling place, even if plausibly unexpected, is clearly correlated with future voting costs (e.g., if the new polling place is moved farther away). Thus, distinguishing whether a persistently lower turnout reflects habit formation or a lasting shift of voting costs may be impossible. However, we are able to test the necessary condition for habit formation, to wit: if voting is habit forming, then a decline in turnout due to the relocation shock must carry over to subsequent elections. We show that the necessary condition for habit formation can be rejected as (inattentive) voters who abstain from voting when subject to reassignment return to voting in the ensuing elections, thus recovering the drop in aggregate turnout.

The next section describes the institutional background and the experimental setting. Section 3.3 outlines the conceptual framework guiding our empirical analysis. We present a simple theoretical model based on a "calculus of voting" framework, in which citizens ground their voting decision in a rational evaluation of the possible alternatives. Section 3.4 outlines the empirical strategy. Section 3.5 presents our main results. Section 3.6 relates our findings to results of previous studies and discusses policy implications. Section 3.7 concludes.

3.2 Institutional Background

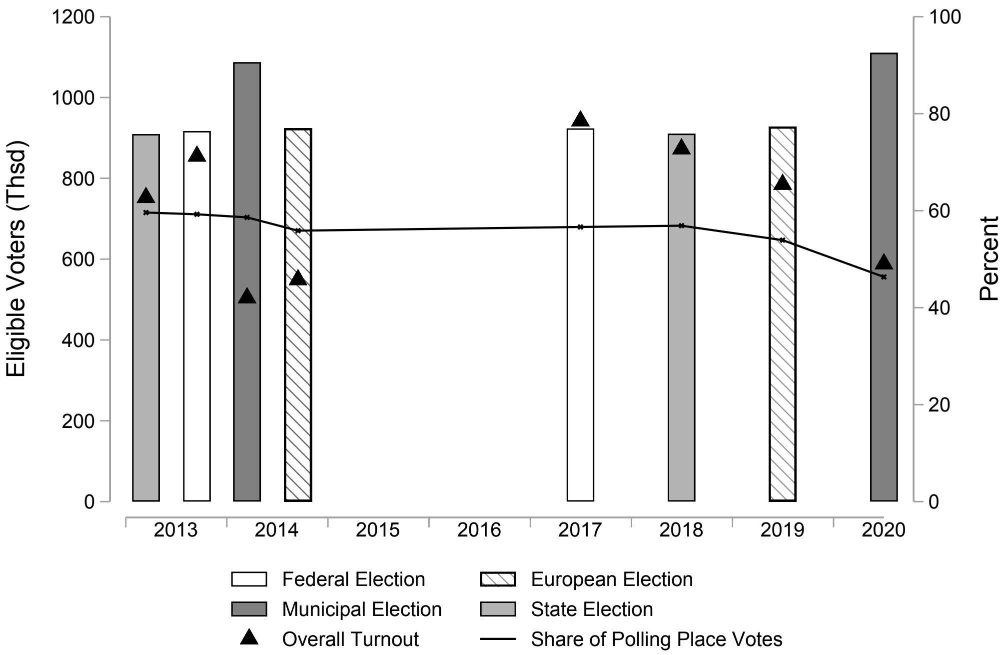
3.2.1 Elections in Munich

Our panel covers the outcomes of eight elections that were held in Munich between the years 2013 and 2020. These include elections to four legislative bodies that reflect the federal system of Germany: the *Bundestag* (German federal parliament), which constitutes the main body of the central government, the Bavarian *Landtag* (state parliament), the *Stadtrat* (Munich city council), which governs the city alongside the mayor, and the European Parliament, which effectively exercises some of the power of the federal government since Germany is a member of the European Union. All elections follow the principles of proportional representation but differ with respect to the electoral rules applied to achieve proportional representation. In Section C.3, we briefly describe the key features and differences of the electoral processes.

Eligible voters are automatically entered on the electoral roll without having to make a specific request. Then, every person on the roll receives an election notification via mail (no later than 21 days before the election) containing information about the election date and time, the location of the polling place, barrier-free access for the disabled or the elderly, and on the possibility of requesting a polling card (*Wahlschein*). There is no explicit information about any *changes* of the polling location—neither in the election documents nor in any separate notification. This contrasts with the US, where changes in precinct borders typically trigger the requirement to notify affected voters [Cantoni, 2020]. Eligible voters may cast a ballot in person at their assigned polling place on Election Day or request a polling card, which entitles them to vote by mail. A polling card must be requested no later than two days before the election. In principle, the polling card also entitles to vote at another polling place in the city (e.g., if the original polling place does not provide barrier-free access), but typically more than 98 percent of ballots cast using polling cards are votes by mail. And more than 90 percent of voters requesting a polling card actually end up casting a vote.

Figure 3.1 illustrates the timeline of the eight elections included in our panel. Two elections were held in both 2013 and 2014 (but not on the same day), and one election per year took place from 2017 to 2020. Depicted are the number of eligible voters on the electoral roll (vertical bars, left axis) as well as total turnout and the share of votes cast at the designated polling place (right axis). The number of eligible voters is distinctively higher in municipal elections, in which EU-foreigners with residence in Munich are also entitled to vote and added to the electoral roll.[1] Additionally, total turnout tends to increase over time when comparing the same election type while the share of polling place votes ranges between 50 and 60 percent of all ballots and shows a slight decline over time.[2]

Figure 3.1: Timeline and turnout of elections held between 2013 and 2020



Notes: The figure presents the number of eligible voters (vertical bars) as well as total turnout (triangles) and the share of polling place votes (solid line) for the eight elections included in our sample. The shading of the bars reflects the different election types. Between 2013 and 2020, two State Elections, two Federal Elections, two European Elections, and two Municipal Elections were held in Munich. The data are from the Munich Electoral Office (*Wahlamt*).

[1] For instance, in the 2020 Municipal Elections, 17.5 percent of eligible voters were foreign EU-citizens. Foreign EU-citizens who wish to vote in Munich instead of their country of origin in European Elections must lodge a registration request.

[2] With more than half of all votes cast by mail, the 2020 Municipal Election held during the Covid-19 pandemic marks an exception.

3.2.2 Precincts and Polling Places

Every election is organized and administered by the Munich Electoral Office (*Wahlamt*) according to a strict legal framework. Employees of the Electoral Office are nonpartisan civil servants and have no direct incentives to manipulate the electoral process. In every election, the electorate is geographically partitioned into several hundred voting precincts based on eligible voters' registered residential addresses.[3] Precincts constitute the smallest administrative unit in German elections and serve to enable a manageable election process and to facilitate the exercise of citizens' franchise, e.g., by preventing overcrowded polling places.

Figure 3.2 shows the electoral map for the 2018 State Election. The black boundaries identify the 618 precincts, the blue lines delineate the 25 city districts.[4] There is one polling place for every precinct (depicted by a black star). However, it is not uncommon for a single venue, typically a school, to host several polling places for neighboring precincts (four on average). The straight gray lines connect the residential addresses of eligible voters on the official electoral roll to the assigned polling places.

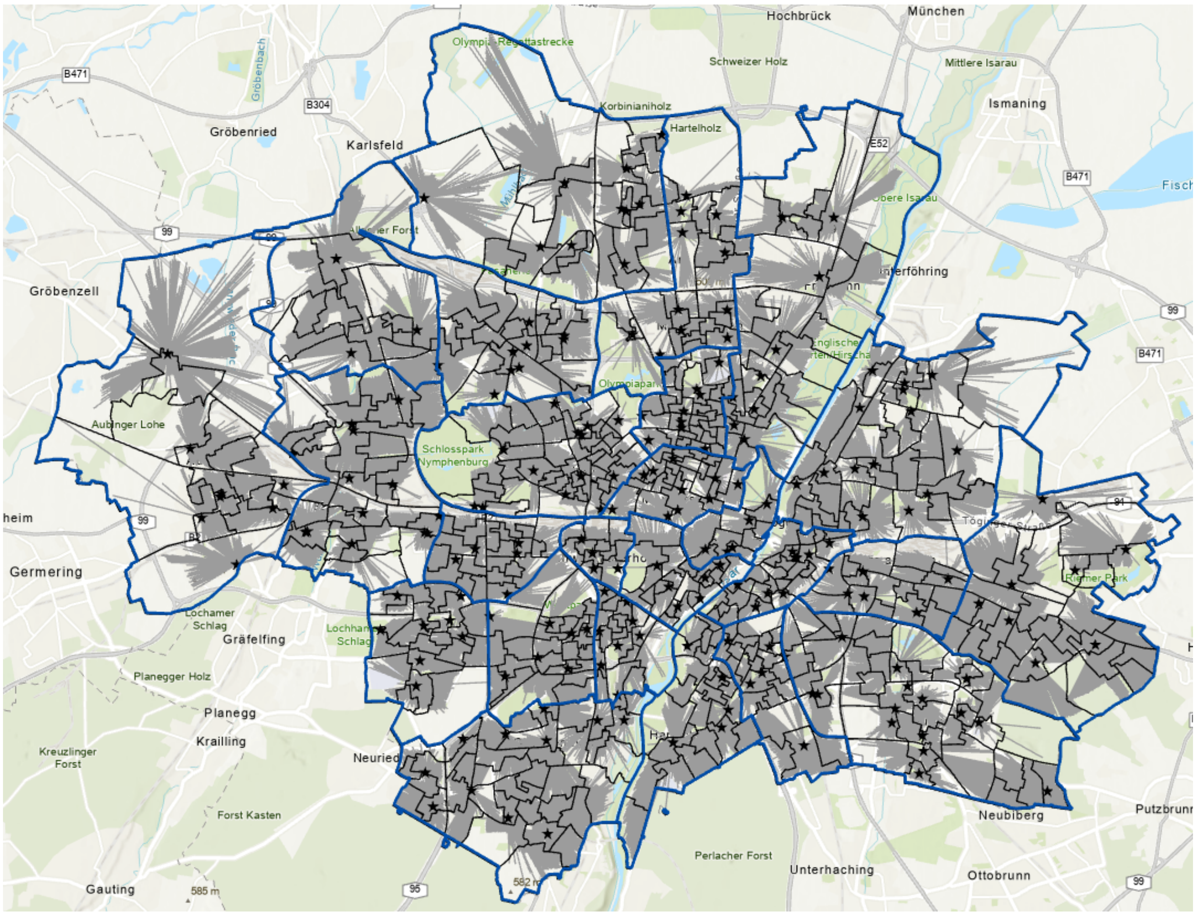
Redrawing Precinct Boundaries One source of variation in the assignment of voters to polling places results from adjustments to precinct boundaries. The law requires that voting precincts be drawn according to local conditions in a manner that participation in the election is "facilitated as much as possible for all eligible voters".[5] It further specifies that a precinct may not accommodate more than 2,500 eligible voters in any election. In practice, the city admits an average number of 1,500 eligible voters per precinct during the elections included in our panel (see Figure C.1.2 for a density plot of precinct sizes across all elections). Each election year, the electoral office evaluates whether a change in the number of eligible voters, population growth, or new housing units require adjustments in the number of precincts or to precinct boundaries to maintain a decent access to the polls. Overall, the total number of precincts remained at 702 in 2013 and 2014 before declining to 617 in 2017, due to the introduction of a new urban planning technology, which allows for a more granular spatial monitoring of the electorate and thus for a more precise delineation of precincts. This resulted in a comprehensive redivision of the city and a significant reduction in the variance

[3] Citizens are required by law to notify the relevant registration office of the city within two weeks of moving into a new flat. This also applies to citizens who move within a municipality.

[4] City districts have constant borders over time. Unlike precincts, districts are directly contested in some elections; for instance, adjacent districts cluster into 4 single-member constituencies in Federal Elections. In Municipal Elections, voters elect a local district committee (*Bezirksausschuss*).

[5] The legal requirements are outlined in the federal, state, and European election law, LWO §10, BWO §12, EUW §12, GLKrWO §13.

Figure 3.2: Electoral map of Munich for the 2018 State Election



Notes: The map delineates the boundaries of the 618 precincts (black lines) and the boundaries of the 25 city districts (blue lines) as of 2018. The locations of polling places are marked by a black star. Gray lines connect the residential addresses of eligible voters in the 2018 State Election to the assigned polling places.

of precinct sizes.[6] The number of precincts remained at 618 in 2018 and 2019 and increased again to 755 in 2020 to accommodate a larger number of eligible voters during municipal elections.

Recruitment of Polling Place Venues A second source of variation in the assignment of voters to polling places results from the recruitment of the venues hosting the polling places. Each election year, the electoral office prepares an information sheet that includes the delineation of the voting precincts and updated requirements for polling places. These requirements include, for instance, an adequate power supply and sufficient mobile network connection. Since 2017, the city has placed priority on selecting venues with barrier-free access for elderly and disabled people. Based on these guidelines, district inspectors (*Bezirksinspektoren)* are charged with the actual recruitment of potential venues, including

[6] Anecdotally, the Electoral Office addressed changes in the number of eligible voters by adjusting the number of poll workers at the polling locations before prior to 2017.

their localization, verification, and the coordination with third parties. Polling venues are typically public or municipal properties, usually schools (about 70 percent), but also retirement homes (15 percent), and ecclesiastical facilities (5 percent)—see Figure C.1.3 for an overview of venue types. While recruitment usually focuses on venues which have already been used in the past, new polling place requirements, competing events on Election Sundays, building closures, or ongoing construction work may leave certain locations unavailable.[7] Overall, we observe 293 distinct venues that hosted polling places in at least one election between 2013 and 2020. The number of operated venues is typically around 200 in any given election.

Despite the changes to precinct boundaries and polling venues, election officials have kept accessibility in terms of distance to polling places fairly constant over time. Figure C.1.4 depicts the median and interquartile range of the street (walking) distance between the addresses of eligible voters and their assigned polling places. The median distance remains at about 715 meters before slightly increasing to roughly 760 meters in 2017.

3.3 Conceptual Framework

To guide our empirical analysis, we present a simple theoretical model drawing on the "calculus of voting" framework, in which citizens base their voting decision on a rational evaluation of their options [Riker and Ordeshook, 1968; Downs, 1957]. For simplicity, we omit individual and election indices in the following. Denote $V \in \{N, P, M\}$ a citizen's voting decision in an election. She can either vote at the polling place (P), which involves cost c_p or vote by mail (M), which entails cost c_m. She may also abstain from voting (N), which generates neither costs nor benefits. Voting yields utility B, which may include the direct benefits from the act of voting itself, e.g., from fulfilling a civic duty, as well as the expected gain if the preferred party wins a greater number of votes in the election. The citizen abstains from voting if and only if her net benefit of voting is (weakly) negative, i.e., $B \leq c_p$ *and* $B \leq c_m$. In contrast, she votes by mail if and only if her net benefit of mail-in voting is positive and polling place voting is relatively more costly, i.e., $B/c_m > 1$ *and* $c_p/c_m > 1$. Similarly, she decides to vote at the polling place if $B/c_p > 1$ *and* $c_p/c_m < 1$. The left diagram in Figure 3.3 plots the benefit of voting relative to the cost of voting by mail against the relative costs of polling place voting. A citizen will vote if and only if her cost-benefit vector lies above the 45-degree line, where $B > c_p$, *or* above the horizontal unity line, where $B > c_m$. If additionally the vector lies to the right of the vertical unity line, where $c_p > c_m$, then she will vote by

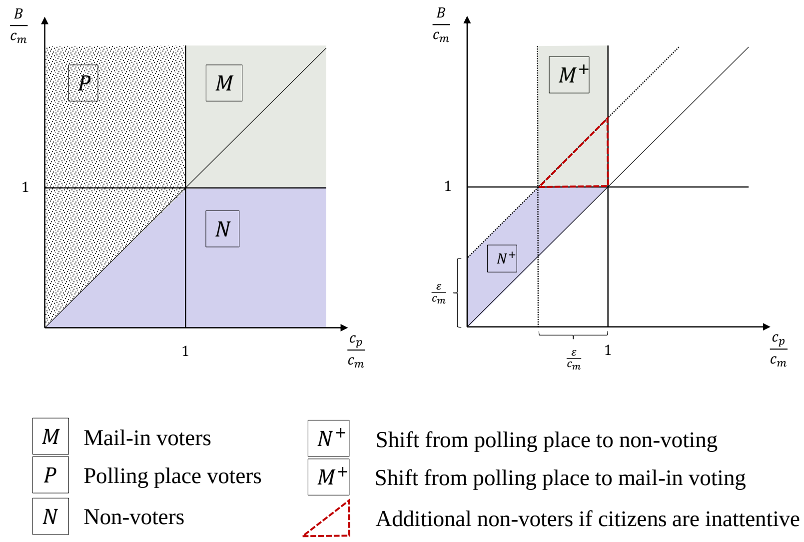
[7] There is no documentation of the reasons why venues become inactive. Anecdotal evidence suggests that, for instance, Munich's school construction program, which included investments of more than 3.8 billion Euros in the refurbishment of educational facilities starting from 2017, affected several polling venues. It is also possible that ecclesiastical institutions schedule religious events on Election Sundays.

mail (M), and chose polling place voting (P) otherwise. In the area below the intersection of the 45-degree line and the horizontal unity line, the net benefit of voting are always negative and the citizen will not vote (N). The shaded areas in the figure illustrate the voting decisions according to different cost-benefit configurations. If one imagines a distribution of Munich's population over the depicted plane, then historically roughly 38 percent of eligible voters lie somewhere in the nonvoting area (N), 33 percent are in polling place voter area (P) and the remaining 29 find themselves in the mail-in voter area (M).

Now, suppose that the electorate is subject to a relocation of the polling place. We anticipate that the reallocation of voters to polling places impacts the costs of voting at the polling place via two distinct mechanisms: *i)* a "transportation effect" and *ii)* a "search effect" [Brady and McNulty, 2011; McNulty et al., 2009]. The transportation effect captures the change in travel costs on Election Day resulting from the change in proximity to the polling place. In Munich, where polling places are usually located within walking distance, travel costs mainly correspond to the time to walk to the polling place. The search effect refers to the additional costs of searching for and learning about the new polling place (holding proximity constant). Search costs may also capture the psychological barrier to engage with the unfamiliar environment.

For illustrative purposes, suppose that the search and the transportation effect (or a combination of both) cause a (net) positive shock to the cost of voting at the polling place, $\varepsilon > 0$. Accordingly, c_p increases to $c'_p = c_p + \varepsilon$. The shock thus increases the *absolute* costs of voting at the polling place and decreases the *relative* costs of voting by mail. Graphically, this corresponds to an upward parallel shift of the diagonal line and a leftward shift of the vertical line, as illustrated in the right diagram of Figure 3.3. As a result, some voters will switch from polling place to mail-in voting (area labeled M^+). This is the case if the reduction in relative cost of voting by mail is large enough that $c'_p/c_m > 1$ and the net benefit of casting a mail-in ballot is positive, $B > c_m$. If the benefit of voting by mail is not sufficient to outweigh the costs, the citizen will switch to nonvoting if the cost shock is large enough to make polling place voting unattractive, i.e., $c'_p > B$ and $c_m > B$. The area labeled N^+ represents the shift from polling place to nonvoting. Therefore, the model predicts that the cost shock engenders a substitution effect between mail-in and polling place voting and a decline in overall turnout. A special circumstance arises, if voters are inattentive to polling place reassignments by not or only carelessly reading the election notification, which is mailed several weeks before Election Day and includes information about the polling location. By the time these voters learn of the polling place relocation, they may have missed the deadline for requesting mail-in ballots. Consequently, inattentiveness attenuates the shift from in-person to mail-in voting—as some voters will choose to go to the new polling place anyway—and amplify the shift towards nonvoting, as some voters who would have voted by mail abstain from casting a ballot. The additional portion of nonvoters is highlighted by the red triangle in the right diagram.

Figure 3.3: Effect of increased cost of polling place voting on voting behavior



Notes: The left diagram illustrates citizens' possible voting behavior—voting at a polling place (P), by mail (M), and not voting (N)—as a function of (individual) benefits (B) and costs of voting at the polling place (c_p) and via mail (c_m). The right diagram illustrates how a positive shock to the cost of voting at the polling place (ε) affects voting behavior. M^+ marks the additional portion of mail-in voters, N^+ marks the additional portion of nonvoters, and the red triangle highlights the additional portion of nonvoters in case citizens are inattentive to polling place reassignments.

To what extent do these adjustments carry over to subsequent elections? The theory suggests two mechanisms that may be at play. First, relocating polling places may alter the cost of voting at the polling place permanently. This is obvious, for instance, when transportation costs increase because a polling place is moved farther away. Similarly, search costs are likely to persist unless people familiarize themselves with the new location between two elections. Thus, the relative cost reduction of mail-in compared to in-person voting is likely to persist and thus to maintain the substitution effect. If the absolute cost increase for voting at the polling place is sufficiently high, then voters may entirely abstain from voting today and in the future. However, the initial election may be different from subsequent ones due to inattentive voters. Some inattentive voters will initially abstain from voting or cast their ballot at the new polling location but revert to mail-in voting in following elections. Consequently, a drop in aggregate turnout may be (partly) recovered and the substitution of in-person for mail-in voting reinforced over time.

A second mechanism that could drive persistent changes in voting behavior is habit formation. Habit formation means that the act of voting itself affects the probability of voting in the future—holding voter traits, such as the sense of civic duty or (individual) voting costs, constant (Fujiwara et al., 2016). Applied to our setting, habit formation would imply that a decline in overall turnout due to polling place relocations would carry over to subsequent elections even if the costs of voting were completely restored to pre-treatment conditions. As there are compelling reasons to anticipate that polling place relocation shocks are not transitory but permanently alter voting costs, it is not possible to separate the effects of habit formation from increased costs in our setting. Yet, we are able to test the necessary condition for habit formation, namely: if (non)voting is habit forming, then any initial decline in voter turnout must persist in the subsequent election(s). Empirically, the magnitudes of these effects depend on the distribution of the population over different cost-benefit vectors and the size of the reassignment shock(s).

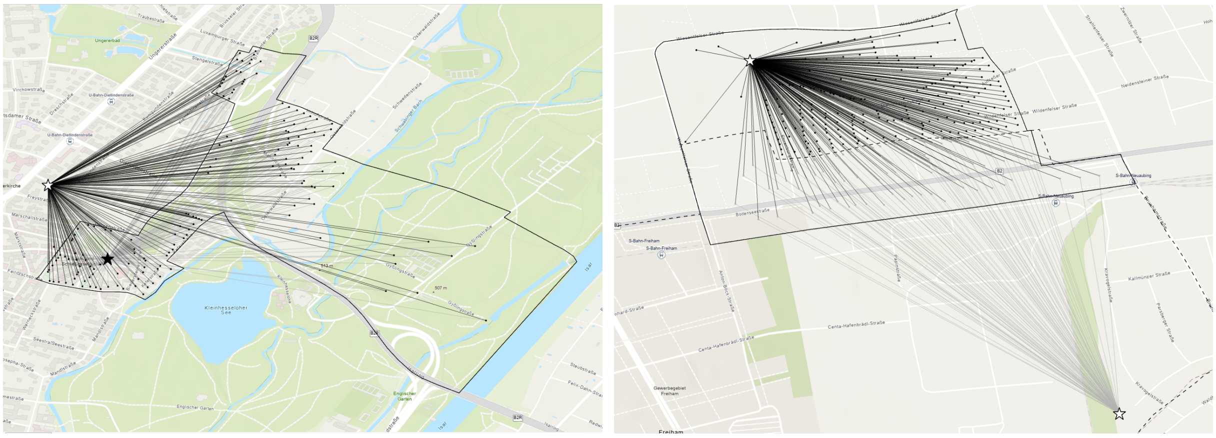
3.4 Empirical Strategy

3.4.1 Polling Place Reassignments

Figure 3.4 illustrates two instances of polling place reassignments which exemplify the two sources of reassignments in our setting. Gray lines connect residential addresses to the corresponding polling places in the 2017 Federal Election. The black lines connect the addresses to their polling place in the 2018 State Election. The solid black border lines delineate the respective precinct of interest. In Panel (1), all voters living in a northern Munich precinct experienced a relocation of their 2017 polling place as the hosting elementary school, marked by the black star, underwent a general renovation and became inoperable for the 2018 election. The new polling place was hosted by a vocational school (indicated by the white star) located six walking minutes (500 meters) from the old polling place. The example shows that recruiting a new polling venue—or the change in the activity status of a venue in general—typically means that all eligible voters living in the affected precinct have to vote at a different polling location than in the previous election. In this case, the average distance to the polling place increased for the affected electorate.

By contrast, Panel (2) illustrates an instance in which only a fraction of a precinct's electorate is treated due to the reconfiguration of its boundaries. The solid black lines mark the borders of the precinct of interest in 2018. The dashed lines delineate the boundaries of another precinct in 2017. Hence, citizens living at the intersection of these two shapes were reassigned from one precinct to the other, resulting in a change in the location of their assigned polling place. The fraction of voters living north of the dashed line were assigned to the same polling place in 2017 and in 2018 and are therefore considered untreated in our setting. Unlike in the preceding example, both polling places remained in operation in 2018 (white stars).

Figure 3.4: Illustration of treatment



(1) Recruitment of different polling venue (2) Redrawing precinct boundaries

Notes: The figure illustrates two instances of polling place reassignments between the 2017 Federal Election and the 2018 State Election. The residential addresses of eligible voters are connected by gray lines to their 2017 polling places and by black lines to their 2018 polling location. The precincts from 2018 are delimited by the solid black borders. In Panel (1), all residential addresses are reassigned due to the recruitment of a different polling venue: from the location marked by a black star to a new location marked by a white star. Panel (2) illustrates a reassignment due to an adjustment in precinct boundaries: the subset of residential addresses at the intersection of the 2018 precinct boundaries (solid black lines) and the 2017 boundaries (dashed black lines) was reassigned from the polling place located in the south to the polling place in the north on the map.

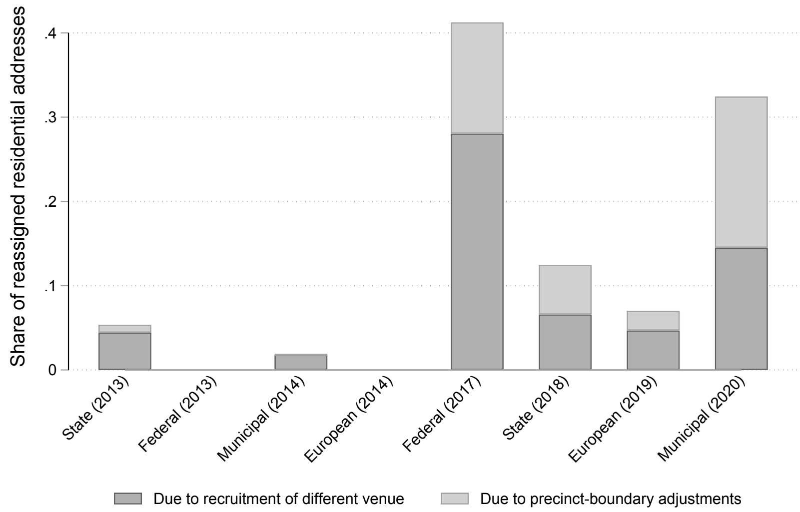
Figure 3.5 documents the fraction of residential addresses that were reassigned to a different polling place than in the previous election. There were no reassignments in the 2013 Federal Election and the 2014 European Election as other elections were held earlier in the same year. In 2017, more than 40 percent of addresses were assigned to a different polling place due to a major consolidation of precincts and updated requirements for polling places.[8]

Figure 3.6 reports the distribution of street distances between residential addresses and polling places (left panel), and the distribution of distance *changes* conditional on a polling place relocation across all elections (right panel). Negative values indicate that the new polling place is situated at a closer distance to an address (compared to the location in the

[8] Most eligible voters are not used to polling place reassignments. Figure C.1.5 depicts how often each eligible voter is affected by a change in the polling place. While most eligible voters are always assigned to the same polling place (more than 40 percent), about a quarter of all eligible voters are reassigned only once or twice, respectively. Less than every tenth eligible voter is reassigned three times or more often in the years between 2009 and 2020.

previous election), positive values correspond to a relocation farther away.[9] For 90 percent of residential addresses, the polling place is no further than 1.4 kilometers away, which roughly corresponds to a 17-minute walk (median: 735 meters). The median difference in distance to the polling place after a reassignment is 30 meters (mean: 55 meters) and the distribution has a skewness of .1. Hence, the distribution is fairly symmetrical, with polling places not systematically closer or farther away after reassignment.

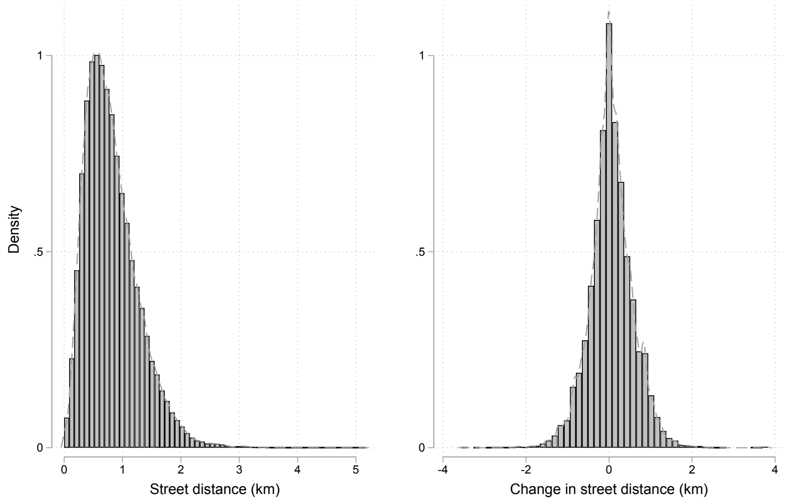
Figure 3.5: Share of addresses assigned to different polling place relative to previous election



Notes: The figure presents the share of addresses of residents on the official electoral rolls, which are assigned to a different polling place than in the previous election. Reassignment can be due to adjustment of precinct boundaries or due to recruitment of a different polling venue.

[9] Figure C.1.6 reports the distributions for straight-line distances. Notice that by definition straight-line distances are no greater than street distances.

Figure 3.6: Density of street distance and change in proximity to the polling place



Notes: The figures present density plots for the street distance between residential addresses of eligible voters and their assigned polling places (left plot, $N = 1,133,136$) and the *change* in distance conditional on assignment to a different polling place compared to the previous election (right plot, $N = 142,062$) for the eight elections between 2013 and 2020.

3.4.2 Data

All information on polling locations, residential addresses, and voter turnout (by mail and in-person at the polling place) comes from administrative sources including official electoral rolls and official election results. We geo-reference polling locations and residential addresses in the eight elections in our panel, as well as in the 2009 Federal Election, which serves as a reference to identify changes in polling place assignments relative to the 2013 state election (the first election in our panel). We identify 152,026 residential addresses from the 2018 electoral roll, of which we are able to match 143,278 to a unique precinct in every election (94.2 percent). 141,612 of these addresses were successfully geo-located (99.0 percent). We also calculate the street distance, defined as the shortest walking distance using the public road network, and the straight-line (Euclidean) distance between every pair of residential address and polling place in every election.[10]

In addition, we leverage time-varying administrative data on structural indicators at the precinct level.[11] These include information on the age structure of the electorate,

[10] We use the geodist STATA package [Picard, 2019] to compute straight-line distances and the osrmtime package [Huber and Rust, 2016], which make use of *Open Source Routing Machine (OSRM)* and of *OpenStreetMaps (OSM)* to find the shortest route (by foot or other means), to calculate street distances.

[11] Precinct-level structural indicators and turnout data are available for download from the city's election review website (*Wahlatlas*): https://www.muenchen.de/rathaus/Stadtinfos/Statistik/Wahlen.html [accessed August 8, 2021]. Official electoral rolls including residential and polling place addresses are provided by the Munich Election Office (*Wahlamt*) upon request.

average duration of residence in Munich, the marital status of residents and their citizenship (German, non-German EU, or non-EU citizenship). We also aggregate annual real estate rental price information compiled by the RWI Institute for Economic Research from square grids with a 1 kilometer length to the precinct level to capture socioeconomic differences among precincts.[12] Unfortunately, mail-in ballots are recorded at the level of administrative delineations that do not coincide with precinct borders. Thus, we are confined to relying on *requests* of polling cards as a proxy for mail-in votes in our empirical analysis. As noted above, about 90 percent of the requested cards are returned as ballots, and more than 98 percent of these are mail-in votes.

To obtain a panel of precincts suitable for estimation, we account for changes in precinct delineation over time. To this end, we harmonize precinct borders to the 2018 configuration, i.e., the share of polling place reassignments and the average distance to the polling place are computed assuming the 2018 (instead of the contemporaneous) precinct borders. Likewise, election-specific precinct characteristics, such as the age structure, the size of the electorate, or the number of votes cast, are converted to 2018 precinct borders using conversion keys provided by the Munich Statistical Office (*Statistisches Amt der Landeshauptstadt München*).[13] This leaves us with a panel of 618 precincts with constant borders, which we observe over eight elections. Figure C.1.7 plots the distribution of treatment intensities, i.e., the share of reassigned addresses, over all precinct-election observations in our panel in which a positive share of residential addresses are assigned to a different polling place. It becomes apparent that in the modal case, a precinct is fully treated, i.e., all its citizens are reassigned (39.8 percent of all instances). Table C.2.1 reports summary statistics of our precinct-level variables.

3.4.3 Main Specifications

We estimate the contemporaneous search and transportation effect by relating turnout to polling place reassignments and changes in average walking distance in the following model:

$$\begin{aligned} turnout^{s}_{pe(t)} = {} & \gamma_1 reassigned_{pe(t)} + \gamma_2 distance_{pe(t)} \\ & + \gamma_3 reassigned_{pe(t-1)} + \gamma_4 distance_{pe(t-1)} \\ & + \mathbf{X}'_{pe(t)}\lambda + \alpha_p + \alpha_{e(t)} + \varepsilon_{pe(t)}, \end{aligned} \tag{3.1}$$

[12] The RWI - Leibniz Institute for Economic Research (formerly Rheinisch-Westfälisches Institut für Wirtschaftsforschung) and its research data center compile granular real estate data obtained from the Internet platform *ImmobilienScout24* for research purposes.

[13] The variables are converted using population or electorate weights. A key assumption is that characteristics are evenly distributed within a precinct. For example, if a precinct is divided in two parts in 2018 (in terms of its electorate), it is assumed that voting behavior has not differed systematically between the two parts in the past.

where $turnout^s_{pe(t)}$ measures the percentage turnout in precinct p in election e held at date t, with $e(t) = 1, 2, ..., 8$, so that elections are ordered chronologically. The superscript s indicates whether turnout refers to turnout at the polling place, via mail, or in total (given as the sum of polling place and mail-in turnout). The variable *reassigned* denotes the share of residential addresses assigned to a different polling place compared to the previous election. Thus, the estimate for γ_1 captures the contemporaneous search effect. *distance* is the natural logarithm of the average street distance between residential addresses and the assigned polling place. By including precinct fixed effects, α_p, we identify the effect of *distance* from precinct-specific deviations from the mean, which are uniquely driven by polling place reassignments. Thus, the transportation effect is captured by γ_2. We also control for the lag terms of reassigned and distance to account for potential serial correlation in treatment that may bias our results. Intuitively, if a voter persistently changes her behavior after a polling place reassignment—for instance, by switching to mail-in voting—a second polling place relocation will not result in further behavioral adjustments. Thus, to the extent that voters are repeatedly reassigned during our observation period, we may underestimate behavioral adjustments to voting cost shocks. **X** is a vector of time-varying covariates at the precinct level: the precinct size (log of number of residents and the share of residents eligible to vote), the age structure of the electorate (share of eligible voters aged 18-24, 25-34, 35-44, 45-59), the share of EU-foreigners in the electorate, the share of native German residents, the share of non-native German residents, the share of single residents, the share of married residents, the average duration of residence (in years), the share of households with children, and the average quoted rent per square meter. We also include election fixed effects, $\alpha_{e(t)}$, to control for election-specific shocks, such as differences in voting propensity due to varying perceived stakes or the weather on Election Day. Precinct fixed effects further account for time-invariant precinct characteristics, such as its size (in terms of area), its remoteness, or its settlement structure (to the extent that it remains stable over our observation period).

The two main identifying assumptions for interpreting the estimation of contemporaneous treatment effects in Equation (3.1) as causal are that *(i)* polling place reassignments and changes in distance are uncorrelated with other unaccounted for factors that may affect turnout, and that *(ii)* polling place reassignments themselves are not driven by the expectation of changes in turnout. Although these assumptions are not directly testable, we provide a number of robustness checks, including a balancing exercise, a placebo test, and a pretrend analysis, suggesting that our results can be interpreted as causal.

We weight precinct-level observations with the number of eligible voters. This allows us to recover the conditional mean association between turnout and polling place reassignments at the individual level. In the baseline specifications, we cluster standard errors at the precinct level to account for the correlation of model errors over time. We also test the robustness of our results to alternative assumptions about the variance-covariance matrix in Section 3.5.3.

To investigate the persistence of behavioral changes due to polling place reassignment, we conduct an event study focusing on the window around the *first* time a precinct is treated in our sample. The event-study design allows us to examine to what extent voters may be permanently dissuaded from voting and whether there are lasting substitution effects between in-person and mail-in voting. Let E_p denote the election in which precinct p is treated for the first time (the event). We regress turnout on election dummies $D^k_{pe(t)}$ relative to the event E_p, control variables, as well as precinct and election fixed effects $(\delta_p, \delta_{e(t)})$:

$$turnout^s_{pe(t)} = \sum_{k=-K}^{-2} \mu_k^{lead} D^k_{pe(t)} + \sum_{k=0}^{L} \mu_k^{lag} D^k_{pe(t)} + \mathbf{X}'_{pe(t)}\phi + \delta_p + \delta_{e(t)} + \upsilon_{pe(t)}, \tag{3.2}$$

with the event-study dummies $D^k_{pe(t)} = \mathbb{1}\{e(t) - E_p = k\}$ and $e(t) = 1,2,...,8$. In our baseline estimates, E_p corresponds to the first election in which the *entire* electorate in a precinct is affected by a polling place reassignment. In the baseline, we also trim precinct time series from the point at which a second relocation occurs to ensure that we capture the impact of an individual reassignment rather than a series of changes. We test our results for robustness to alternative specifications in the subsequent section.

As a number of recent contributions have pointed out, two-way fixed effect (TWFE) event study (or difference-in-difference) approaches, similar to the specification in Equation (3.2), may still yield biased estimates when treatment effects vary over time (see e.g., Athey and Imbens, 2021; de Chaisemartin and D'Haultfœuille, 2020; Borusyak et al., 2021; Goodman-Bacon, 2021; Sun and Abraham, 2020). The main reason for this is that the TWFE estimator uses already-treated precincts as a control group for newly-treated precincts, thereby violating the parallel trend assumption in the presence of treatment effect dynamics. To account for this threat to identification, we also perform alternative approaches proposed by Callaway and Sant'Anna [2021], Roth and Sant'Anna [2021*a*], and Sun and Abraham [2020]. For instance, Callaway and Sant'Anna [2021] suggest a two-step estimation strategy by first estimating "group-time average treatment effects", where groups are defined according to the first time units (precincts) are treated, before aggregating the treatment effects by relative time using a propensity-score weighting method.

3.4.4 Balancing Test

Under our identifying assumption, precincts with and without polling place reassignments share similar determinants of voter participation, on average. Consequently, the correlation between observable precinct characteristics and reassignments should be negligible and statistically insignificant *conditional* on election and precinct fixed effects. We test this in Table 3.1. Each cell contains OLS estimates from a separate regression, with rows corresponding to precinct characteristics. The dependent variable in Column (1) is a dummy identifying precincts with a nonzero share of reassignments. The estimates are very small and not statistically significant, suggesting that the likelihood of *any* number of voters being reassigned to a different polling location is unrelated to observables. The dependent variable in Column (2) is the share of addresses assigned to a different polling place. Only one estimate appears marginally significant. Columns (3) and (4) distinguish between the reasons for reassignment, i.e., change in precinct boundaries or recruitment of a different polling venue, respectively. The estimates indicate no evidence that precinct characteristics are systematically related to the likelihood of reassignment for either reason. Finally, Column (5) regresses the log of average street distance on precinct characteristics. Out of seventeen estimates, only two estimates cross the threshold of statistical significance. Nonetheless, F-tests cannot reject the hypotheses that the estimates are jointly equal to zero in any column, indicating that the fixed effects perform well in eliminating the residual correlation between treatment and precinct characteristics. Therefore, the balancing test supports our identifying assumption.

Table 3.1: Balance test on precinct characteristics

	(1)	(2)	(3)	(4)	(5)
	dummy (reassigned >0)	share reassigned	share reassigned (precinct boundaries)	share reassigned (recruitment)	log street distance
residents (thsd)	-0.013	0.055	0.030	0.025	-0.004
	(0.045)	(0.035)	(0.028)	(0.031)	(0.031)
single residents (thsd)	0.016	0.108*	0.068	0.040	0.037
	(0.076)	(0.060)	(0.046)	(0.056)	(0.055)
married residents (thsd)	-0.103	0.070	0.002	0.067	-0.059
	(0.113)	(0.085)	(0.057)	(0.076)	(0.075)
native German residents (thsd)	-0.133	0.035	-0.033	0.067	-0.001
	(0.098)	(0.077)	(0.044)	(0.071)	(0.081)
non-native German residents (thsd)	-0.065	0.165	0.050	0.115	-0.185*
	(0.169)	(0.125)	(0.087)	(0.108)	(0.102)
foreign residents (thsd)	0.040	0.083	0.076	0.008	0.022
	(0.060)	(0.053)	(0.046)	(0.043)	(0.044)
residents eligible to vote (thsd)	-0.017	0.038	-0.039	0.077	-0.029
	(0.074)	(0.057)	(0.040)	(0.054)	(0.054)
eligible voters aged 18-24 (thsd)	-0.073	0.067	0.014	0.054	0.255
	(0.250)	(0.203)	(0.131)	(0.177)	(0.165)
eligible voters aged 25-34 (thsd)	0.105	0.149	-0.061	0.209*	0.142
	(0.138)	(0.113)	(0.067)	(0.108)	(0.115)
eligible voters aged 35-44 (thsd)	-0.075	0.130	-0.030	0.160	-0.052
	(0.173)	(0.139)	(0.085)	(0.129)	(0.123)
eligible voters aged 45-59 (thsd)	-0.253	0.156	-0.030	0.186	-0.114
	(0.175)	(0.144)	(0.103)	(0.127)	(0.122)
eligible voters aged 60+ (thsd)	-0.046	-0.026	0.006	-0.033	-0.167**
	(0.113)	(0.095)	(0.071)	(0.078)	(0.084)
Germans in the electorate (thsd)	-0.061	0.067	-0.020	0.088	-0.058
	(0.084)	(0.066)	(0.039)	(0.062)	(0.069)
EU-foreigners in the electorate (thsd)	0.001	0.063	-0.014	0.077	0.036
	(0.093)	(0.067)	(0.046)	(0.066)	(0.050)
households with children (%)	-0.001	0.003	0.003	0.001	0.004
	(0.004)	(0.004)	(0.002)	(0.003)	(0.003)
average quoted rent per sqm	0.000	0.002	-0.001	0.003	0.001
	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)
average duration of residence	-0.001	-0.001	0.000	-0.002	-0.003
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
F-test [p-value]	0.66 [0.85]	0.49 [0.96]	1.04 [0.42]	0.55 [0.93]	1.09 [0.36]
observations	4,944	4,944	4,944	4,944	4,944
precinct FE	✓	✓	✓	✓	✓
election FE	✓	✓	✓	✓	✓

Notes: Each cell in Columns (1) through (5) reports OLS estimates from a separate regression on precinct characteristics (in rows). All regressions include precinct and election fixed effects. The dependent variables are a dummy identifying precincts with a nonzero share of reassignments (Column (1)), the share of addresses assigned to a different polling place (Column (2)), the share of reassignments due to adjustment to precinct boundaries (Column (3)), the share of reassignments due to the recruitment of a different polling place (Column (4)), and the log of average street distance to the polling location (Column (5)). Regressions are weighted with the number of eligible voters. Standard errors are clustered at the precinct level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

3.5 Results

3.5.1 Search and Transportation Costs

Table 3.2 reports the estimation results of Equation (3.1). Panels A and B show the results for polling place turnout and turnout via mail, respectively. Panel C reports the net effect on total turnout. Column (1) includes only the share of reassigned residential addresses and the fixed effects. Column (2) adds precinct covariates. Column (3) further includes the lag term of reassignment. The estimate of *reassigned* in this column thus captures the average impact of a relocation on turnout. Column (4) reports the full specification including log street distance and the lag terms of reassignment and distance. Column (5) removes the lag terms to test the sensitivity of the estimates of contemporaneous reassignment and distance. Finally, we run a falsification test by relating contemporaneous turnout to *future* reassignments and distance to the polling place in addition to current and past values. It may that current and future reassignments share common causes that also determine voter participation. For instance, population growth may necessitate additional adjustments of precinct boundaries, and perhaps citizens in these precincts have a systematically different voting behavior. Thus, a relation between future reassignments and current turnout would suggest that these persistent confounders afflict our core estimates. The results of the placebo treatment are presented in Column (6).

In line with our expectations, the effect of reassignment on polling place turnout is negative and significant at the 1 % level in all specifications (Panel A). Controlling for lagged reassignments and covariates, the relocation of a polling place reduces in-person voting by .75 percentage points on average (Column (3)). Evaluated at the mean, this corresponds to a reduction of roughly 2.2 percent. Adding distance in Column (4) breaks down the reduction into the search effect and the transportation effect. Holding distance to the polling place and other factors constant, polling place relocation reduce in-person voting by .46 percentage points (1.4 percent at the mean), on average. The transportation effect also appears statistically significant: increasing the street distance to the polling place by 10 percent (equivalent to roughly 71 meters at the mean) reduces polling place turnout by .34 percentage points (equivalent to a one-percent decline at the mean). Thus, about 60 percent of reduction can be attributed to the search effect. The estimates also imply that a polling place would have to move approximately 13 percent closer to the voter to counterbalance the negative impact of the search effect for in-person voting, on average. The estimates of the contemporaneous search and transportation effect are insensitive to excluding the lag terms, suggesting that serial correlation in reassignments does not bias our results (Column (5)). The placebo treatment estimates reported in Column (6) further show that future polling place relocations do not affect current turnout in any panel. Thus, we find no evidence for unobserved persistent confounders.

The impact on mail-in turnout in Panel B mirrors the effect on polling place voting. On average, reassignments increase mail-in turnout by .29 percentage points (Column (3) of Panel B). However, only the transportation effect is statistically significant in the full specification (Column (4)). Increasing the distance to the polling place by 10 percent raises mail-in voting by 2.4 percentage points (equivalent to 8.4 percent at the mean). Thus, we find evidence for a substitution of in-person voting for mail-in voting after a polling place relocation. Yet, holding distance constant, the search cost effect only slightly compensates the drop in polling place turnout by increasing participation via mail. Similarly, a hypothetical relocation that results in a greater distance to the polling place leads to a larger decrease in polling place turnout than it increases mail-in turnout. This is in line with the theory predicting only a partial substitution as some voters will switch to nonvoting because the (individual) costs of voting by mail are higher than the perceived benefits or because inattentiveness regarding polling place relocations causes some voters to miss the deadline for requesting mail-in ballots.

The net effect of polling place reassignment on overall participation is indeed sizable and statistically significant. On average, turnout declines by .46 percentage points (Column (3) of Panel C). Both search and transportation costs drive the effect: holding distance constant, a polling place reassignment reduces overall turnout by .38 percentage points, which is equivalent to .6 percent at the mean, (Column (4) of Panel C). Thus, about 80 percent of the overall effect is due to the search effect. Increasing the distance to the polling place by 10 percent depresses voter turnout by approximately .1 percentage points, which corresponds to a .2 percent reduction at the mean. The estimates imply that the magnitude of the search cost effect on overall participation is equivalent to an increase in travel distance by 38 percent. Notice that the estimate of the contemporaneous search effect on overall turnout also reflects inattentiveness, i.e., votes that would have been cast by mail if individuals had noticed their polling place relocation in time. For instance, the estimates of the lag terms of *reassigned* suggest that there is some increase in mail-in voting stemming from relocations in the past. This could indicate that inattentive voters revert to mail-in voting in the election *after* the reassignment. The event-study analysis in the subsequent section allows to shed more light on this potential driver of declining turnout. In sum, the evidence so far shows pronounced transportation and search effects in the short-run, consistent with theoretical predictions and previous research [Brady and McNulty, 2011; McNulty et al., 2009].

Table 3.2: Search and transportation costs: baseline specification

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Turnout at the Polling Place						
reassigned	-0.76***	-0.74***	-0.75***	-0.46***	-0.44***	-0.60***
	(0.14)	(0.12)	(0.13)	(0.11)	(0.11)	(0.14)
reassigned, $t-1$			-0.62***	-0.44***		-0.58***
			(0.12)	(0.12)		(0.13)
log street distance				-3.44***	-3.48***	-3.60***
				(0.23)	(0.23)	(0.34)
log street distance, $t-1$				-0.02		0.24
				(0.19)		(0.24)
reassigned, $t+1$						0.04
						(0.13)
log street distance, $t+1$						-0.15
						(0.21)
R^2	0.96	0.96	0.96	0.97	0.97	0.96
Panel B: Turnout via Mail (requested)						
reassigned	0.26*	0.28**	0.29**	0.08	0.06	0.31**
	(0.15)	(0.13)	(0.13)	(0.12)	(0.12)	(0.15)
reassigned, $t-1$			0.52***	0.36***		0.52***
			(0.13)	(0.13)		(0.15)
log street distance				2.41***	2.59***	2.59***
				(0.24)	(0.23)	(0.33)
log street distance, $t-1$				0.35*		0.10
				(0.18)		(0.23)
reassigned, $t+1$						0.05
						(0.12)
log street distance, $t+1$						-0.06
						(0.17)
R^2	0.93	0.94	0.94	0.95	0.95	0.95
Panel C: Overall Turnout						
reassigned	-0.51***	-0.45***	-0.46***	-0.38***	-0.38***	-0.29**
	(0.16)	(0.12)	(0.12)	(0.12)	(0.11)	(0.14)
reassigned, $t-1$			-0.10	-0.08		-0.06
			(0.13)	(0.13)		(0.15)
log street distance				-1.03***	-0.90***	-1.00***
				(0.20)	(0.21)	(0.25)
log street distance, $t-1$				0.33*		0.33
				(0.19)		(0.21)
reassigned, $t+1$						0.09
						(0.12)
log street distance, $t+1$						-0.20
						(0.16)
R^2	0.98	0.99	0.99	0.99	0.99	0.99
observations	4,944	4,944	4,944	4,944	4,944	4,326
controls		✓	✓	✓	✓	✓

Notes: Dependent variables are voter turnout (0–100) at the polling place (Panel A), by mail (Panel B), and overall (Panel C). Mail-in voting is approximated by the number of requests for of polling cards (*Wahlscheine*). All specifications include election and precinct fixed effects. Precinct controls include the log of the number of residents, the share of residents eligible to vote, the share of eligible voters aged 18-24, 25-34, 35-44, 45-59, respectively, the share of EU-foreigners in the electorate, the share of native German residents, the share of non-native German resident, the share of single residents, the share of married residents, the average duration of residence (in years), the share of households with children, and the average quoted rent per square meter. Regressions are weighted with the number of eligible voters. Standard errors are clustered at the precinct level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

3.5.2 Pretrends and Persistence of the Relocation Shock

The key assumption of our empirical analysis maintains that polling place reassignments occur randomly conditional on precinct and election fixed effects. A central threat to validity are differential trends in turnout among precincts, depending on whether or not polling places relocations occurred. Hypothetically, the election office could systematically consolidate neighboring precincts that have historically shown greater shifts from in-person to mail-in voting to reduce the costs of operating polling places. In this case, our OLS estimate for the effect of reassignments may simply reflect a pre-existing trend rather than the substitution effect of a cost shock to voting at the polling place. The parallel-trend assumption is not directly testable. However, the event-study approach allows us to examine the existence of differential trends preceding the treatment.

Figure 3.7 plots the event-study results for turnout at the polling place, via mail, and overall. The event is defined as the first election a precinct is treated in our sample. In the baseline, we consider this to be the case when all residential addresses are reassigned to a new polling place. As emphasized above, we exclude all precinct-election observations beyond any second relocation so that we pick up the effects of only one instance of reassignments in every precinct. Of our 618 precincts, 278 are treated at some point. For most treated precincts the event occurs in the 2017 Federal Election (60 percent), 14 percent (13 percent) experience the reassignment shock in the 2020 Municipal Election (2018 State Election), and the remainder are treated in other elections.

Reassuringly, the results do not show evidence of pretrends in any of our outcome variables: all pre-event dummies are very small in magnitude and statistically indistinguishable from zero. By contrast, we find that polling place turnout falls by 1.15 (SE = .24) and mail-in turnout increases by .58 (SE = .24) percentage points immediately after a polling place relocation. This is in line with the substitution effect ensuing a reduction in relative costs of mail-in voting due to a polling place relocation. The bottom plot shows that the effect is not strong enough to completely offset the reduction in overall participation: total turnout declines on average by .57 (SE = .17) percentage points in the event election. Thus, compared to the earlier results estimated for the full sample, the event-study estimates for contemporaneous turnout are slightly more pronounced, suggesting a greater reduction in polling place turnout, a stronger substitution towards mail-in voting, and a slightly larger decline in aggregate turnout.

The estimates further show that the substitution of polling place voting for mail-in voting persists in the two subsequent elections. This is consistent with the theory predicting a persistent substitution effect resulting from a permanent change in the relative costs of voting. Interestingly, the net effect on total turnout appears to be statistically indistinguishable from zero in all elections following the event. While a portion of treated voters switch to nonvoting upon reassignment, the decline in turnout is already recovered in the following election. One interpretation is that the initial shock to the costs to polling place voting subsides over time. For instance, the search cost effect may diminish, as voters become familiar with the new polling place and uncertainty about its location and accessibility decreases. Another explanation is that the initial decline is largely driven by inattentive voters, who do not read the election notification (or do not read it carefully) and miss the deadline for mail-in voting before noticing that the polling place has been moved. Inattentive voters who would have switched to mail-in voting will either decide to vote at the new polling place anyway or forgo voting in the event election. But aware about the reassignment, these voters turn to mail-in voting in subsequent elections. The estimates support this interpretation, as total turnout recovers after the event and mail-in (polling place) voting exhibits a slight upward (downward) trend in the subsequent elections.

Finally, our results reject the hypothesis of habit formation in voting behavior. If (non)voting were actually habit forming, we would expect a lasting decline in turnout after the initial decline—even if the costs of voting were entirely restored to pre-event levels. Our estimates clearly do not support this pattern. However, in our setting, the decline in turnout—and consequently the test of the habit formation hypothesis—is likely to be disproportionately driven by inattentive voters. As this subset of the population is not necessarily representative of the general electorate, we cannot rule out with certainty that habit formation is still a relevant determinant of voting behavior for the average citizen.

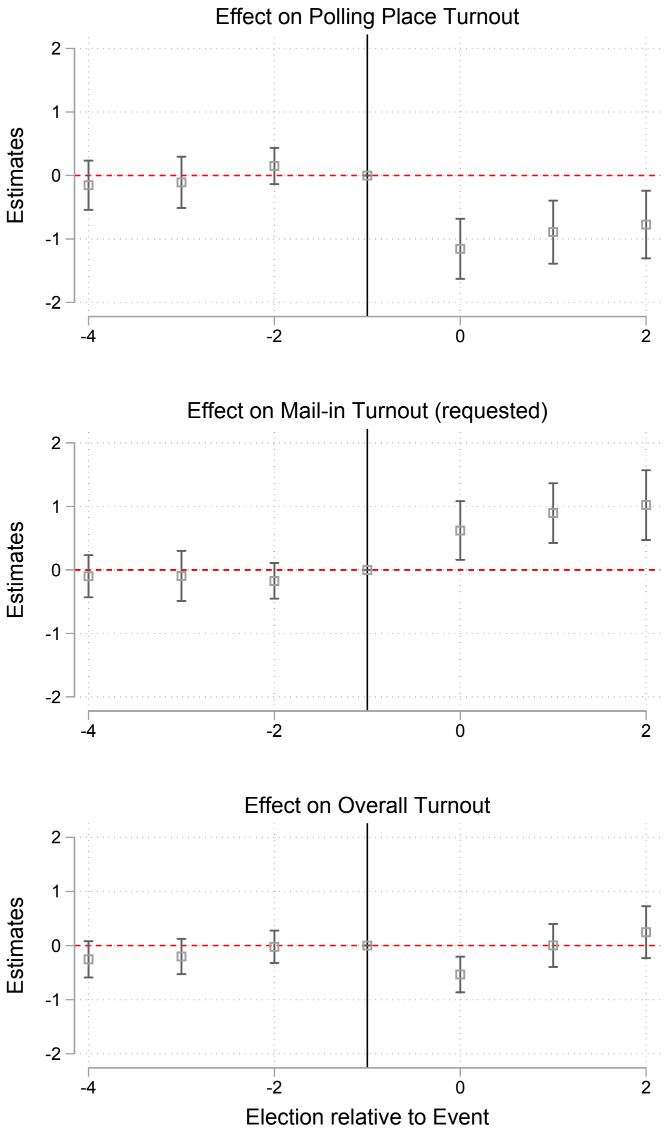
The full set of our event-study results are reported in Table 3.3. We first verify that our baseline estimates of the search and transportation effects (Equation (3.1)) on turnout hold for the subsample used in the event study (Column (1)). In Column (2), we present the event-study results corresponding to estimates reported in Figure 3.7. In Column (3), we additionally control for the log of street distance to absorb the transportation effect resulting from the polling place relocation. Since on average, a reassignment causes the distance to a citizen's polling place to increase, it is not surprising that post-event estimates now appear slightly closer to zero. Yet, the coefficients remain statistically significant, with the exception of the event-dummy in Panel (B), which captures the initial impact of a polling place relocation on mail-in votes. Thus, it appears that, holding transportation costs constant, a polling place relocation reduces polling place turnout but does not affect mail-in turnout. The shift towards mail-in voting only occurs in the election(s) following the event. This result further supports to the hypothesis of inattentive citizens, who would have switched to voting by mail, but do not notice the relocation until after the deadline for requesting mail-in ballots has passed. We also estimate the event study using the full sample instead of trimming the time series once a second treatment occurs. The estimates presented in Column (4) show that the results remain robust. In Column (5), we consider a different definition of the event. More specifically, the event corresponds to the first election in which at least 50 percent of a precinct is affected by polling place reassignments. The effect sizes are slightly attenuated but remain statistically significant. Finally, we estimate the model with a balanced sample. This reduces the number of observations by roughly 500 and the number of treated precincts from 278 to 114, of which 90 percent occur in the 2017 Federal Election and 10 percent in the 2018 State Election. The results reported in Column (6) confirm the previous estimations. Only the negative treatment effect on overall turnout in Panel (C) appears statistically insignificant, possibly due to the loss of statistical power due to the restricted sample.

In Figure C.1.8, we replicate the results of Table 3.3, Column (4) with several novel difference-in-differences estimators for staggered timing of the treatment.[14] Column (1) of Figure C.1.8 shows the results applying the estimator suggested by Roth and Sant'Anna [2021*a*], Column (2) reports the estimators proposed by Callaway and Sant'Anna [2021], and Column (3) the one by Sun and Abraham [2020]. In our setting, treatment accumulates in the 2017 Federal Election, which is right in the middle of our observation period. Hence, estimators give a high weight to this cohort and heterogeneity of treatment is only a minor concern.

Our setting has the unique feature that it is almost equally likely that a reassignment of a polling place increases or decreases the distance to the new polling place. In Figure C.1.9, we re-estimate the specification from Table 3.3, Column (2), and split the sample of the treated precincts into two subgroups: (i) precincts, where the new polling place is on average closer and (ii) precincts, where the new polling place is on average further away. The left panel shows that if the distance to the assigned polling place is lower after the relocation shock, search costs and reduced transportation costs, balance each other out. The treatment effect for this group can be set off by a reduction of the distance of about 250 meters (from 804 meters on average originally). In the treatment year nor in-person voting nor mail-in voting is affected. In subsequent elections, in-person turnout increases slightly as the search effect is transitory and the reduction in transportation costs is persistent. The right panel shows that if the distance to the assigned polling place is higher after the relocation shock, in-person voting declines by 2.01 percentage points and is only partially off set by mail-in voting (1.19 percentage points). So, an increase in distance of 337 meters on average decreases in-person turnout by approximately 2 percentage points (from 642 meters on average originally). The drop in in-person voting is persistent due to a permanent increase in transportation costs. The effect on the overall turnout is with −.82 percentage points higher than in the previous estimations. The off setting in the subsequent elections by an increase

[14] We used the staggered R-package by Roth and Sant'Anna [2021*b*].

Figure 3.7: Event-study illustration



Notes: The figure presents the event-study results from regressing turnout (at the polling place, via mail, and overall, respectively) on a set of election-date dummies around the event, which is defined as the first time the entire precinct is reassigned to a new polling place (Equation (3.2)). Regressions are weighted with the number of eligible voters. Confidence intervals reported at the 95% level. The full results of the underlying regressions appear in Column (2) of Table 3.3.

Table 3.3: Event study

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Turnout at the Polling Place						
reassigned	-0.63***					
	(0.15)					
log street distance	-3.45***		-3.43***			
	(0.31)		(0.26)			
$t-4$		-0.18	-0.22	-0.19	-0.13	-0.55*
		(0.20)	(0.19)	(0.20)	(0.16)	(0.28)
$t-3$		-0.09	-0.15	-0.11	-0.01	-0.49
		(0.21)	(0.21)	(0.21)	(0.18)	(0.30)
$t-2$		0.14	0.15	0.14	0.21	0.21
		(0.14)	(0.14)	(0.14)	(0.15)	(0.20)
t		-1.15***	-0.74***	-1.16***	-0.97***	-1.85***
		(0.24)	(0.22)	(0.24)	(0.17)	(0.40)
$t+1$		-0.91***	-0.67***	-0.86***	-0.79***	-1.70***
		(0.25)	(0.22)	(0.22)	(0.22)	(0.36)
$t+2$		-0.77***	-0.51**	-0.60***	-0.59**	-0.99***
		(0.27)	(0.24)	(0.22)	(0.26)	(0.35)
R^2	0.97	0.96	0.97	0.96	0.96	0.96
Panel B: Turnout via Mail (requested)						
reassigned	0.08					
	(0.16)					
log street distance	2.57***		2.64***			
	(0.29)		(0.27)			
	(0.24)					
$t-4$		-0.09	-0.06	-0.07	-0.02	0.20
		(0.17)	(0.17)	(0.17)	(0.14)	(0.21)
$t-3$		-0.11	-0.07	-0.09	-0.06	0.30
		(0.21)	(0.20)	(0.20)	(0.18)	(0.29)
$t-2$		-0.17	-0.18	-0.17	-0.08	-0.03
		(0.15)	(0.14)	(0.14)	(0.17)	(0.20)
t		0.58**	0.26	0.59**	0.44**	1.51***
		(0.24)	(0.23)	(0.24)	(0.18)	(0.35)
$t+1$		0.89***	0.71***	0.79***	0.76***	1.40***
		(0.24)	(0.22)	(0.21)	(0.22)	(0.34)
$t+2$		1.01***	0.81***	0.70***	0.91***	1.39***
		(0.28)	(0.26)	(0.24)	(0.27)	(0.36)
R^2	0.95	0.95	0.95	0.95	0.95	0.95
Panel C: Overall Turnout						
reassigned	-0.55***					
	(0.14)					
log street distance	-0.88***		-0.79***			
	(0.24)		(0.25)			
$t-4$		-0.27	-0.28	-0.26	-0.15	-0.35
		(0.17)	(0.17)	(0.17)	(0.15)	(0.24)
$t-3$		-0.20	-0.22	-0.20	-0.07	-0.19
		(0.17)	(0.17)	(0.17)	(0.16)	(0.26)
$t-2$		-0.03	-0.03	-0.03	0.13	0.18
		(0.16)	(0.16)	(0.16)	(0.15)	(0.26)
t		-0.57***	-0.48***	-0.57***	-0.53***	-0.34
		(0.17)	(0.17)	(0.17)	(0.14)	(0.27)
$t+1$		-0.02	0.04	-0.07	-0.03	-0.30
		(0.20)	(0.20)	(0.19)	(0.19)	(0.30)
$t+2$		0.24	0.30	0.10	0.32	0.39
		(0.24)	(0.24)	(0.21)	(0.25)	(0.30)
R^2	0.99	0.99	0.99	0.99	0.99	0.99
observations	4,350	4,350	4,350	4,500	4,090	3,518
event: 100% reassigned		✓	✓	✓		✓
full sample				✓		
event: >50% reassigned					✓	
balanced panel						✓

Notes: Dependent variables are voter turnout (0–100) at the polling place (Panel A), by mail (Panel B), and overall (Panel C). Mail-in voting is approximated by the number of requests for of polling cards (*Wahlscheine*). All specifications include election and precinct fixed effects and control for the following precinct covariates: the log of the number of residents, the share of residents eligible to vote, the share of eligible voters aged 18-24, 25-34, 35-44, 45-59, respectively, the share of EU-foreigners in the electorate, the share of native German residents, the share of non-native German resident, the share of single residents, the share of married residents, the average duration of residence (in years), the share of households with children, and the average quoted rent per square meter. The specification in Column (1) additionally controls for the lag of *reassigned* and the lag of *log street distance* (output suppressed). Regressions are weighted with the number of eligible voters. Standard errors are clustered at the precinct level and reported in parentheses. $^{***} p < 0.01, ^{**} p < 0.05, ^{*} p < 0.1$.

in mail-in voting, can again be explained by inattentiveness. Apparently, inattentiveness only plays a role for this subgroup. The difference of the effect on mail-in voting can be explained by different levels of mail-in voting before the treatment, which are probably rooted in different average distances to the polling place before the treatment. In fact, in the group where the relocation shock decreases the distance to the polling place, the original mail-in voting turnout was 38.94 percent, while in the other group, mail-in turnout was more than 2 percentage points lower (36.34 percent).

3.5.3 Robustness of the Results

Reason for Reassignment One potential concern is that the different reasons for polling place reassignments yield systematically different behavioral responses. This would suggest that voters anticipate changes due to a reconfiguration of precinct boundaries and changes due to the recruitment of a different venue to varying degrees. It may also be that a part of the electorate is systematically more likely to experiencing one type of reassignment, casting doubt on the (quasi-)randomness of treatment. Moreover, voters living near precinct borders be more likely reassigned due to revisions of precinct boundaries. If these voters differ systematically with respect to other determinants of electoral turnout, this could in turn afflict our estimates of interest. To test whether the different causes of reassignments could be a source of concern, we re-estimate Equation (3.1) differentiating the reassignments by reason. The results are shown in Table 3.4. Column (1) reports the baseline results for comparison. The estimates in Column (2) show that the different reasons for polling place reassignment do not drive the effect of a reassignment unequally. The t-tests for equality of the estimates (p-values reported in square brackets) indicate that the estimates are not statistically different from each other with respect to all outcomes (Panels A, B and C). This supports the assumption that voters do not anticipate or react differently to polling place reassignments depending on the reason for the change.

Error Correlation within Election-Districts Another potential concern is that model errors are correlated within city districts. This may happen because adjustments to the boundaries of adjacent precincts are not made across but solely within a district. Moreover, it is not uncommon for the polling places of several precincts (within a district) to be located in the same building. In these cases, a change in venue activity status will affect multiple precincts simultaneously. To account for this, we re-estimate Equation (3.1), correcting standard errors for two-way clusters at the level of precincts (to account for error correlation over time) and at the level of districts in each election (to account for within-district-election correlation). Column (3) presents the estimates with two-way cluster-robust standard errors. The standard errors of our variables of interest increase slightly but their statistical significance remains unaffected. We also re-estimate our event-study specification with two-way cluster-robust standard errors, which does not reduce the statistical significance of the estimates compared to the baseline specification.

Accounting for Constituencies Unlike precincts, city districts are directly contested in some elections. In state and federal elections, for instance, the 25 districts are combined into several single-member constituencies in which the parties' candidates compete for seats in the respective parliament. In municipal elections, citizens also elect a local district committee (in addition to the city council and the mayor). If there are systematic differences in voting incentives across districts—for instance, because citizens anticipate very close races in some constituencies—this may pose a threat to validity of our estimates of interest. Thus, we account for potential cross-district variation by estimating Equation (3.1) including a full set of district-election fixed effects. This ensures that comparisons are only made within district-election cells. The results in Column (4) show that our estimates of interest and their statistical significance are insensitive to the alternative specification.

Linear Time Trends We also test the robustness of our results to the inclusion of precinct-specific time trends. In the aggregate, we observe a slight shift towards mail-in voting over time, which was somewhat reinforced by the introduction of a simplified online application procedure for requesting polling cards in 2017. To account for possible differential trends among precincts, we re-estimate Equation (3.1), including a linear precinct-specific yearly trend. The results presented in Column (5) suggest that our results remain robust to this specification.

Excluding Election during Covid-19 Pandemic We also estimate the model excluding the 2020 Municipal Election, which was held at the onset of the Covid-19 pandemic in March. Uncertainty about contagion risks and limited hygiene concepts led to a historically low polling place turnout. As precincts may be hit by varying degrees by the crisis and voting behavior may not adapt uniformly in the city, we estimate the baseline equations without the 2020 election. Our results still hold, as shown in Column (6).

Alternative Distance Measures We also consider alternative measures of the transportation cost effect in Table C.2.2. In our baseline, we use the logarithmic street distance (walking distance) between residential addresses and their assigned polling place (replicated in Column (1)). Column (2) uses the linear street distance and Column (3) includes the linear street distance together with a quadratic term. The logarithmic and the linear street distance in Columns (1) and (2) show very similar estimates in all panels. Hence, the effects of an additional kilometer and a doubled distance are comparable. This indicates that the effect is not driven by precincts with a very high or very low average distance to the polls. The quadratic distance in Column (3) shows that an additional meter tends to reduce the effect size. In Columns (4) through (6), we perform the same exercise but replace the street distance with the average straight-line (Euclidean) distance between the residential addresses and the polling place. With exception of the first specification the estimates increase slightly as the straight-line distance is, by definition, shorter than the street distance. Importantly, the search cost effect (*reassigned*) remains robust to alternative measurement of the transportation effect across all specifications.

Table 3.4: Search and transportation costs: robustness

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Polling Place Turnout						
reassigned	-0.46***		-0.46***	-0.39***	-0.59***	-0.61***
	(0.11)		(0.12)	(0.12)	(0.15)	(0.14)
log street distance	-3.38***	-3.37***	-3.38***	-3.41***	-3.50***	-3.57***
	(0.23)	(0.23)	(0.23)	(0.22)	(0.30)	(0.30)
reassigned (precinct boundaries)		-0.54***				
		(0.18)				
reassigned (recruitment)		-0.42***				
		(0.14)				
R^2	0.97	0.97	0.97	0.97	0.98	0.96
T-test for equality of estimates		-0.56 [0.57]				
Panel B: Turnout via Mail (requested)						
reassigned	0.08		0.08	0.07	0.20	0.32**
	(0.12)		(0.17)	(0.12)	(0.16)	(0.15)
log street distance	2.37***	2.36***	2.37***	2.46***	2.55***	2.47***
	(0.24)	(0.24)	(0.25)	(0.22)	(0.30)	(0.31)
reassigned (precinct boundaries)		0.13				
		(0.20)				
reassigned (recruitment)		0.06				
		(0.15)				
R^2	0.95	0.95	0.95	0.96	0.96	0.95
T-test for equality of estimates		0.31 [0.76]				
Panel C: Overall Turnout						
reassigned	-0.38***		-0.38***	-0.31***	-0.40***	-0.29**
	(0.12)		(0.14)	(0.12)	(0.14)	(0.14)
log street distance	-1.01***	-1.00***	-1.01***	-0.95***	-0.96***	-1.10***
	(0.20)	(0.20)	(0.20)	(0.19)	(0.25)	(0.25)
reassigned (precinct boundaries)		-0.41**				
		(0.20)				
reassigned (recruitment)		-0.36***				
		(0.13)				
R^2	0.99	0.99	0.99	0.99	0.99	0.99
T-test for equality of estimates		-0.22 [0.82]				
observations	4,944	4,944	4,944	4,944	4,944	4,326
election FE	✓	✓	✓		✓	✓
precinct FE	✓	✓	✓	✓	✓	✓
2-way cluster			✓			
election-district FE				✓		
linear trend					✓	
excluding 2020 Election						✓

Notes: Dependent variables are voter turnout (0–100) at the polling place (Panel A), by mail (Panel B), and overall (Panel C). Mail-in voting is approximated by the number of requests for of polling cards (*Wahlscheine*). All specifications control for lag of *reassigned* and the lag of *log street distance* in addition to the following precinct covariates: the log of the number of residents, the share of residents eligible to vote, the share of eligible voters aged 18-24, 25-34, 35-44, 45-59, respectively, the share of EU-foreigners in the electorate, the share of native German residents, the share of non-native German resident, the share of single residents, the share of married residents, the average duration of residence (in years), the share of households with children, and the average quoted rent per square meter. Regressions are weighted with the number of eligible voters. Standard errors are clustered at the precinct level (except in Column (3)) and reported in parentheses. In Column (3), standard errors are corrected for two-way clusters at the level of precincts (to account for model error correlation over time) and at the level of districts in each election (to account for within-district-election correlation). $^{***} p < 0.01, ^{**} p < 0.05, ^{*} p < 0.1$.

3.5.4 Heterogeneity

Table 3.5 shows the heterogeneity across socio-economic characteristics. We examine the effect for age groups, by the share of households having children, and by the average square-meter price for renting. While Column (1) repeats the baseline results, Columns (2) through (6) show different age group effects, Column (7) shows the effect of children, Column (8) shows heterogeneity by wealth, and Column (9) combines all groups. For the age groups, heterogeneous effects are especially observed for the younger (aged 18 to 24 and 25 to 34) and the oldest (60 years and more) group. In line with our expectations, the treatment effect is much weaker for the youngest group. For them, it might be the first election, and therefore they do not perceive a change in the polling place, because they do not know the polling station of the last election. Moreover, they might be eager to start their democratic participation and therefore go to the polling place even if it involves relatively high costs. For the second youngest group, half of the effect is substituted by mail-in voting. Inattentive voters are thus much less present in this age group. For the oldest group, we observe no difference at the polling place. However, this group does not substitute by mail-in voting. This then leads to a much stronger overall absence from voting.

The effect at the polling place is smaller for a higher share of households with children. We hypothesize that this group might be more familiar with nearby schools that serve as polling places and that, in order to teach democratic values to their children, this group visits the polling place instead of voting by mail. The opposite holds for higher rents. As average square-meter prices for renting increase, the effect at the polling station decreases and is strongly substituted by mail-in voting. This finding can be rationalized by higher opportunity costs. Instead of searching for the new polling place, this group votes by mail, which is less time consuming. A lower overall effect is found for both groups.

All these variables might be correlated with each other and also with unobserved characteristics. For instance, the share of children is higher in certain age groups than in others and higher rents are correlated with education. However, when all effects are combined, only the overall estimate for the oldest age group remains statistically significant, while different effects are observed for the other groups at the polling place and by mail-in voting.

Table 3.5: Heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Turnout at the Polling Place									
reassigned	-0.46***	-0.46***	-0.43***	-0.44***	-0.45***	-0.45***	-0.46***	-0.30**	-0.41***
	(0.11)	(0.11)	(0.12)	(0.12)	(0.11)	(0.12)	(0.11)	(0.14)	(0.14)
log street distance	-3.44***	-3.44***	-3.45***	-3.45***	-3.44***	-3.44***	-3.42***	-3.45***	-3.41***
	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)
reassigned#eligible voters aged 18-24		0.31***							0.30***
		(0.09)							(0.09)
reassigned#eligible voters aged 25-34			-0.14						0.37*
			(0.10)						(0.21)
reassigned#eligible voters aged 35-44				-0.10					
				(0.11)					
reassigned#eligible voters aged 45-59					0.03				
					(0.12)				
reassigned#eligible voters aged 60+						0.01			0.37*
						(0.11)			(0.19)
reassigned#households with children							0.39***		0.51***
							(0.11)		(0.14)
reassigned#average quoted rent per sqm								-0.24**	-0.08
								(0.11)	(0.12)
R^2	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
Panel B: Turnout via Mail (requested)									
reassigned	0.08	0.08	0.01	0.04	0.08	0.00	0.08	-0.21	-0.12
	(0.12)	(0.12)	(0.12)	(0.13)	(0.12)	(0.12)	(0.13)	(0.15)	(0.16)
log street distance	2.41***	2.41***	2.44***	2.42***	2.41***	2.44***	2.39***	2.43***	2.41***
	(0.24)	(0.24)	(0.24)	(0.24)	(0.24)	(0.24)	(0.24)	(0.24)	(0.24)
reassigned#eligible voters aged 18-24		-0.18**							-0.24***
		(0.09)							(0.09)
reassigned#eligible voters aged 25-34			0.35***						-0.40*
			(0.12)						(0.23)
reassigned#eligible voters aged 35-44				0.23**					
				(0.10)					
reassigned#eligible voters aged 45-59					0.06				
					(0.12)				
reassigned#eligible voters aged 60+						-0.35***			-0.69***
						(0.10)			(0.18)
reassigned#households with children							-0.32***		-0.45***
							(0.12)		(0.15)
reassigned#average quoted rent per sqm								0.42***	0.20
								(0.12)	(0.14)
R^2	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
Panel C: Overall Turnout									
reassigned	-0.38***	-0.38***	-0.42***	-0.40***	-0.37***	-0.45***	-0.38***	-0.50***	-0.52***
	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.14)	(0.14)
log street distance	-1.03***	-1.03***	-1.01***	-1.03***	-1.03***	-1.01***	-1.03***	-1.02***	-1.00***
	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)	(0.19)	(0.20)	(0.20)	(0.20)
reassigned#eligible voters aged 18-24		0.13*							0.06
		(0.07)							(0.07)
reassigned#eligible voters aged 25-34			0.21**						-0.03
			(0.10)						(0.20)
reassigned#eligible voters aged 35-44				0.13					
				(0.09)					
reassigned#eligible voters aged 45-59					0.10				
					(0.10)				
reassigned#eligible voters aged 60+						-0.35***			-0.32**
						(0.09)			(0.16)
reassigned#households with children							0.07		0.06
							(0.11)		(0.15)
reassigned#average quoted rent per sqm								0.18*	0.12
								(0.11)	(0.13)
R^2	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
observations	4,944	4,944	4,944	4,944	4,944	4,944	4,944	4,944	4,944

Notes: Dependent variables are voter turnout (0–100) at the polling place (Panel A), by mail (Panel B), and overall (Panel C). Mail-in voting is approximated by the number of requests for of polling cards (*Wahlscheine*). The variables of the interaction term are standardized. All specifications control for lag of *reassigned* and the lag of *log street distance* in addition to the following precinct covariates: the log of the number of residents, the share of residents eligible to vote, the share of eligible voters aged 18-24, 25-34, 35-44, 45-59, respectively, the share of EU-foreigners in the electorate, the share of native German residents, the share of non-native German resident, the share of single residents, the share of married residents, the average duration of residence (in years), the share of households with children, and the average quoted rent per square meter. Regressions are weighted with the number of eligible voters. Standard errors are clustered at the precinct level and reported in parentheses. $^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.1$.

3.6 Discussion

3.6.1 Comparison with previous research

Previous findings on the effect of polling place reassignments on voting behavior provide an important benchmark for our results. There exist no other studies investigating how lasting these effects are. Thus, we focus on contemporaneous effects in the following. We estimate that, on average, reassignments result in a decline of in-person voting by .75 percentage points which is partially offset by an .29 percentage points increase in mail-in voting, leading to an overall decline of .46 percentage points–or .74 percent evaluated at the mean. Brady and McNulty [2011] find a similar partial substitution of mail-in voting for in-person voting following polling place reassignments in the 2003 Los Angeles gubernatorial recall election. However, the estimated effect magnitudes are significantly higher, with polling place turnout declining by 3.0 percentage points and overall turnout falling by 1.8 percentage points, or 3 percent relative to an overall turnout of 61.2 percent. Given that Brady and McNulty analyze a setting in which the number of polling places was significantly reduced (and thus distances to the polls increased), the greater decline in turnout is unsurprising. Still, we cannot rule out the possibility that our estimates suffer from attenuation bias due to imperfect measurement as we rely on the share of reassigned addresses instead of reassigned individuals. Accounting for changes in distance, Brady and McNulty find that about 60 percent of the reduction in polling place turnout is due to the search effect. This estimate is almost identical to our finding. Keeping in mind that our setting also features relocations that result in a closer distance to the polling place, this result indicates that the search effect is stronger overall in their setting. In fact, the authors find that the magnitude of the search effect is approximately equivalent to moving the polling place about one mile (1.6 kilometers) further away. By contrast, our estimates imply that the size of the search effect is comparable increasing the distance by about 100 meters, which is more than an order of magnitude smaller than Brady and McNulty's estimate. One explanation for this discrepancy is that voters use different modes of transportation to get to the polling locations, with Los Angeles voters primarily driving while Munich voters typically walking. Thus, the effects in terms of travel time are much closer. Another explanation is that the magnitude of the search effect itself is influenced by the distance to the polling place. If the new polling place is farther away, it is more likely located in an unfamiliar neighborhood. Consequently, the costs of getting acquainted with the new environment are higher. Since Brady and McNulty estimate the search effect in combination with *greater* distances to polling locations, the search effect is likely to be more pronounced than in Munich, where increases and decreases in distance are roughly equal.

McNulty et al. [2009] analyze a 2006 school budget referendum and estimate that the reducing the number of polling places caused a turnout decline of 7 percentage points Due to the negligible number of mail-in ballots, the authors focus on polling place voting only. Again, this substantial drop in turnout may be due to the fact that the travel distances to poling locations increased. At the same time, the results suggest that the effects of reassignments crucially depend on the context. The additional cost of voting in less salient or lower-stake elections, such as a school referendum, may have a greater impact on voting decisions than in higher-stake elections.

3.6.2 Policy implications

Election administrators' goal in Munich is to facilitate access to polling places as much as possible. Accessibility has been primarily understood in terms of precinct sizes, proximity to the polls, and (in more recent years) barrier-free access for individuals with physical impairments. Our results suggest that changing polling locations, even for the purpose of improving accessibility, constitutes an overlooked hurdle to voting. On average, reassigned voters are less likely to cast a ballot leading to a drop in aggregate turnout. We identified two main reasons for this result. First, the decision to vote appears only marginally affected by the change in the distance to the polls and primarily driven by the mere change in polling location (search effect). Second, inattentiveness to reassignments push individuals to temporarily abstain from turning out. Both channels could be mitigated by minimizing the number of reassignments by actively considering reassignments a threat to accessibility. Moreover, if voters choose not to vote because they missed the deadline to request mail-in ballots, an information treatment could alleviate the effect; for instance, by notifying citizens of polling place relocations separately from the usual election notification. In a correlational study in the context of the 2001 mayoral race in the city of Atlanta (US), Haspel and Knotts [2005] provide suggestive evidence that postcards sent to voters whose polling place had been moved increased the likelihood of casting a ballot by reminding citizens to vote.

3.7 Conclusion

Voting is the backbone of democracy. Yet, the likelihood of a pivotal vote is dwarfed by any positive cost of voting. Thus, even small and seemingly innocuous shocks to voting costs may affect aggregate electoral turnout. We exploit a natural experiment in the city of Munich (Germany) to evaluate how the relocation of polling places affects democratic participation. We find that moving a polling place has a disenfranchising effect, depressing precinct-level turnout by .46 percentage points, on average. The decline in polling place turnout by .75 percentage points is partially compensated by an increase in mail-in votes by

.29 percentage points. These effects can be explained by a combination of increased search costs due to unfamiliarity with the new polling place and transportation costs due to altered proximity to the polls. Further analyses show that the decline in overall turnout is transitory while the substitution of polling place for mail-in voting persists after the relocation of the polling place. This finding is consistent with the presence of inattentive voters, who only notice the polling place reassignment after the closing date for requesting mail-in ballots. Inattentive voters who would have switched to mail-in voting as their preferred choice either turn out at the new polling place anyway or abstain from voting. But with the awareness about the change, these voters revert to mail-in voting in ensuing elections, recovering the temporary drop in overall participation. Thus, rather than producing a (non)voting habit, reassignments provoke a persistent substitution of in-person for mail-in voting, consistent with rational choice models of electoral turnout.

Though the effect size of .46 seems rather small, the effect size if all voters are shifted 1 kilometer away from their polling place would correspond to a decline in voting of 1.5 percentage point and the decline at the polling place would be close to 5 percentage points. The reduction at the polling place can serve as an upper bound of the effect as mail-in voting might be less used. What was shown for instance for the group of the more than 60 year olds.

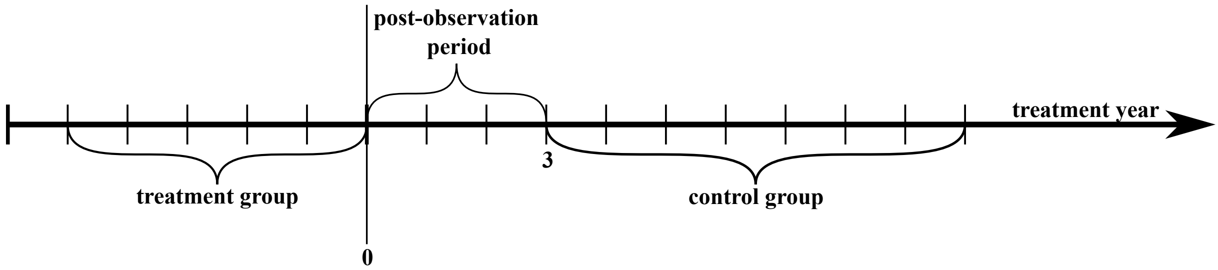
Appendices

Appendix A

Appendix to Chapter 1

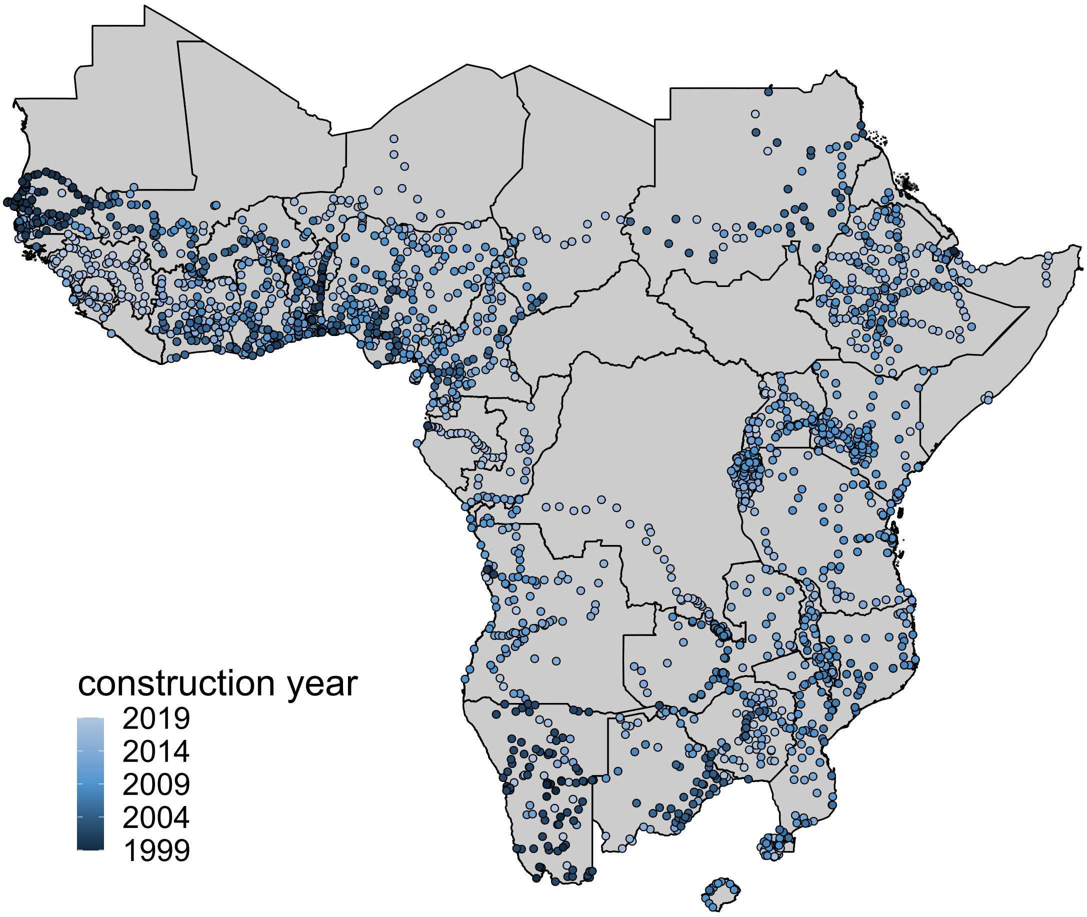
A.1 Additional Figures

Figure A.1: Timeline identification



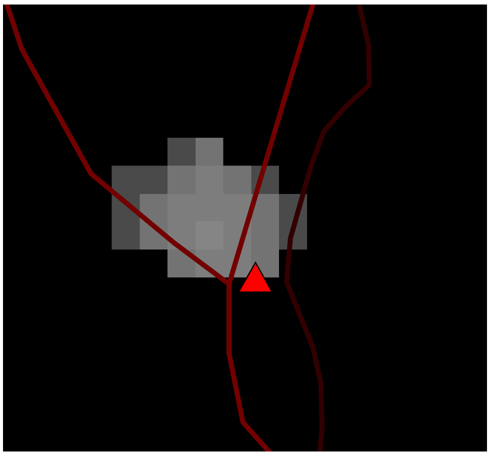
Notes: The figure gives an overview, of the timeline of our identification strategy. Treated are all towns having already an access point to the national infrastructure, when the Internet becomes available countrywide. As we investigate the effect of Internet availability for three years after the treatment, the control group contains all towns that get an access point in any year after theses three years.

Figure A.2: Access points and their construction years

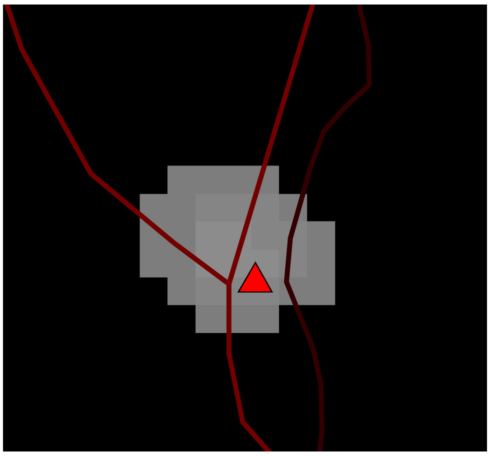


Notes: The figure depicts the location and construction date of all SSA access points. Brighter blue dots correspond to later constructed access points.

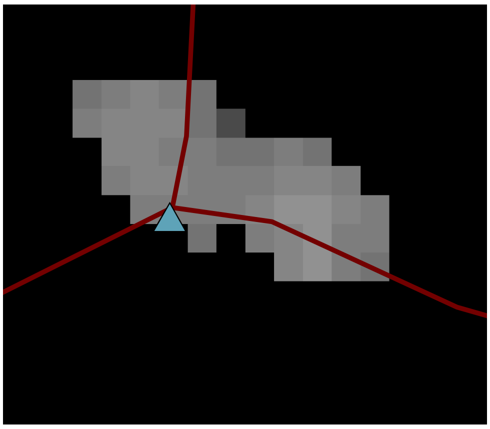
Figure A.3: Development of illuminated towns in Benin



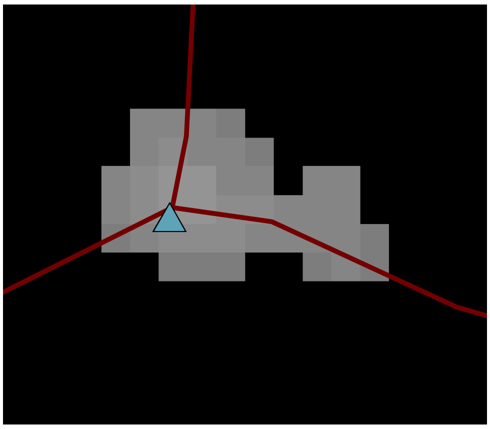
(1) Treated town in 2001



(2) Treated town in 2004



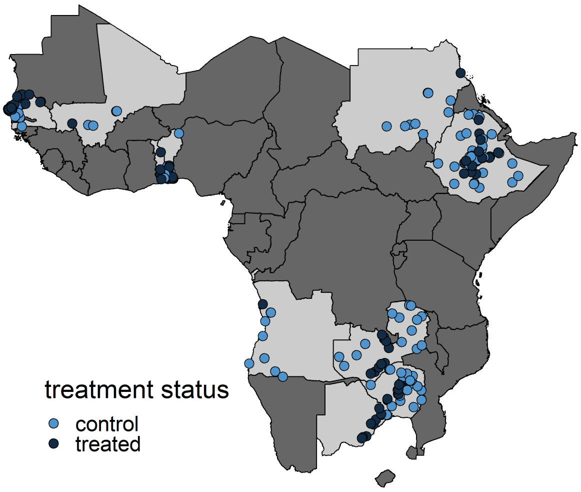
(3) Control group town in 2001



(4) Control group town in 2004

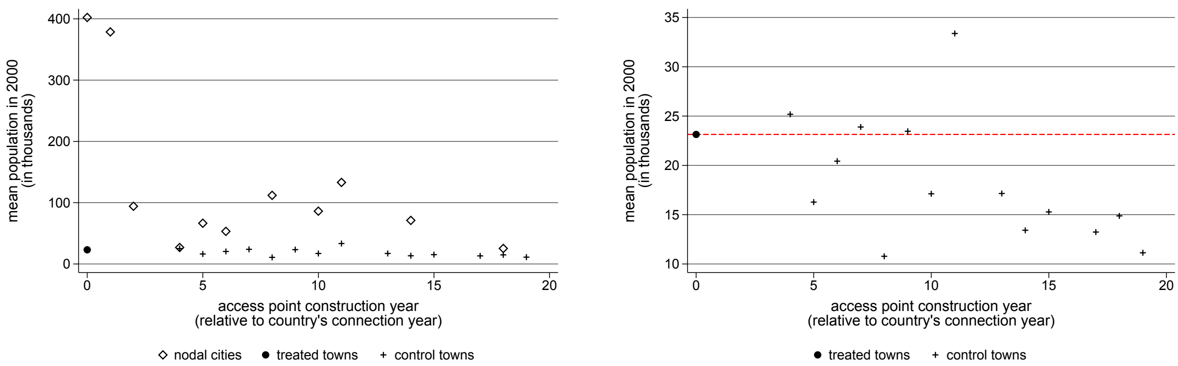
Notes: The panels show a treatment and control group town from Benin, with gray NTLs pixels from 2001 and 2004. Access points are marked with a triangle (red if constructed until 2001 and blue if constructed afterwards). The dark red line represents a major road connecting and the darker red line the railway. The black-to-white scale indicates light intensity, with brighter colors reflecting higher light intensities.

Figure A.4: Countries and towns location in the estimation sample



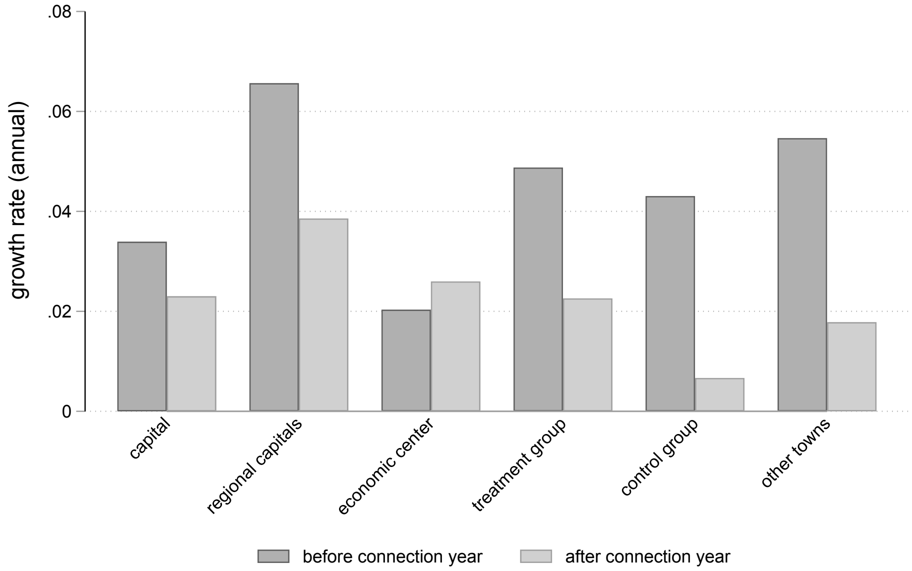
Notes: This figure depicts the countries in our analysis (brighter gray) and for each country the towns in the treatment and control group.

Figure A.5: Population size of connected cities and towns by year (relative to connection year)



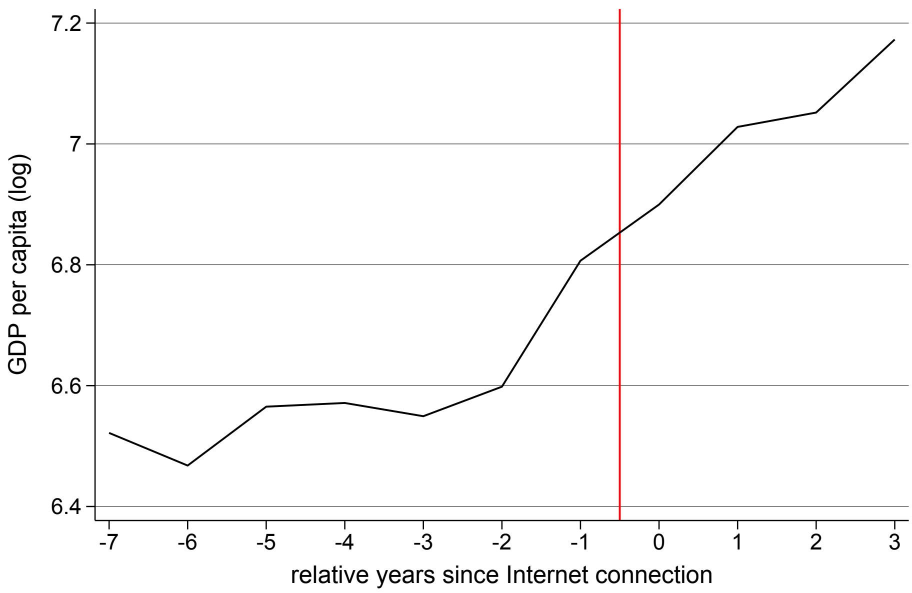
Notes: The figure depicts the average population size of connected cities and towns by year relative to the connection year. On the left, the black dot in the lower left corner represents the treated towns, while the control towns are represented by the plus symbol and the nodal cities by a diamond. For treated towns and nodal cities that were connected in earlier years than the arrival of an SMC are shown in year zero as well for clarity. On the right, the treatment and control group are shown in more detail without nodal cities.

Figure A.6: Comparison of city types (annual growth rates)



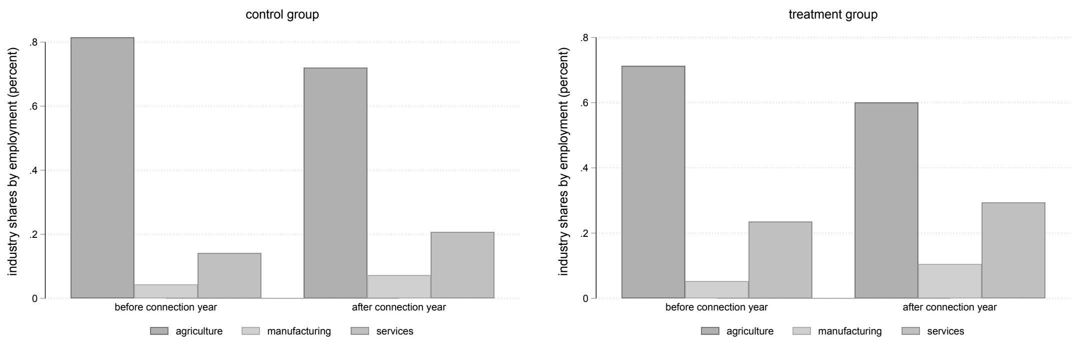
Notes: This figure shows the annual growth rates of the different city types measured by NTL intensity.

Figure A.7: Trend of GDP per capita growth



Notes: This figure shows the national GDP per capita growth before and after the Internet connection was established.

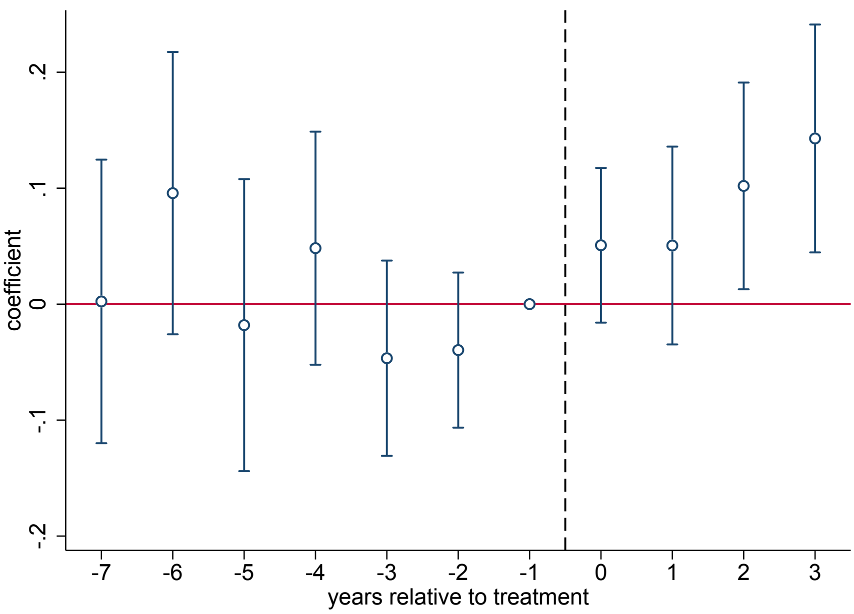
Figure A.8: Comparison of industry shares



(1) (2)

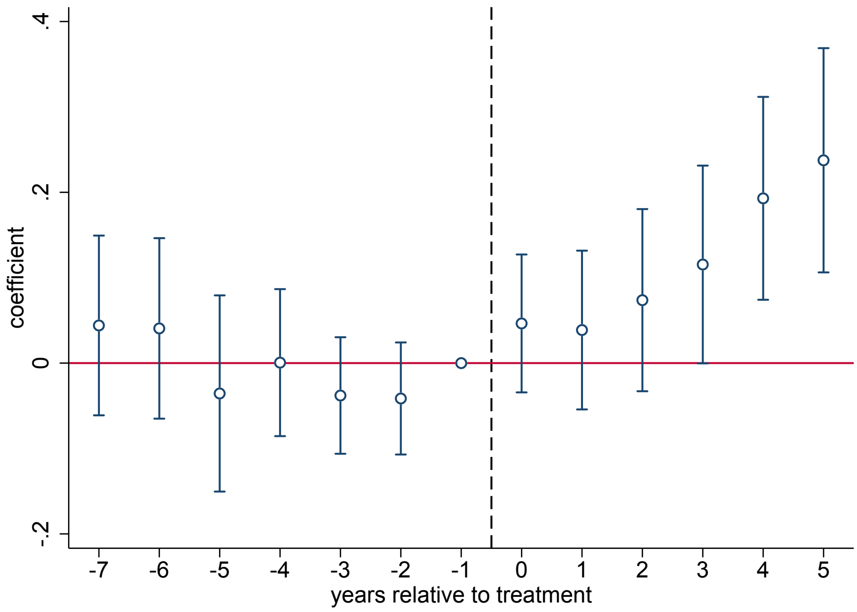
Notes: The figures depict the changes in the industry shares before and after the Internet connection for the treatment (1) and control group (2).

Figure A.9: Robustness event study (classical fixed effects)



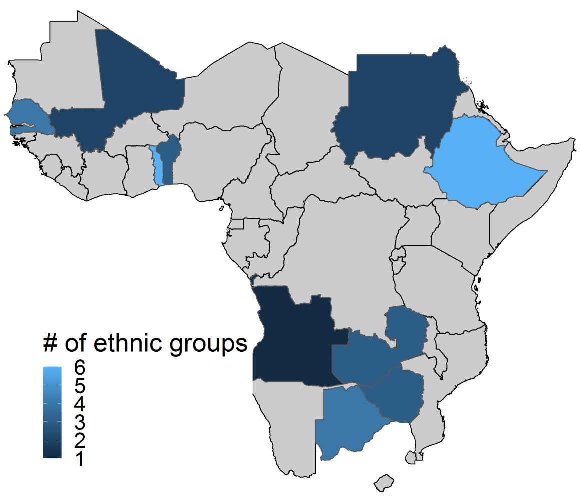
Notes: For external validity, more relaxed fixed effects are applied and therefore more countries are included. Robust standard errors clustered by town. Confidence intervals reported at the 95% level.

Figure A.10: Event-study coefficients with longer post-treatment period



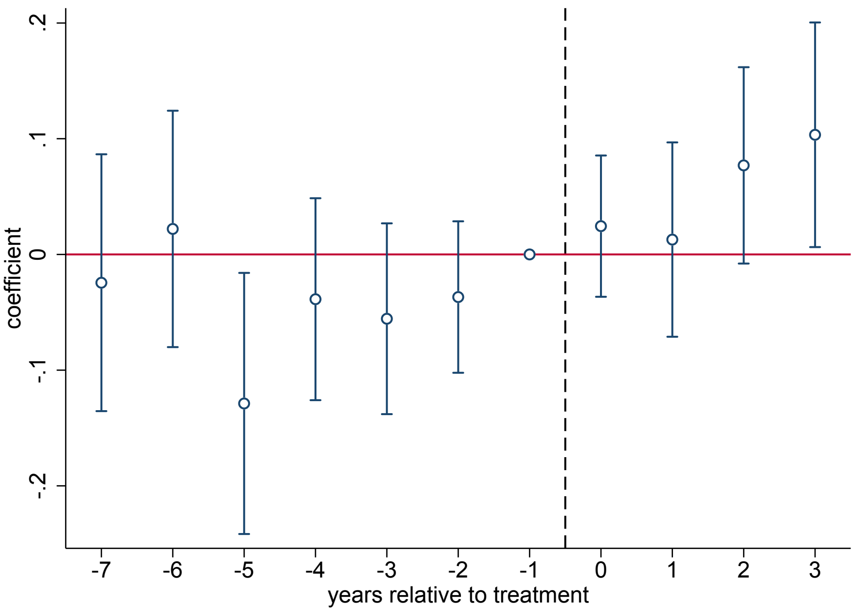
Notes: Coefficients for event study specification with five post-treatment years. Robust standard errors clustered by town. Confidence intervals reported at the 95% level.

Figure A.11: Ethnic groups



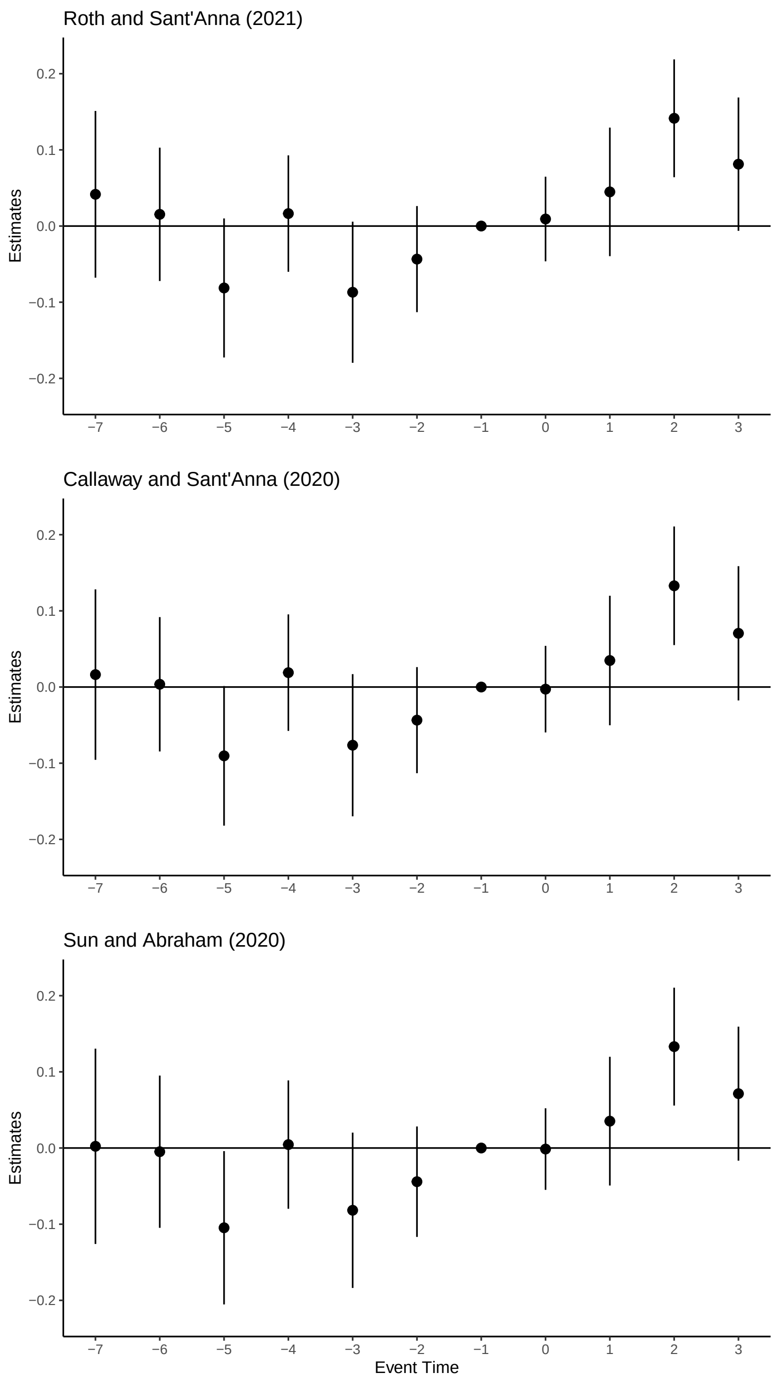
Notes: The figure shows for each SSA country in our analysis how many different ethnic groups were provided with at least one access point before the arrival of an SMC. Brighter blue colors indicate more different ethnic groups. Gray indicates countries not included in our analysis.

Figure A.12: Robustness event study (external validity)



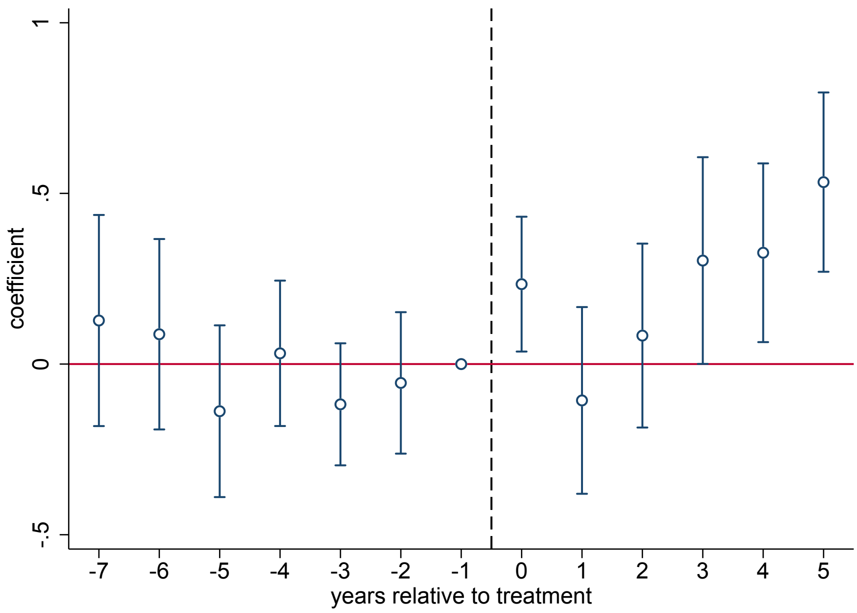
Notes: For external validity, less restrictive fixed effects are applied and therefore more countries are included. Robust standard errors clustered by town. Confidence intervals reported at the 95% level.

Figure A.13: Robustness event study (external validity)



Notes: The figure presents the event study results estimate with the suggested procedures by Roth and Sant'Anna [2021*a*], Callaway and Sant'Anna [2021], and Sun and Abraham [2020]. Control variables are not included. Robust standard errors clustered by town. Confidence intervals reported at the 95% level. Results are estimated with the staggered R-package by Roth and Sant'Anna [2021*b*].

Figure A.14: Heterogeneity event study (coastal countries)



Notes: Estimated on a sample including only coastal SSA countries. Robust standard errors clustered by town. Confidence intervals reported at the 95% level.

A.2 Additional Tables

Table A.1: Connection years

Country	Connection year	Connected by	SMC landing point	Upgrade year
Namibia	1999	Neighboring country		2012
Djibouti	1999	Sub-marine cable	Djibouti City	2009
Senegal	2000	Sub-marine cable	Dakar	2010
Angola	2001	Sub-marine cable	Sangano	2012
Benin	2001	Sub-marine cable	Cotonou	2012
Ghana	2001	Sub-marine cable	Accra	2010
Cameroon	2001	Sub-marine cable	Douala	2012
Gabon	2001	Sub-marine cable	Libreville	2012
Nigeria	2001	Sub-marine cable	Lagos	2010
Ivory Coast	2001	Sub-marine cable	Abidjan	2010
Sudan	2003	Sub-marine cable	Port Sudan	2010
Mali	2004	Neighboring country		2010
Botswana	2004	Neighboring country		2009
Zimbabwe	2004	Neighboring country		2011
Burkina Faso	2005	Neighboring country		2010
Togo	2005	Sub-marine cable	Lomé	2012
Gambia	2005	Sub-marine cable	Banjul	2012
Chad	2005	Neighboring country		2012
Central African Republic (CAR)	2005	Neighboring country		2012
Guinea-Bissau	2005	Sub-marine cable	Suro	2012
Mozambique	2006	Sub-marine cable	Maputo	2009
Lesotho	2006	Neighboring country		2010
Niger	2006	Neighboring country		2012
Malawi	2007	Neighboring country		2010
Ethiopia	2007	Neighboring country		2012
Zambia	2007	Neighboring country		2011
Swaziland	2008	Neighboring country		2009

Notes: The table reports the connection years of all SSA countries being connected before 2009. Source: *Submarine Cable Maps* and *Africa Bandwidth Maps*.

Table A.2: Summary statistics

VARIABLES	(1) mean	(2) sd	(3) min	(4) p25	(5) p50	(6) p75	(7) max	(8) N
population	20,581.39	17,933.61	0.00	8,501.50	16,019.00	30,114.00	82,602.00	220.00
distance to any regional capital	85.45	80.54	1.67	26.98	66.57	129.70	407.28	220.00
distance to the capital	231.65	203.73	1.67	75.81	170.73	355.42	987.20	220.00
distance to the coastline	426.58	307.37	0.00	154.25	427.69	632.76	1,175.48	220.00
distance to next river	56.84	56.65	0.00	15.16	43.99	86.89	411.27	220.00
distance to next port	195.34	272.67	8.23	28.40	74.31	177.52	1,207.12	75.00
distance to the road network	2.58	12.07	0.00	0.00	0.00	0.00	112.57	220.00
distance to the railroad network	57.26	96.44	0.00	0.00	3.80	82.08	440.13	220.00
distance to the electricity grid	13.44	40.58	0.00	0.00	0.00	3.80	350.51	220.00
number of lit pixels	43.35	33.26	1.00	24.00	35.00	53.00	288.00	220.00
summed light intensity	463.04	529.12	21.00	161.50	285.00	530.50	4,026.00	220.00
average light intensity	7.50	5.97	0.26	3.22	5.25	10.07	29.38	220.00
GSM coverage	0.62	0.47	0.00	0.00	1.00	1.00	1.00	220.00
distance to next AP in 2019	1.26	2.52	0.00	0.00	0.00	1.21	9.43	220.00

Notes: The table reports summary statistics of the estimation sample.

Table A.3: Robustness (population)

population (ln, gpw)	(1)	(2)	(3)	(4)	(5)
post x treated	0.0116	-0.00283	0.0218	0.0124	0.0102
	(0.0183)	(0.00805)	(0.0374)	(0.0277)	(0.0191)
GSM coverage	0.00699	-0.0120	0.0120	0.00104	0.00545
	(0.0119)	(0.00904)	(0.0261)	(0.0287)	(0.0234)
Observations	2,420	1,765	830	610	440
R-squared	0.999	1.000	0.997	0.999	1.000
#countries	10	10	10	10	10
#towns	220	220	220	220	220
share treated	.445	.445	.445	.445	.445
town FE	✓	✓	✓	✓	✓
country x year FE	✓	✓	✓	✓	✓
w/o capital+landing point	✓	✓	✓	✓	✓
w/o regional capitals	✓	✓	✓	✓	✓
w/o population >100k	✓	✓	✓	✓	✓
years	-7 to 3	2000-connection year+3	1995;2000;2005;2010	2000;2005;2010	one pre + one post

Notes: Population measured as the logarithmic sum of of inhabitants per square kilometer within town area and a 2 kilometer buffer from *Gridded Population of the World* (GPW). Columns show different periods. GSM mobile coverage is calculated as the percentage share of town area covered with signal. All specifications are estimated on a sample restricted by excluding landing-point cities and capitals, regional capitals, and cities with more than 100,000 inhabitants and include town and country-year fixed effects. Robust standard errors clustered by town reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.4: Robustness (population control)

	combined		intensive		extensive	
	(1)	(2)	(3)	(4)	(5)	(6)
post x treated	0.0703**	0.0661**	0.0513**	0.0503**	0.0516*	0.0473*
	(0.0349)	(0.0318)	(0.0231)	(0.0229)	(0.0282)	(0.0244)
GSM coverage	0.0486	0.0461	0.0477**	0.0471*	0.0281	0.0255
	(0.0342)	(0.0336)	(0.0240)	(0.0241)	(0.0263)	(0.0255)
population (ln, gpw)		0.359*		0.0793		0.375**
		(0.190)		(0.113)		(0.148)
observations	2,420	2,420	2,420	2,420	2,420	2,420
R-squared	0.943	0.944	0.947	0.947	0.924	0.925
#countries	10	10	10	10	10	10
#towns	220	220	220	220	220	220
share treated	.445	.445	.445	.445	.445	.445
town FE	✓	✓	✓	✓	✓	✓
country x year FE	✓	✓	✓	✓	✓	✓
w/o capital+landing point	✓	✓	✓	✓	✓	✓
w/o regional capitals	✓	✓	✓	✓	✓	✓
w/o population >100k	✓	✓	✓	✓	✓	✓

Notes: Light intensity in the combined models is measured as the logarithmic sum of light intensities of DMPS-OLS pixels within town area and a 2 kilometer buffer as in Table 1.1. The measure for the intensive margin is logarithmic mean light intensity and for the extensive margin logarithmic sum of pixels (coded as 1 if a pixel is lit). GSM mobile coverage is calculated as the percentage share of town area covered with signal. Population is measured as inhabitants per square kilometer within town area and a 2 kilometer buffer from *Gridded Population of the World* (GPW). All specifications are estimated on a sample restricted by excluding landing-point cities and capitals, regional capitals, and cities with more than 100,000 inhabitants and include town and country-year fixed effects. Robust standard errors clustered by town reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.5: Electricity grid (from *Afrobarometer*)

electricity grid	(1)	(2)	(3)	(4)	(5)	(6)
post x treated	0.000387	-0.0359	0.0411	0.0579	-0.0731	-0.0914
	(0.103)	(0.0688)	(0.114)	(0.0766)	(0.211)	(0.173)
GSM coverage	0.0623	0.0205	0.0580	0.00348	0.107	-0.00385
	(0.111)	(0.0901)	(0.115)	(0.106)	(0.171)	(0.158)
observations	270	270	250	250	102	102
R-squared	0.680	0.806	0.675	0.784	0.720	0.814
#countries	6	6	6	6	4	4
#towns	94	94	88	88	37	37
share treated	.351	.351	.307	.307	.351	.351
town FE	✓	✓	✓	✓	✓	✓
country x year FE	✓	✓	✓	✓	✓	✓
weights		✓		✓		✓
w/o capital+landing point			✓	✓		
w/o nodal cities					✓	✓

Notes: Access to the electricity grid was aggregated at the town/city level and comes from Afrobarometer (rounds 1 to 4). GSM mobile coverage is calculated as the percentage share of town area covered with signal. All specifications are estimated on a sample restricted by excluding landing point cities and capitals, regional capitals, and cities with more than 100,000 inhabitants and include town and country-year fixed effects. Robust standard errors clustered by town reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.3 Example: Case of Benin

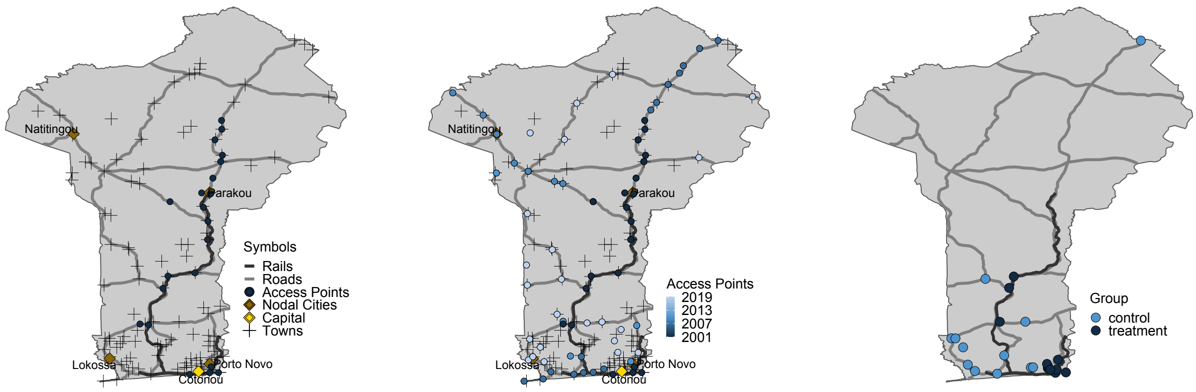
Benin is a good example for how the national backbone was rolled out and how it influenced Internet usage. It is one of the countries that was connected by the SAT-3 SMC, which brought an international connection of 45 Mbps [Chabossou, 2007]. The rollout of the national backbone was planned by Benin Telecoms SA, the fixed-line monopolist which manages the gateway to the national Internet, operates as the national carrier, and administers the national domain (.bj). Benin Telecoms SA is state-owned and offers permanent ADSL connections with up to 2 Mbps [Agyeman, 2007].

Infrastructure Rollout Following Chabossou [2007], the SAT-3 SMC landed in Cotonou, Benin's biggest city, the location of the seat of government, and 40 kilometers away from Benin's capital, Porto-Novo. Close by, in Abomey-Calavi Benin's hub is located as well. These cities form with Godomey Benin's largest agglomeration with nearly 2.5 million inhabitants (about a third of Benin's population). From there, first, a connection to Parakou with a 425 kilometers optical fiber cable was constructed in 2001. Parakou is Benin's next largest economic center with more than 150,000 inhabitants in the 2002 census and the capital of the Borgou department. This connection was constructed along Benin's railway line and roads network (Figure A.3.1) and connected further smaller towns on its way, e.g., Savalou with 30,000 inhabitants. Next, from Parakou connections to the borders to Niger, in the north-east, and Burkina Faso, in the north-west, were constructed along the road network, transforming Benin to a sub-regional digital hub interconnecting Togo, Nigeria, Burkina Faso, and Niger. The first kilometers of the fiber-optic backbone and access points were still constructed until 2001. Consequently, Benin Telecoms SA investment in the telecommunications sector peaked in 2001 with more than 80 billion US-Dollar. The connection towards Burkina Faso and Togo was constructed through Natitingou, the capital of the Atakora department. Again, connecting also further smaller towns, such as Kandi or Djougou, incidentally. Only later on, further rural towns were connected when constructing backbone circles to make the network more reliable, e.g., Nikki, Ségbana, and Banikoara.

Internet Usage All transmission happens via Benin Telecoms SA. They offer data transmission networks to mostly commercial clients (banks, hotels, ministries, etc.) in packets.[1] Having grown exponentially, thousands of cybercafés offer Internet access. While international institutions, major corporations, service providers, and some cybercafés have

[1] Network interconnectivity enables new providers to use the incumbent's infrastructure instead of having to invest greatly to build an own one, which incentivizes competitive adaptation. There are, in addition to the former monopolist, which still owns the infrastructure, three licensed providers. However, there are about 50 providers operating without a licence and there is no adequate framework for regulation.

Figure A.3.1: Rollout in Benin



(1) Access points in 2001 (2) All access points (3) Treatment/control group

Notes: The figure outlines the rollout of access points in Benin. Besides access points, the maps include the capital city, nodal cities, and all towns. Railroads and roads are included as well. In the left panel, the early rollout with access points being constructed until the arrival of the SMC in 2001 is shown. The middle panel depicts further access points and their respective construction years. The right panel shows the towns of your analysis divided into treatment and control group.

permanent links, home access remains very limited [Chabossou, 2007]. Still, in 2007 only 25 percent of people in Benin's population have used the Internet at least one time. Access is mainly at cybercafés (21 percent) or at the workplace (2.2 percent) while Internet at home remains a luxury. Though, workplace Internet usage is low, it indicates that firms are great adopters of broadband Internet. Among the groups of higher education, Internet usage is also a lot higher. Therefore, we expect local growth through firm's productivity to increase induced by broadband Internet.

A.4 Estimation Sample

We focus on early SMCs bringing Internet connections at basic speeds to SSA in the early 2000s. Therefore, we do not consider countries which were connected after 2008 for the first time, when the next generation of SMCs (which allowed for much higher speeds) landed. This leaves 27 countries, which are listed in Table A.1. Among the first countries that were connected are Djibouti, where an SMC landed in 1999, Namibia, which was connected by a trans-national fiber cable from South Africa in 1999, and Senegal, where an SMC landed in 2000.[2] In 2001, nine more countries were connected by a single SMC, the SAT-3 cable. In the following years until 2008, 17 more countries got an SMC connection or were connected through a neighboring country.

However, not all countries that were connected until 2008 had constructed a national backbone infrastructure before the respective SMC or the connection through a neighboring country arrived. In this case, the treatment group is missing as there are no towns with national backbone access right after the connection. This reduces the number of countries in our analysis to 23.[3] Moreover, eleven countries established only in nodal cities access points before Internet became available countrywide.[4] Therefore, there are no towns in the treatment group and we cannot estimate on these countries. Finally, we cannot consider Namibia in our analysis because it did not construct further access points after getting the Internet connection. Therefore, we are unable to define a control group. This leaves 12 countries for our analysis.

Due to the staggered arrival of SMCs, this sample represents an unbalanced panel. In our main specification, we take a conservative approach and estimate on a balanced panel. Therefore, we truncate the data to attain a balanced panel. Malawi and Mozambique only have two post-treatment years. They were connected in 2007 and 2006, respectively, and got upgraded by an SMC with more capacity in 2010 and 2009, respectively. Thus, only three years lie between the first connection to the Internet and the Internet capacity upgrade for both countries. Hence, estimating on a balanced sample with three post-treatment years leaves us with a sample of ten countries.

[2] Djibouti and Senegal were connected as single SSA countries through bigger international multi-country SMCs. Djibouti was connected with SeaMeWe-3, which connected Northern and Western Europe with Eastern Asia and Australia. Senegal was connected with Atlantis-2, which went from Spain and Portugal through the Canary Islands to Brazil and Argentina and landed on Sengal's shores on the way.

[3] Central African Republic has not yet constructed a national backbone infrastructure. In Lesotho, the access points were established in 2009 three years after being connected through South Africa. In Djibouti, the first access points were established in 2007, which is eight years after the first SMC connection. Nigeria established its first access points in 2003, which is two years after the arrival of the first SMC.

[4] Guinea-Bissau, Lesotho, and Swaziland established all access points until today only in nodal cities.

A.5 Further Robustness Checks

Spatial Correlation To account for potential spillover effects, we cluster standard errors at the state level for robustness. However, it might be the case that spatial correlation between the towns' location requires correction of the standard errors. Following Conley [1999] we re-estimate Equation (1.1) correcting standard errors for spatial correlation. Results show statistical significance at the 1% level (Table A.5.1).

Table A.5.1: Spatial correlation

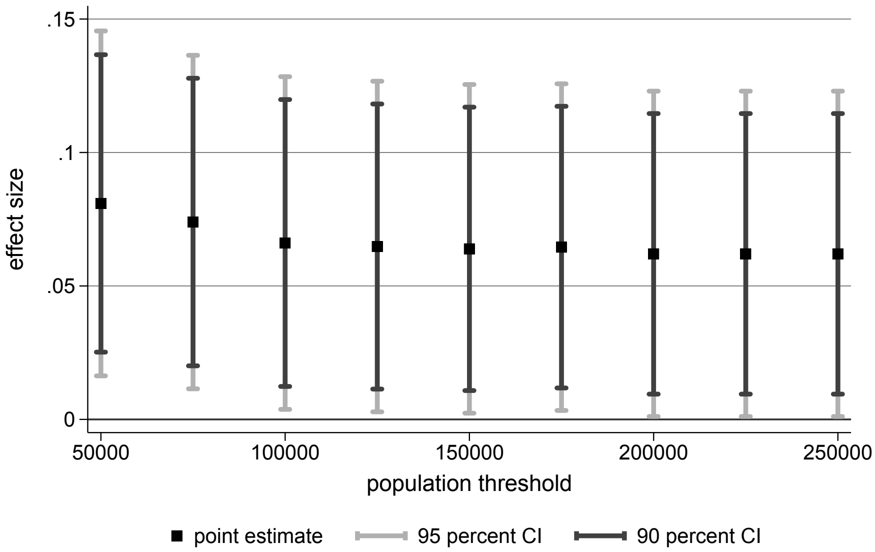
light intensity (log)	(1)	(2)	(3)	(4)	(5)
post x treated	0.0462***	0.0532***	0.0591***	0.0633***	0.0703***
	(0.0179)	(0.0190)	(0.0215)	(0.0226)	(0.0232)
GSM coverage					0.0486**
					(0.0236)
observations	3,190	3,069	2,563	2,420	2,420
R-squared	0.002	0.003	0.003	0.004	0.006
town FE	✓	✓	✓	✓	✓
country x year FE	✓	✓	✓	✓	✓
w/o capital+landing point		✓	✓	✓	✓
w/o regional capitals			✓	✓	✓
w/o population >100k				✓	✓

Notes: Light intensity is measured as the logarithmic sum of light intensities of DMPS-OLS pixels within town area and a 2 kilometer buffer. GSM mobile coverage is calculated as the percentage share of town area covered with signal. Sample successively restricted by excluding landing-point cities and capitals, regional capitals, and cities with more than 100,000 inhabitants and all specifications include town and country-year fixed effects. Conley standard errors to account for spatial correlation reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Definition of Nodal Cities For the main results, we defined all towns with more than 100,000 inhabitants as nodal cities. However, this threshold is chosen arbitrarily. Therefore, when estimating the specification in Column (2) of Table 1.1, we vary the population threshold as a further robustness check. Figure A.5.1 shows that the estimate remains independently of the chosen population threshold. There is a small tendency of a declining effect from a threshold of 50,000 over 75,000 to 100,000 inhabitants.

Definition of Internet Access For the main results, Internet access was defined for towns with an access point to the national backbone within 10 kilometers as within this distance Internet should be accessible. However, this threshold is not sharp with respect to Internet access. Therefore, when estimating the specification in Column (2) of Table 1.1, we vary this distance from 0 to 50 kilometers as a further robustness check in Figure A.5.2. For very low distances, Internet access might be higher over the whole town's area. For very high distances, Internet access can still be provided with an additional fiber-cable rollout. This rollout is not in place in all towns and cannot be observed with our data. Moreover, one

Figure A.5.1: Robustness (nodal cities)

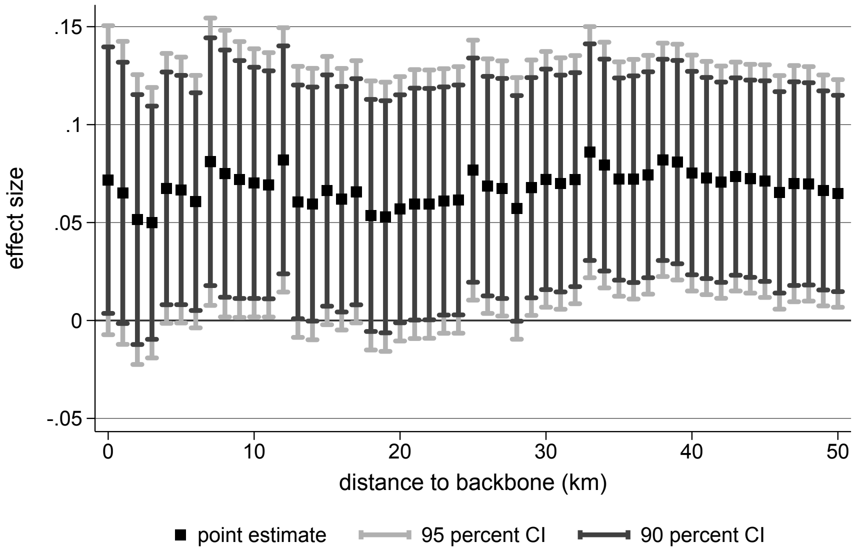


Notes: Variation of population thresholds are shown. Coefficients for the specification in Column (2) of Table 1.1. Robust standard errors clustered by town.

should note that the distance to the access point influences the sample. When allowing for higher distances to define Internet access, the control group shrinks. This means that also some countries drop out of the sample if only treated towns remain. The composition of treatment and control group is not constant as well. It is important that the treatment group contains only towns that can use the Internet, while in the control group Internet should not be accessibly. First, for low distances, the latter might not hold anymore. Hence, the ATT compares a treatment group with a control group which contains actually treated towns. Second, for high distances, the treatment group might contain some towns without actual Internet access. At the same time, the control group shrinks in size as only very few towns remain that did not have an access point in a certain higher distance at the beginning. These towns might also be less developed and less growing because of their unfortunate location. Hence, though the treatment group is contaminated in this case, the ATT might find high effects.

Definition of Control Group A further concern might be that towns being connected through an access point which was constructed many years after the first Internet connection are not comparable to the treated towns which were connected through an access point constructed before the first Internet connection. However, Table A.5.2 shows that when restricting the year when control towns were connected does not have a strong impact on the estimate. In contrast to the a priori concern, economic and statistical significance increase when only including towns that were connected shortly after a

Figure A.5.2: Robustness (Internet access)



Notes: Variation of the distance to the next access point to the national backbone are shown. Coefficients for the specification of Column (2) of Table 1.1. Robust standard errors clustered by town.

countrywide Internet connection was established to the control group. The last column repeats the main effect estimate.

Table A.5.2: Robustness (connected control towns)

light intensity (log)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
post x treated	0.277***	0.267***	0.115*	0.121**	0.111**	0.110**	0.107**	0.0980**	0.0700*	0.0625*	0.0703**
	(0.0999)	(0.1000)	(0.0604)	(0.0538)	(0.0480)	(0.0471)	(0.0457)	(0.0458)	(0.0401)	(0.0378)	(0.0349)
GSM coverage	0.0758	0.0790	0.0947*	0.102**	0.0696	0.0688	0.0687	0.0682	0.0592	0.0520	0.0486
	(0.0642)	(0.0598)	(0.0528)	(0.0480)	(0.0428)	(0.0427)	(0.0422)	(0.0421)	(0.0398)	(0.0382)	(0.0342)
observations	1,265	1,364	1,573	1,650	1,793	1,804	1,837	1,848	2,123	2,233	2,420
R-squared	0.954	0.951	0.950	0.950	0.949	0.949	0.949	0.949	0.944	0.944	0.943
#countries	10	10	10	10	10	10	10	10	10	10	10
#cities	115	124	143	150	163	164	167	168	193	203	220
share treated	.852	.79	.685	.653	.601	.598	.587	.583	.508	.483	.445
town FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
country x year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
w/o capital+landing point	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
w/o regional capitals	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
w/o population >100k	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
backbone border	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020

Notes: Light intensity is measured as the logarithmic sum of light intensities of DMPS-OLS pixels within town area. GSM mobile coverage is calculated as the percentage share of town area covered with signal. Sample restricted by excluding landing point cities and capitals, regional capitals, and cities with more than 100,000 inhabitants. Robust standard errors clustered by town reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Longer Post-Treatment Trends and Shorter Pre-Treatment Trends For robustness, we show that the results do not depend on the chosen window around the treatment year. Table A.5.3 shows in Column (2) results for a longer post-treatment period. This reduces the sample size to six countries which were connected that early for the first time that they have at least five post-treatment years before a speed-upgrade SMC arrived. Column (3) reduces the pre-treatment period by two years to five pre-treatment years. The number of countries and towns remains as the data on NTL goes a lot further back in time than the connection year of the first country. In both cases, the point estimate and its level of statistical significance increase in comparison to the baseline specification of Column (1). Column (2) shows that growth rates increase further even five years after the treatment. Column (3) indicates that there is no divergence in the years prior to the treatment. In fact, treated and control group towns have a marginal tendency to converge before the treatment and diverge strongly after the treatment.

Table A.5.3: Robustness (estimation window and unbalanced panel)

light intensity (log)	(1)	(2)	(3)	(4)	(5)	(6)
post x treated	0.0703**	0.156***	0.0904***	0.0615*	0.0890**	0.0791**
	(0.0349)	(0.0555)	(0.0343)	(0.0333)	(0.0400)	(0.0327)
GSM coverage	0.0486	0.0743	0.0579	0.0472	0.0544	0.0509
	(0.0342)	(0.0484)	(0.0367)	(0.0314)	(0.0381)	(0.0335)
observations	2,420	1,729	1,980	2,690	2,739	2,196
R-squared	0.943	0.926	0.948	0.945	0.939	0.949
#countries	10	6	10	12	11	12
#towns	220	133	220	247	222	247
share treated	.445	.436	.445	.417	.459	.417
town FE	✓	✓	✓	✓	✓	✓
country x year FE	✓	✓	✓	✓	✓	✓
w/o capital+landing point	✓	✓	✓	✓	✓	✓
w/o regional capitals	✓	✓	✓	✓	✓	✓
w/o population >100k	✓	✓	✓	✓	✓	✓
balanced panel	✓	✓	✓			
pre-treatment years	7	7	5	7	7	5
post-treatment years	3	5	3	3	5	3

Notes: Light intensity is measured as the logarithmic sum of light intensities of DMPS-OLS pixels within town area and a 2 kilometer buffer. GSM mobile coverage is calculated as the percentage share of town area covered with signal. Sample successively restricted by excluding landing point cities and capitals, regional capitals, and cities with more than 100,000 inhabitants. Robust standard errors clustered by town reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Unbalanced Panel Estimating on a balanced panel has the advantage of not depending on the sample composition close to the period boundaries. Relaxing this restriction, however, allows to estimate on a bigger sample and therefore shows that the results have high external validity. Table A.5.3 repeats the estimations from before on an unbalanced sample. Therefore, in the baseline specification, the sample increases by two countries (Column (4)). In Column (5), the sample shrinks only by one country, instead of four as in Column (2) in the balanced sample. In Column (6), the sample size again remains at the higher level. The estimates are only slightly lower in comparison to the balanced sample. Again, it can be observed that the main estimate increases from Column (4) to Column (5), when a longer post-treatment period is applied. The same holds for the comparison of Columns (4) and (6).

Parts of SSA as Fixed Effects Table A.5.4 show the results when re-estimating Equation (1.1) with part-year fixed effects instead of country-year fixed effects. The parts are East, Southern, West, and Central Africa. This specification allows for more countries, as it is not necessary for a single country to have a treatment and control group. On the other hand, a comparison is made within the parts of SSA, such that the growth path of different parts is considered. Again, nodal cities are removed in the stepwise procedure. As before, the estimate increases column by column. It is statistically significant at the 1% level in all specifications. Moreover, the point estimate has a higher level. In our refereed specification (Column (5)), it is .125.

Table A.5.4: Robustness (parts of SSA as fixed effects)

light intensity (log)	(1)	(2)	(3)	(4)	(5)
post x treated	0.0904***	0.0997***	0.118***	0.122***	0.125***
	(0.0266)	(0.0283)	(0.0317)	(0.0337)	(0.0337)
GSM coverage					0.0320
					(0.0265)
observations	4,895	4,697	3,718	3,553	3,553
R-squared	0.965	0.951	0.942	0.923	0.923
#countries	16	15	15	14	14
#towns	445	427	338	323	323
share treated	.364	.34	.334	.313	.313
town FE	✓	✓	✓	✓	✓
part x year FE	✓	✓	✓	✓	✓
w/o capital+landing point		✓	✓	✓	✓
w/o regional capitals			✓	✓	✓
w/o population >100k				✓	✓

Notes: Light intensity is measured as the logarithmic sum of light intensities of DMPS-OLS pixels within town area and a 2 kilometer buffer. GSM mobile coverage is calculated as the percentage share of town area covered with signal. Sample successively restricted by excluding landing-point cities and capitals, regional capitals, and cities with more than 100,000 inhabitants and all specifications include town and country-year fixed effects. Robust standard errors clustered by town reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Mobile Coverage Lags When controlling for mobile coverage, we therefore control for the difference in having a different ICT infrastructure available. As it might take some time for an infrastructure to affect economic outcomes as we have seen for Internet availability, we also include different lags for mobile coverage instead of current mobile coverage. Table A.5.5 shows that mobile coverage induces economic growth with a lag of one year. All other lags remain insignificant. However, in all lag specifications, the main effect is robust.

Table A.5.5: Robustness (mobile coverage lags)

light intensity (log)	(1)	(2)	(3)	(4)	(5)	(6)
post x treated	0.0703**	0.0707**	0.0647*	0.0646*	0.0646*	0.0610*
	(0.0349)	(0.0349)	(0.0346)	(0.0343)	(0.0343)	(0.0342)
GSM coverage	0.0486					
	(0.0342)					
GSM coverage (lag 1)		0.0734**				
		(0.0359)				
GSM coverage (lag 2)			0.0235			
			(0.0353)			
GSM coverage (lag 3)				0.0491		
				(0.0335)		
GSM coverage (lag 4)					0.0484	
					(0.0346)	
GSM coverage (lag 5)						0.0384
						(0.0362)
observations	2,420	2,420	2,420	2,420	2,420	2,420
R-squared	0.943	0.943	0.943	0.943	0.943	0.943
#countries	10	10	10	10	10	10
#cities	220	220	220	220	220	220
share treated	.445	.445	.445	.445	.445	.445
City FE	✓	✓	✓	✓	✓	✓
Country x Year FE	✓	✓	✓	✓	✓	✓
w/o capital+landing point	✓	✓	✓	✓	✓	✓
w/o regional capitals	✓	✓	✓	✓	✓	✓
w/o population >100k	✓	✓	✓	✓	✓	✓

Notes: Light intensity is measured as the logarithmic sum of light intensities of DMPS-OLS pixels within town area and a 2 kilometer buffer. Mobile coverage as share of built-up area with the most basic technology (GSM). All specifications are estimated on a sample restricted by excluding landing point cities and capitals, regional capitals, and cities with more than 100,000 inhabitants and include town and country-year fixed effects. Robust standard errors clustered by town reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.6 Mobile Coverage

In our preferred specification, we control for mobile coverage. When controlling for mobile coverage, the main estimate increases slightly and is estimated more precisely. Next, we re-estimate Equation (1.1) with mobile coverage as outcome variable and follow the stepwise removal of nodal cities. In Table A.6.1, Internet access is negatively associated with mobile coverage in all specifications. This means that control group towns catch up to the treatment group with respect to the coverage of the mobile network. This effect is irrespective of controlling for population (Column (5)) and light intensity (Column (6)), which are even jointly not significant as control variables (Column (7)).

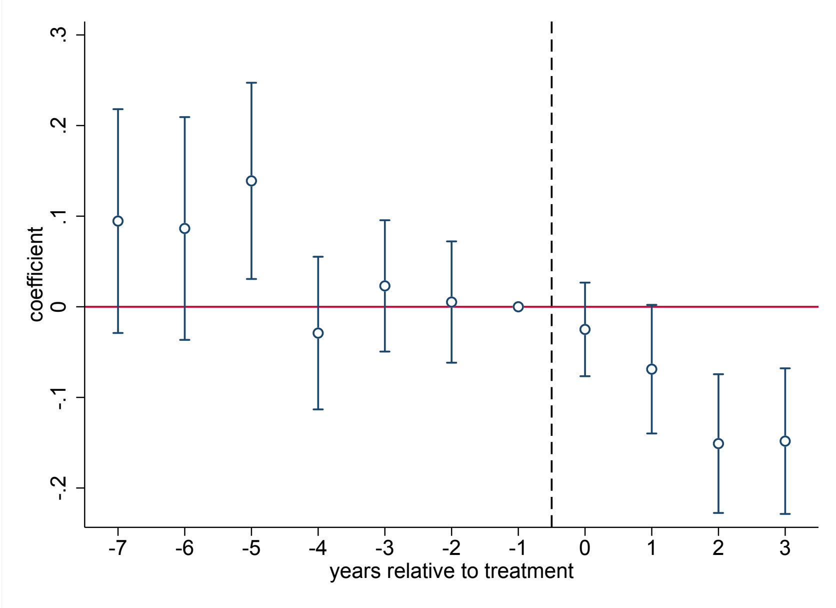
Table A.6.1: Mobile coverage

mobile coverage	(1)	(2)	(3)	(4)	(5)	(6)	(7)
post x treated	-0.123***	-0.134***	-0.155***	-0.142***	-0.144***	-0.145***	-0.146***
	(0.0333)	(0.0342)	(0.0364)	(0.0374)	(0.0368)	(0.0374)	(0.0369)
population (ln, gpw)					0.141		0.127
					(0.196)		(0.193)
light intensity						0.0413	0.0393
						(0.0294)	(0.0291)
observations	3,190	3,069	2,563	2,420	2,420	2,420	2,420
R-squared	0.816	0.816	0.816	0.817	0.817	0.817	0.817
#countries	10	10	10	10	10	10	10
#towns	290	279	233	220	220	220	220
share treated	.493	.473	.468	.445	.445	.445	.445
town FE	✓	✓	✓	✓	✓	✓	✓
country x year FE	✓	✓	✓	✓	✓	✓	✓
w/o capital+landing point		✓	✓	✓	✓	✓	✓
w/o regional capitals			✓	✓	✓	✓	✓
w/o population >100k				✓	✓	✓	✓

Notes: Mobile coverage as share of built-up area with the most basic technology (GSM). All specifications are estimated on a sample restricted by excluding landing point cities and capitals, regional capitals, and cities with more than 100,000 inhabitants and include town and country-year fixed effects. Robust standard errors clustered by town reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

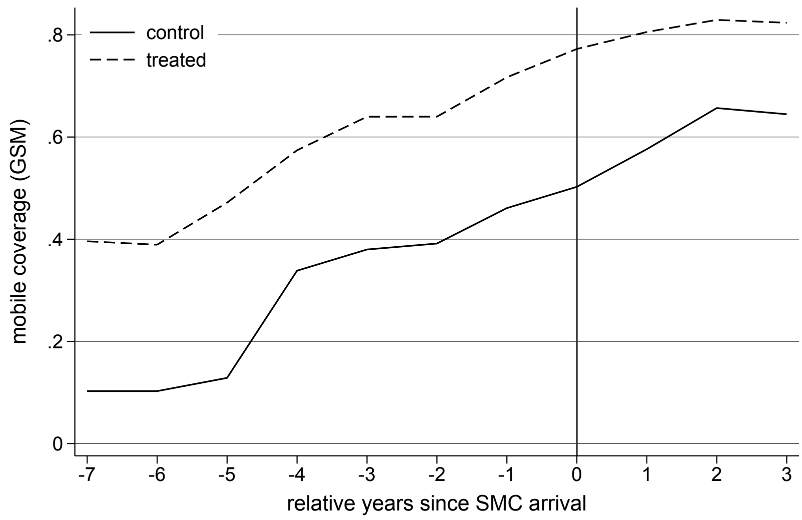
Figure A.6.1 indicates that in the early years before the Internet connections, the treatment group had a stronger rollout of the mobile network. While in the years right before the connection, the rollout speed was similar between the treatment and control group, after the connection, the control group caught up with respect to mobile coverage. This interpretation is in line with Figure A.6.2 which plots the mobile coverage separated between the treatment group and the control group towns without any fixed effects. While for treated towns the rollout stops slightly above 80 percent coverage shortly after the Internet connection, control group towns continue with the rollout in a linear manner and therefore catch up in the years after the Internet connection was established.

Figure A.6.1: Event-study coefficients for mobile coverage



Notes: Coefficients for event study specification of Column (4) of Table A.6.1. Robust standard errors clustered by town. Confidence intervals reported at the 95% level.

Figure A.6.2: Trends for mobile coverage



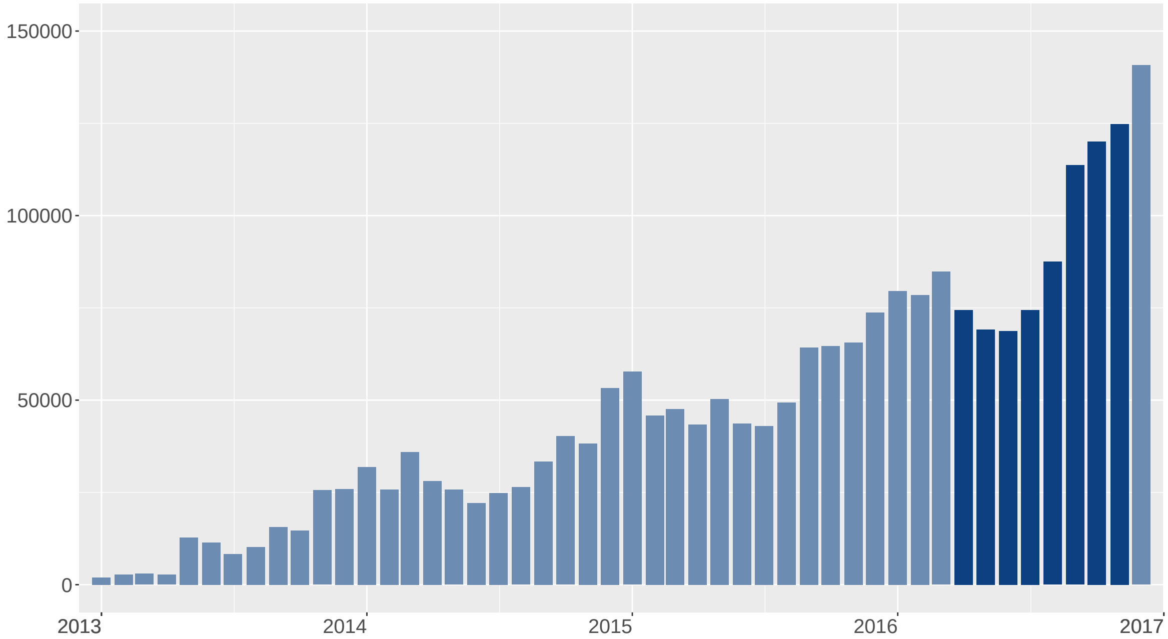
Notes: The figure depicts the average growth mobile coverage of the towns in the treatment and control group over a period of eleven years (seven before and three after the treatment year). The measurement is the towns' area covered by GSM technology.

Appendix B

Appendix to Chapter 2

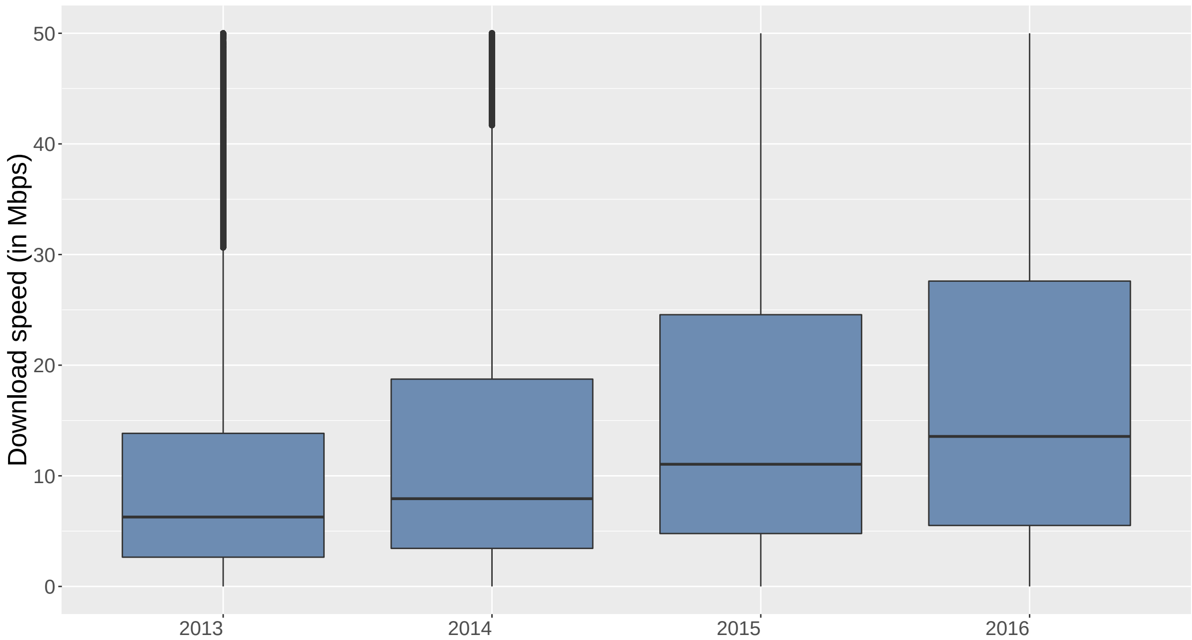
B.1 Additional Figures

Figure B.1.1: Number of speed tests from 2013 to 2016 (darker blue analysis period)



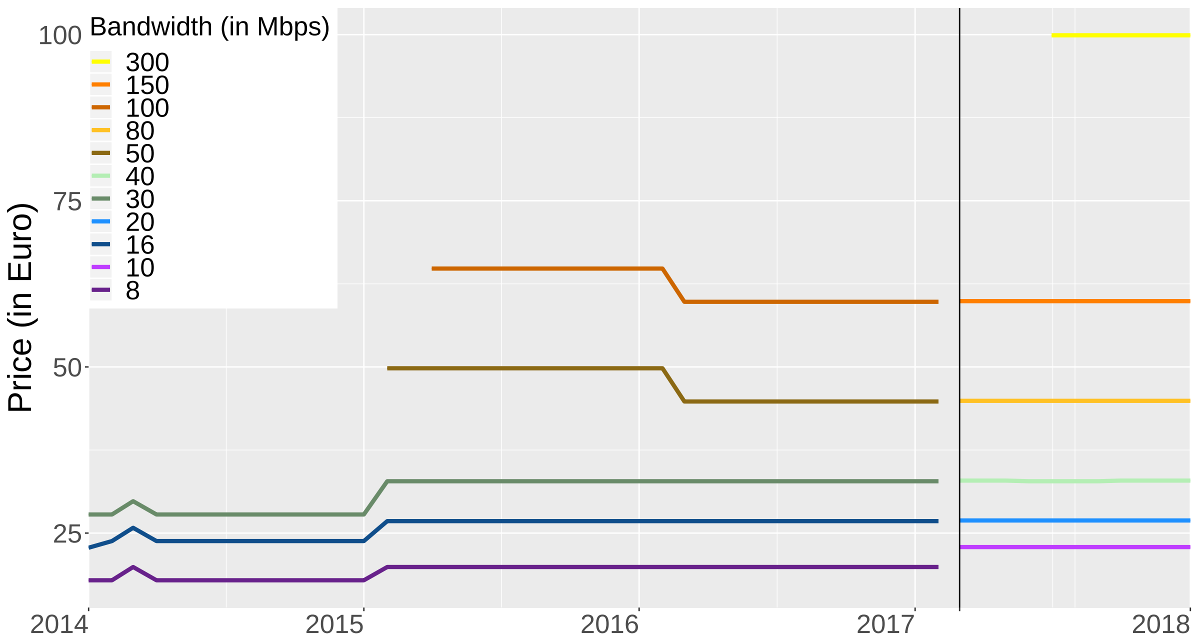
Notes: Data from RTR-NetTest from the Austrian Regulatory Authority for Broadcasting and Telecommunications

Figure B.1.2: Boxplot of speed tests from 2013 to 2016



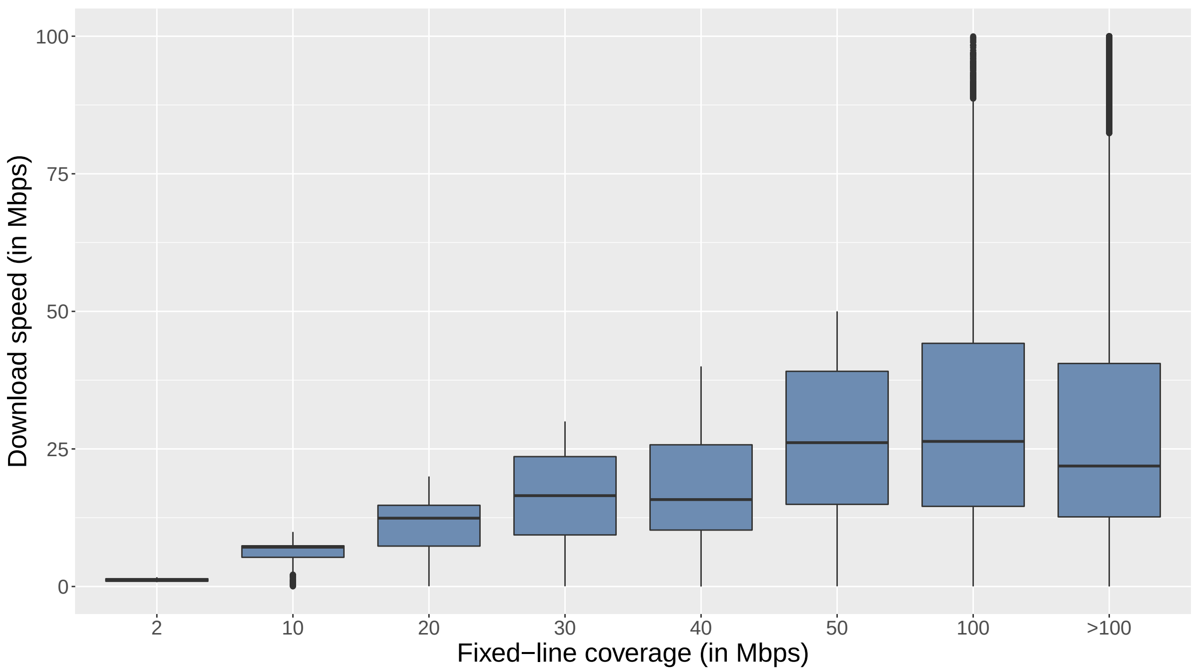
Notes: Data from RTR-NetTest from the Austrian Regulatory Authority for Broadcasting and Telecommunications

Figure B.1.3: A1 Telekom contracts (download speed and prices from 2014 to 2017)



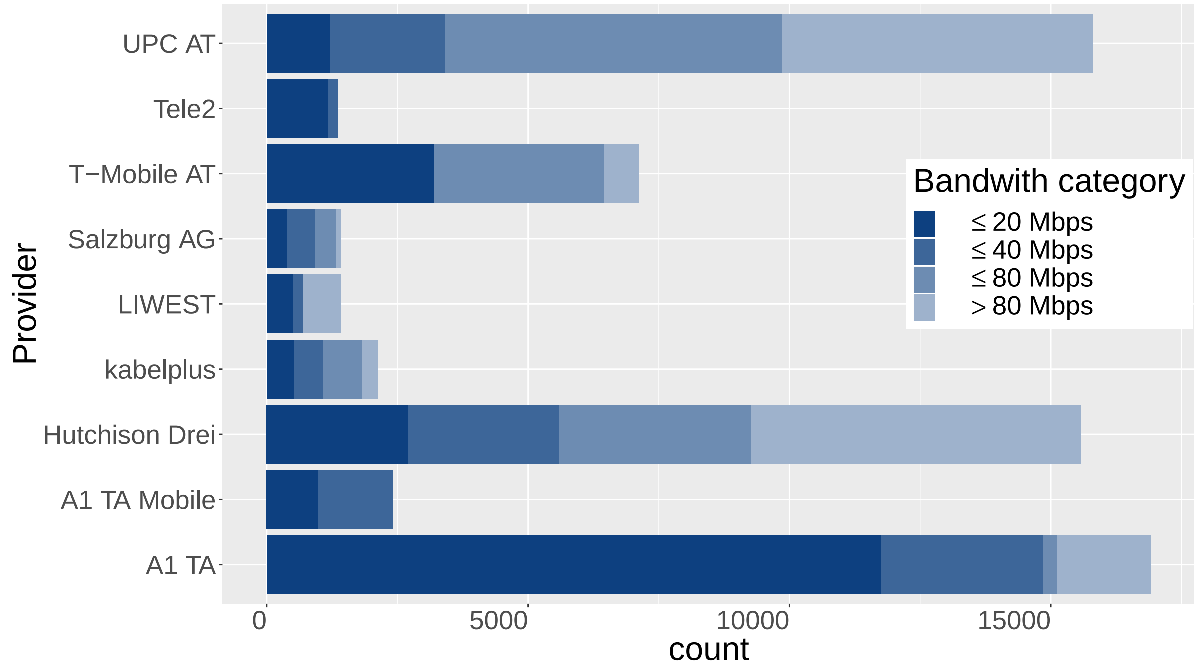
Notes: Data from *AK-Tarifwegweiser*

Figure B.1.4: Distribution of download speed by fixed-line coverage



Notes: Data from RTR-NetTest from the Austrian Regulatory Authority for Broadcasting and Telecommunications and *Broadband Atlas* provided by the Ministry for Transport, Innovation and Technology (bmvit)

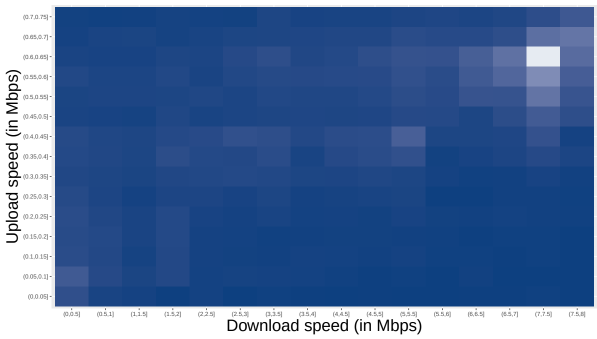
Figure B.1.5: Number of tests for each provider, classified into four bandwidths alternatives



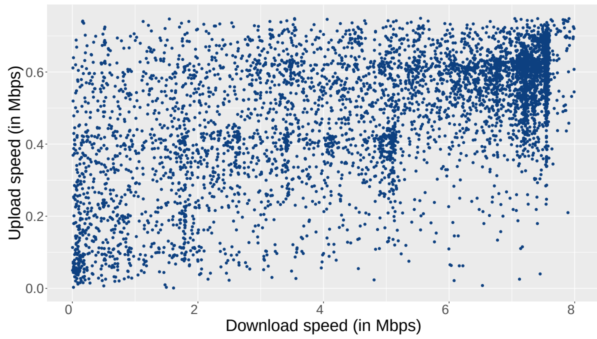
Notes: Data from RTR-NetTest from the Austrian Regulatory Authority for Broadcasting and Telecommunications

Figure B.1.6: Measured download and upload

(1) Heatmap



(2) Scatter plot

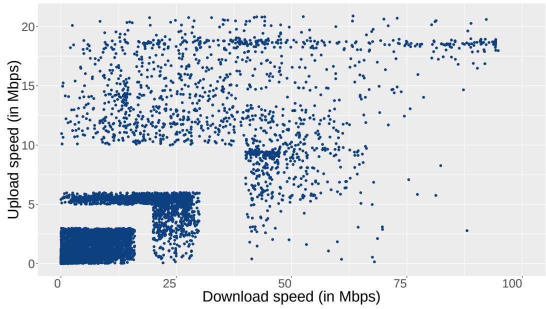


Notes: The figures show a heat map and a scatter plot for A1 Telekom tests with a measured download speed below 8 Mbps and a measured upload speed below .75 Mbps. These values refer to the lowest A1 Telekom contract in my data (Table B.2.3). The data in these figures was reduced to A1 Telekom tests and with respect to the download and upload speed due to higher visibility. The heat map depicts a bright hot spot just below the contractual bandwidths. Additionally, a horizontal tail shows that there are many tests within a certain upload speed corridor but with very different download speeds. However, a similar vertical tail cannot be found. Therefore, it is more important to consider the upload bandwidth when filtering the noise, although the download bandwidth is presumably more important to the consumer. Moreover, the figure displays why it is important to filter the noise: There are many tests above 5 Mbps download speed and below .5 upload speed. It is very likely that there exists an old contract with these contractual speeds. However, I have no information on the price of this contract. Therefore, it is better to remove these tests than assigning a wrong contract to them. Data from RTR-NetTest from the Austrian Regulatory Authority for Broadcasting and Telecommunications.

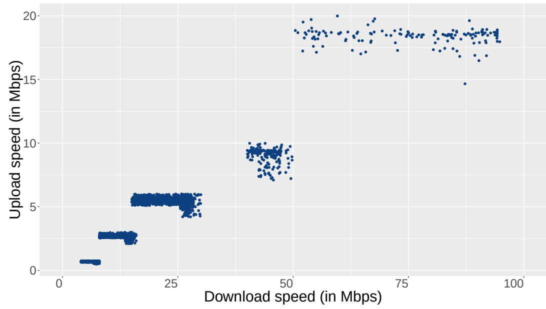
Figure B.1.7: Measured download speed for tests with fixed-line and mobile providers

A1

(1) Whole sample

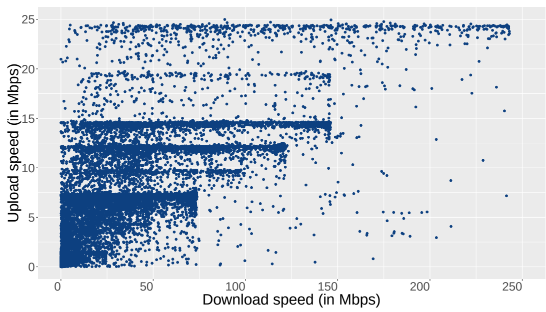


(2) Reduced sample

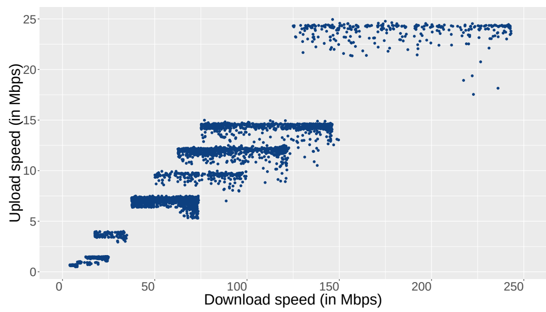


UPC

(3) Whole sample

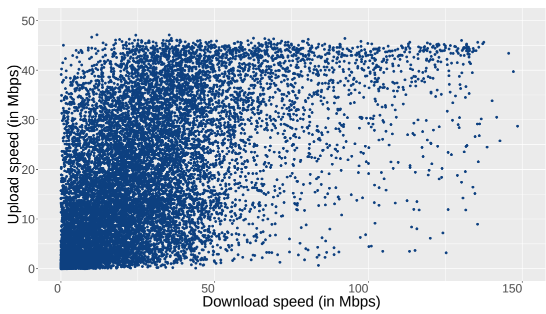


(4) Reduced sample

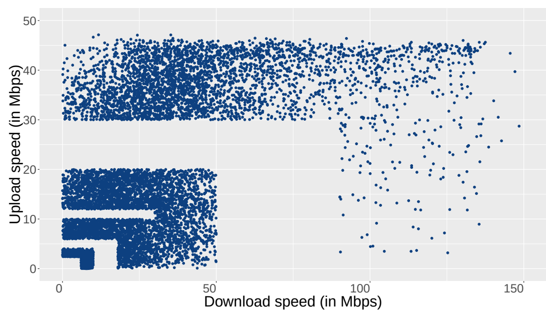


Drei

(5) Whole sample



(6) Reduced sample



Notes: The figures plot the tests for the most important providers for copper, cable, and mobile Internet. The data is already reduced to locations with full fixed-line coverage. In the left column, the whole sample is shown. In the right column, only the tests close to a known contract are shown (reduced sample). This column illustrates the corridors when filtering the noise. The observations with a measured download or upload speed which does not fall into one of the corridors are removed. In these tests, the speeds were measured imprecisely.Data from RTR-NetTest from the Austrian Regulatory Authority for Broadcasting and Telecommunications.

B.2 Additional Tables

Table B.2.1: Broadband coverage (by technologies and bandwidths, in percent) in Austria, overall and for rural areas, for 2012 and 2017, compared to EU 28

	NGA	VDSL	FTTP	DOCSIS 3.0	>30 Mbps	>100 Mbps
Austria (2012): total	69.5	50.5	6.3	35.3	–	–
Austria (2012): rural	14.4	0.0	1.2	13.4	–	–
Austria (2017): total	90.0	82.2	13.5	52.8	81.1	57.2
Austria (2017): rural	45.0	22.1	5.4	20.5	–	–
EU 28 (2017): total	80.1	53.4	26.8	44.7	79.0	50.8
EU 28 (2017): rural	46.9	32.5	11.3	10.8	–	–

Notes: Data from the Europe's Digital Progress (Country) Report

Table B.2.2: Contractual prices (in Euro) by provider and contractual download (in Mbps)

	8	10	15	16	20	25	30	35	40	50	70	75	100	125	150	250
A1 TA	19.90			26.80			32.80			44.80			59.80			
Tele2	17.80				24.90		29.90									
UPC AT	24.90			31.80	24.90	26.15		29.90				29.90	29.90	39.90	39.90	58.03
Hutchison Drei		18.00					25.00			35.00					45.00	
Salzburg AG			19.90				29.00			39.00	49.00				79.00	
kabelplus				19.90			29.90					39.90			69.90	99.90
A1 TA Mobile					26.33				39.90							
LIWEST					19.90			29.90					39.90		59.90	
T-Mobile AT					19.99					29.99					49.99	

Notes: Data from *AK-Tarifwegweiser*

Table B.2.3: Contractual upload by provider and contractual download (in Mbps)

	8	10	15	16	20	25	30	35	40	50	70	75	100	125	150	250
A1 TA	0.75			3.00			6.00			10.00			20.00			
Tele2	0.75				4.00		4.00									
UPC AT	0.70			1.00	1.00	1.50		4.00				7.50	10.00	12.50	15.00	25.00
Hutchison Drei		4.00					10.00			20.00					50.00	
Salzburg AG			1.50				3.00			5.00	7.00				10.00	
kabelplus				1.00			3.00					7.50			15.00	25.00
A1 TA Mobile					5.00				10.00							
LIWEST					3.00			6.00					10.00		20.00	
T-Mobile AT					5.00					10.00					30.00	

Notes: Data from *AK-Tarifwegweiser*, complemented with upload information from the providers' websites

Table B.2.4: Number of choices (in the whole sample)

	alternative							
	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps	Total
A1 TA	34987	8646	833	4791	3182	4642	0	57081
Hutchison Drei	0	0	0	0	7003	19681	15332	42016
LIWEST	653	210	0	754	0	0	0	1617
Salzburg AG	414	538	419	117	0	0	0	1488
T-Mobile AT	0	0	0	0	8623	8317	1701	18641
Tele2	1991	560	0	0	0	0	0	2551
UPC AT	2474	2725	7486	6547	0	0	0	19232
kabelplus	659	862	975	461	0	0	0	2957
Total	41178	13541	9713	12670	18808	32640	17033	145583

Notes: This table shows how the choices distribute between alternatives and choices. The numbers are shown for the whole set.

Table B.2.5: Number of choices (in the reduced sample)

	alternative							
	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps	Total
A1 TA	11619	5849	641	1699	4897	5214	0	29919
Hutchison Drei	0	0	0	0	3276	14416	8109	25801
LIWEST	123	93	0	88	0	0	0	304
Salzburg AG	167	322	195	64	0	0	0	748
T-Mobile AT	0	0	0	0	5046	7130	799	12975
Tele2	531	518	0	0	0	0	0	1049
UPC AT	450	883	3005	2404	0	0	0	6742
kabelplus	365	355	324	82	0	0	0	1126
Total	13255	8020	4165	4337	13219	26760	8908	78664

Notes: This table shows how the choices distribute between alternatives and choices. The numbers are shown for the reduced set.

Table B.2.6: Number of choices (with full fixed-line coverage in 2015)

	alternative							
	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps	Total
A1 TA	4336	2143	201	692	1596	1630	0	10598
Hutchison Drei	0	0	0	0	1257	4832	3627	9716
LIWEST	119	86	0	85	0	0	0	290
Salzburg AG	161	318	193	54	0	0	0	726
T-Mobile AT	0	0	0	0	1789	2804	343	4936
Tele2	341	159	0	0	0	0	0	500
UPC AT	230	734	2596	2222	0	0	0	5782
kabelplus	317	292	280	65	0	0	0	954
Total	5504	3732	3270	3118	4642	9266	3970	33502

Notes: This table shows how the choices distribute between alternatives and choices. The numbers are shown for the reduced set, restricted to full fixed-line coverage in 2015.

Table B.2.7: Summary of control variables with respect to full fixed-line coverage in 2015

coverage	<100 Mbps	≥ 100 Mbps
population (<5000)	0.56	0.17
population (<10000)	0.15	0.13
population (<100.000)	0.25	0.54
population (>100.000)	0.29	0.71
age (young)	0.20	0.19
age (middle)	0.49	0.51
age (old)	0.31	0.30
male	0.49	0.48
newborns	0.88	0.78
education (low)	0.23	0.23
education (middle)	0.52	0.49
education (high)	0.09	0.12

Notes: This table shows how observations with and without full fixed-line coverage in 2015 differ. Number are calculated from the whole sample. Population is divided into the following categories: less than 5000, less than 10,000, less than 100,000, at least 100,000 in 2014. Comparison is also provided for the number of newborns, the share of males, education level (compulsory, secondary and tertiary) and age distribution (young is defined as not older than 20 and old is defined as at least 60 years old).

Table B.2.8: Nested logit regression results

	(1)	(2)
price	-0.176***	-0.150***
	(0.00116)	(0.00106)
dissimilarity parameters		
$\tau_{fixed-line}$	1.988	1.866
	(0.01343)	(0.01363)
τ_{mobile}	2.413	2.197
	(0.03055)	(0.02692)
basic controls	✓	✓
sample	whole	reduced
observations	449,274	234,514
cases	64,182	33,502

Note: The regression table shows the nested logit results for the whole (Column (1)) and the reduced sample (Column (2)) with four fixed-line and three mobile alternatives. The sample is reduced to tests that had full fixed-line coverage (at least 100 Mbps) in 2015 to ensure that all alternatives are eligible and that the results are comparable. Weights with respect to the number of mobile and fixed-line contracts throughout Austria and robust standard errors are applied. It is controlled for whether the device from which the test was performed was connected via wireless LAN or LAN. Furthermore, the first zip-code digit and the mobile upload coverage from the *Broadband Atlas* are applied as geographical covariates. Standard errors are shown in parentheses. * p<0.05, ** p<0.01, *** p<0.001

Table B.2.9: Own- and cross-price elasticities (whole sample)

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-2.559	.755	.876	1.299	.258	.524	.203
≤ 40 Mbps	1.347	-4.425	1.04	1.353	.264	.54	.241
≤ 80 Mbps	1.117	.734	-5.031	1.603	.239	.467	.227
> 80 Mbps	1.092	.636	1.081	-6.205	.225	.469	.245
mobile ≤ 20 Mbps	1.12	.632	.809	1.126	-3.358	.869	.852
mobile ≤ 80 Mbps	.832	.479	.591	.886	.319	-4.259	1.553
mobile > 80 Mbps	.307	.204	.275	.437	.302	1.498	-4.547

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the whole sample (Table 2.1, Column (1)).

Table B.2.10: Own- and cross-price elasticities of the reduced sample with more control variables on municipality level

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-3.005	1.2	.946	1.19	.286	.704	.093
≤ 40 Mbps	1.318	-4.533	1.019	1.177	.31	.805	.189
≤ 80 Mbps	1.052	1.012	-5.527	1.652	.262	.659	.176
> 80 Mbps	.97	.866	1.238	-6.987	.249	.697	.204
mobile ≤ 20 Mbps	1.032	1.004	.848	1.079	-3.701	1.208	.748
mobile ≤ 80 Mbps	.664	.692	.574	.823	.317	-4.216	1.57
mobile > 80 Mbps	.109	.202	.191	.29	.25	1.991	-4.459

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample and with some control variables on the zip code level (Table 2.1, Column (3)).

Table B.2.11: Own- and cross-price elasticities (control variables on zip code level)

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-3.011	1.2	.936	1.195	.288	.711	.093
≤ 40 Mbps	1.321	-4.553	1.03	1.181	.311	.808	.189
≤ 80 Mbps	1.046	1.025	-5.52	1.638	.261	.657	.177
> 80 Mbps	.977	.868	1.221	-6.994	.253	.704	.201
mobile ≤ 20 Mbps	1.041	1.009	.845	1.095	-3.721	1.215	.748
mobile ≤ 80 Mbps	.672	.695	.571	.833	.318	-4.242	1.577
mobile > 80 Mbps	.11	.202	.191	.286	.25	1.996	-4.465

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample and with the complete set of control variables on the zip code level (Table 2.1, Column (4)).

Table B.2.12: Own- and cross-price elasticities (unweighted mean prices)

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-3.182	1.297	.946	1.256	.309	.751	.101
≤ 40 Mbps	1.418	-4.87	1.022	1.326	.34	.866	.216
≤ 80 Mbps	1.097	1.071	-5.804	1.79	.267	.689	.181
> 80 Mbps	1.045	1.001	1.26	-7.575	.272	.762	.243
mobile ≤ 20 Mbps	1.125	1.117	.834	1.167	-3.985	1.275	.847
mobile ≤ 80 Mbps	.727	.767	.588	.898	.339	-4.583	1.711
mobile > 80 Mbps	.118	.23	.185	.334	.279	2.103	-4.785

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with unweighted mean prices.

Table B.2.13: Own- and cross-price elasticities (no weights)

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-2.467	.722	.598	.724	.468	.874	.023
≤ 40 Mbps	.829	-3.68	.641	.759	.495	.974	.047
≤ 80 Mbps	.659	.601	-4.397	1.149	.435	.828	.051
> 80 Mbps	.594	.536	.877	-5.66	.42	.888	.096
mobile ≤ 20 Mbps	.618	.564	.524	.668	-2.556	1.321	.309
mobile ≤ 80 Mbps	.382	.371	.337	.482	.437	-2.907	1.019
mobile > 80 Mbps	.016	.028	.033	.081	.165	1.641	-2.917

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with no weights.

Table B.2.14: Own- and cross-price elasticities (control for incumbent choice)

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-2.594	1.59	.404	1.065	.275	.423	.04
≤ 40 Mbps	2.158	-5.885	.879	1.688	.387	.884	.166
≤ 80 Mbps	.492	.778	-6.525	3.609	.353	1.04	.319
> 80 Mbps	.599	.7	1.708	-8.364	.335	1.293	.396
mobile ≤ 20 Mbps	1.055	1.052	1.057	2.086	-4.611	1.621	1.041
mobile ≤ 80 Mbps	.365	.579	.777	2.086	.386	-5.06	2.036
mobile > 80 Mbps	.044	.134	.287	.755	.318	2.519	-5.694

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample and with a control variable if a tester had a A1 Telekom contract.

Table B.2.15: Own- and cross-price elasticities (random sample split in halves)

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-2.978	1.198	.941	1.194	.283	.678	.09
≤ 40 Mbps	1.338	-4.537	1.017	1.197	.309	.777	.187
≤ 80 Mbps	1.049	.995	-5.584	1.71	.263	.659	.187
> 80 Mbps	1	.888	1.311	-7.154	.251	.696	.219
mobile ≤ 20 Mbps	1.025	.99	.853	1.061	-3.691	1.184	.793
mobile ≤ 80 Mbps	.664	.68	.589	.824	.32	-4.269	1.635
mobile > 80 Mbps	.106	.196	.2	.302	.262	1.981	-4.486

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with a random subset of only half of the observations.

Table B.2.16: Own- and cross-price elasticities (other half of the random subset)

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-3.003	1.209	.921	1.18	.283	.718	.093
≤ 40 Mbps	1.314	-4.564	.974	1.212	.312	.843	.194
≤ 80 Mbps	1.027	.981	-5.536	1.781	.258	.661	.17
> 80 Mbps	.941	.881	1.308	-7.072	.249	.726	.232
mobile ≤ 20 Mbps	1.025	1.028	.831	1.094	-3.702	1.208	.729
mobile ≤ 80 Mbps	.669	.724	.566	.857	.31	-4.239	1.524
mobile > 80 Mbps	.113	.219	.192	.348	.251	2.045	-4.625

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with the other half of the random subset.

Table B.2.17: Own- and cross-price elasticities (5 percent random subset)

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-2.792	1.081	.835	1.246	.241	.628	.097
≤ 40 Mbps	1.062	-3.971	.893	1.076	.263	.779	.233
≤ 80 Mbps	.945	1.002	-5.102	1.631	.229	.573	.097
> 80 Mbps	.912	.787	1.074	-6.021	.19	.563	.131
mobile ≤ 20 Mbps	.907	.988	.75	.974	-3.312	1.047	.573
mobile ≤ 80 Mbps	.605	.766	.502	.759	.269	-3.853	1.315
mobile > 80 Mbps	.123	.301	.112	.222	.196	1.77	-4.033

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with a random subset that only contains 5 percent of the original observations.

Table B.2.18: Own- and cross-price elasticities (at least 10 tests per municipality)

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-3.075	1.224	.966	1.237	.28	.714	.095
≤ 40 Mbps	1.318	-4.588	1.016	1.243	.308	.823	.194
≤ 80 Mbps	1.037	.994	-5.607	1.788	.259	.672	.186
> 80 Mbps	.971	.897	1.338	-7.155	.25	.715	.231
mobile ≤ 20 Mbps	1.012	1.019	.869	1.12	-3.747	1.218	.775
mobile ≤ 80 Mbps	.665	.712	.597	.858	.314	-4.288	1.574
mobile > 80 Mbps	.113	.215	.213	.345	.262	2.053	-4.685

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with a subset containing only municipalities were at least 10 tests were performed.

Table B.2.19: Own- and cross-price elasticities (at least 20 tests per municipality)

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-3.184	1.246	1.023	1.287	.286	.724	.104
≤ 40 Mbps	1.314	-4.69	1.064	1.291	.314	.841	.212
≤ 80 Mbps	1.04	1.009	-5.68	1.83	.263	.681	.197
> 80 Mbps	.969	.913	1.382	-7.214	.253	.714	.239
mobile ≤ 20 Mbps	1.017	1.044	.917	1.165	-3.825	1.224	.782
mobile ≤ 80 Mbps	.67	.739	.635	.892	.319	-4.394	1.584
mobile > 80 Mbps	.124	.238	.238	.372	.268	2.069	-4.841

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with a subset containing only municipalities were at least 20 tests were performed.

Table B.2.20: Own- and cross-price elasticities (at least 50 tests per municipality)

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-3.34	1.252	1.099	1.431	.287	.719	.119
≤ 40 Mbps	1.284	-4.772	1.13	1.414	.315	.82	.23
≤ 80 Mbps	1.014	1	-5.682	1.94	.259	.674	.218
> 80 Mbps	.963	.919	1.44	-7.205	.252	.695	.255
mobile ≤ 20 Mbps	1.003	1.062	.976	1.28	-3.919	1.204	.816
mobile ≤ 80 Mbps	.687	.764	.715	1.001	.33	-4.585	1.546
mobile > 80 Mbps	.146	.273	.296	.458	.292	2.007	-5.068

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with a subset containing only municipalities were at least 50 tests were performed.

Table B.2.21: Own- and cross-price elasticities (less than 1000 tests per municipality)

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-3.073	1.249	1.031	1.152	.282	.733	.069
≤ 40 Mbps	1.359	-4.767	1.137	1.192	.316	.871	.188
≤ 80 Mbps	1.077	1.069	-5.716	1.671	.263	.699	.166
> 80 Mbps	1.051	.99	1.504	-7.571	.244	.677	.12
mobile ≤ 20 Mbps	1.049	1.067	.939	.974	-3.88	1.311	.834
mobile ≤ 80 Mbps	.676	.738	.631	.695	.324	-4.358	1.711
mobile > 80 Mbps	.075	.191	.18	.143	.254	2.089	-4.372

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with a subset containing only municipalities were less than 1000 tests were performed.

Table B.2.22: Own- and cross-price elasticities (less than 1500 tests per municipality)

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-2.853	1.162	.895	1.151	.26	.666	.076
≤ 40 Mbps	1.272	-4.362	.96	1.168	.286	.783	.181
≤ 80 Mbps	.979	.941	-5.335	1.714	.237	.642	.18
> 80 Mbps	.917	.841	1.279	-6.812	.227	.68	.222
mobile ≤ 20 Mbps	.974	.968	.812	1.044	-3.564	1.167	.742
mobile ≤ 80 Mbps	.63	.676	.565	.814	.294	-4.089	1.539
mobile > 80 Mbps	.088	.193	.197	.319	.237	1.932	-4.322

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with a subset containing only municipalities were less than 1500 tests were performed.

Table B.2.23: Own- and cross-price elasticities of the reduced sample for Vienna

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-3.796	1.037	1.786	1.974	.223	.565	.117
≤ 40 Mbps	1.054	-5.291	1.847	2.003	.252	.692	.272
≤ 80 Mbps	.887	.89	-5.468	2.364	.229	.576	.246
> 80 Mbps	.747	.74	1.829	-6.449	.203	.532	.287
mobile ≤ 20 Mbps	.83	.916	1.728	1.979	-4.257	.953	.688
mobile ≤ 80 Mbps	.601	.724	1.273	1.524	.271	-5.048	1.289
mobile > 80 Mbps	.18	.414	.805	1.178	.289	1.89	-6.568

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with a subset containing only tests from Vienna.

Table B.2.24: Own- and cross-price elasticities of the reduced sample for all locations but Vienna

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-2.535	1.16	.634	.876	.276	.677	.07
≤ 40 Mbps	1.312	-3.939	.665	.839	.302	.778	.126
≤ 80 Mbps	1.244	1.136	-5.659	.938	.291	.735	.089
> 80 Mbps	1.226	1.038	.68	-7.662	.308	.876	.1
mobile ≤ 20 Mbps	.983	.934	.52	.749	-3.222	1.19	.699
mobile ≤ 80 Mbps	.615	.623	.344	.565	.304	-3.66	1.531
mobile > 80 Mbps	.077	.124	.051	.078	.224	1.909	-3.717

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with a subset containing only outside Vienna.

Table B.2.25: Own- and cross-price elasticities of the reduced sample for all of the cities (Graz, Linz, Salzburg, Innsbruck)

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-2.758	1.206	.419	1.1	.326	.703	.18
≤ 40 Mbps	1.242	-3.88	.461	1.071	.353	.703	.18
≤ 80 Mbps	.981	1.043	-5.506	1.691	.364	.519	.078
> 80 Mbps	1.015	.96	.66	-6.394	.36	.656	.129
mobile ≤ 20 Mbps	.904	.937	.427	1.064	-3.134	1.052	.711
mobile ≤ 80 Mbps	.7	.682	.228	.736	.381	-3.814	1.258
mobile > 80 Mbps	.272	.263	.052	.212	.401	1.933	-4.89

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with a subset containing only tests from Graz, Linz, Salzburg, and Innsbruck.

Table B.2.26: Own- and cross-price elasticities of the reduced sample for remaining locations (without Vienna, Graz, Linz, Salzburg, and Innsbruck)

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-2.524	1.184	.692	.773	.268	.712	.047
≤ 40 Mbps	1.386	-4.081	.745	.76	.298	.837	.099
≤ 80 Mbps	1.319	1.195	-5.857	.818	.28	.817	.081
> 80 Mbps	1.333	1.12	.753	-8.295	.291	.903	.054
mobile ≤ 20 Mbps	1.036	.971	.565	.629	-3.357	1.33	.684
mobile ≤ 80 Mbps	.627	.628	.38	.452	.301	-3.676	1.597
mobile > 80 Mbps	.048	.087	.043	.031	.185	1.899	-3.447

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with a subset containing only tests outside Vienna, Graz, Linz, Salzburg, and Innsbruck.

Table B.2.27: Own- and cross-price elasticities of the reduced sample for April and May

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-2.912	1.092	.961	1.287	.275	.621	.127
≤ 40 Mbps	1.348	-4.426	.986	1.236	.307	.669	.173
≤ 80 Mbps	1.056	.853	-5.292	1.695	.275	.598	.152
> 80 Mbps	1.077	.83	1.339	-7.502	.257	.719	.253
mobile ≤ 20 Mbps	1.003	.895	.928	1.088	-3.57	1.077	.821
mobile ≤ 80 Mbps	.638	.554	.576	.873	.303	-4.247	1.859
mobile > 80 Mbps	.127	.139	.142	.29	.23	1.844	-4.071

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with a subset containing tests from April and May.

Table B.2.28: Own- and cross-price elasticities of the reduced sample for June and July

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-3.465	1.476	1.13	1.334	.307	.769	.068
≤ 40 Mbps	1.535	-5.354	1.156	1.358	.384	1.013	.219
≤ 80 Mbps	1.214	1.174	-6.499	2.124	.281	.737	.191
> 80 Mbps	1.093	1.053	1.628	-8.549	.279	.896	.332
mobile ≤ 20 Mbps	1.184	1.402	.996	1.265	-4.452	1.371	.869
mobile ≤ 80 Mbps	.75	.953	.684	1.09	.35	-5.058	1.779
mobile > 80 Mbps	.078	.243	.208	.462	.269	2.128	-4.903

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with a subset containing tests from June and July.

Table B.2.29: Own- and cross-price elasticities of the reduced sample for August and September

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-3.088	1.318	.935	1.079	.292	.73	.103
≤ 40 Mbps	1.48	-4.736	1.001	1.061	.312	.857	.211
≤ 80 Mbps	1.179	1.092	-5.879	1.738	.257	.668	.15
> 80 Mbps	1.102	.953	1.454	-7.44	.249	.675	.154
mobile ≤ 20 Mbps	1.197	1.12	.836	.969	-3.947	1.284	.763
mobile ≤ 80 Mbps	.783	.814	.582	.717	.335	-4.416	1.488
mobile > 80 Mbps	.16	.293	.193	.233	.298	2.215	-5.1

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with a subset containing tests from August and September.

Table B.2.30: Own- and cross-price elasticities of the reduced sample for October and November

	≤ 20 Mbps	≤ 40 Mbps	≤ 80 Mbps	> 80 Mbps	mobile ≤ 20 Mbps	mobile ≤ 80 Mbps	mobile > 80 Mbps
≤ 20 Mbps	-2.683	1.025	.797	1.105	.268	.68	.079
≤ 40 Mbps	1.074	-3.99	.875	1.169	.274	.747	.17
≤ 80 Mbps	.836	.862	-4.944	1.584	.243	.638	.198
> 80 Mbps	.766	.767	1.068	-5.93	.227	.623	.195
mobile ≤ 20 Mbps	.835	.807	.719	1.006	-3.168	1.11	.631
mobile ≤ 80 Mbps	.562	.586	.51	.753	.293	-3.683	1.333
mobile > 80 Mbps	.086	.177	.21	.304	.225	1.795	-4.064

Notes: The own- and cross-price elasticities are simulated 100 times with the regression coefficients of the reduced sample which was estimated with a subset containing tests from October and November.

Table B.2.31: Heterogeneous effects across covariates

alternative	fixed-line $\leq 40\ Mbps$	$\leq 80\ Mbps$	>80 Mbpps	mobile $\leq 20\ Mbps$	$\leq 80\ Mbps$	>80 Mbpps
mobile upload	0.0452***	0.0503***	0.0684***	0.0782***	0.141***	0.294***
	(0.000549)	(0.000597)	(0.000652)	(0.000511)	(0.000539)	(0.000924)
first zip code digit=2	-0.190***	-0.513***	-2.118***	-0.862***	-0.846***	-0.0579
	(0.0137)	(0.0171)	(0.0264)	(0.0201)	(0.0172)	(0.0331)
first zip code digit=3	-0.174***	-1.856***	-3.955***	-0.865***	-1.376***	-0.762***
	(0.0156)	(0.0216)	(0.0293)	(0.0225)	(0.0201)	(0.0421)
first zip code digit=4	-0.298***	-2.730***	-1.765***	-0.571***	-0.708***	-0.193***
	(0.0150)	(0.0255)	(0.0316)	(0.0199)	(0.0174)	(0.0386)
first zip code digit=5	0.562***	-1.348***	-3.814***	0.281***	-0.341***	-0.228***
	(0.0147)	(0.0179)	(0.0258)	(0.0186)	(0.0172)	(0.0395)
first zip code digit=6	0.128***	-1.056***	-0.865***	-0.0476*	-0.0757***	0.169***
	(0.0153)	(0.0208)	(0.0259)	(0.0195)	(0.0173)	(0.0395)
first zip code digit=7	-0.863***	-2.455***	-5.250***	-0.696***	-1.304***	-1.594***
	(0.0218)	(0.0313)	(0.0547)	(0.0288)	(0.0253)	(0.0564)
first zip code digit=8	-0.380***	-2.253***	-2.292***	-0.191***	-0.755***	-0.453***
	(0.0164)	(0.0200)	(0.0289)	(0.0208)	(0.0191)	(0.0414)
first zip code digit=9	0.602***	-0.207***	-0.0323	-0.0314	0.483***	1.772***
	(0.0219)	(0.0244)	(0.0359)	(0.0316)	(0.0250)	(0.0572)
wireless LAN	0.0325***	0.0547***	-0.0516***	-0.695***	-0.632***	-0.0803***
	(0.00706)	(0.00752)	(0.0107)	(0.00867)	(0.00759)	(0.0191)
population=low	-0.0901***	0.137***	-0.848***	-0.120***	-0.0620***	-0.299***
	(0.0110)	(0.0152)	(0.0232)	(0.0147)	(0.0122)	(0.0255)
population=medium	0.381***	-0.0471***	-0.951***	0.0215	0.280***	0.510***
	(0.0106)	(0.0110)	(0.0154)	(0.0137)	(0.0127)	(0.0248)
population=high	-0.420***	-0.211***	0.268***	-0.110***	-0.426***	-1.562***
	(0.0155)	(0.0181)	(0.0240)	(0.0205)	(0.0184)	(0.0378)
newborns	-3.284***	-3.318***	0.767***	0.184	-1.782***	-5.587***
	(0.104)	(0.145)	(0.205)	(0.145)	(0.122)	(0.266)
male	22.67***	31.46***	65.00***	17.63***	21.73***	15.36***
	(0.495)	(0.614)	(0.901)	(0.678)	(0.576)	(1.271)
education=low	2.079**	25.40***	-26.60***	-0.128	5.300***	25.70***
	(0.671)	(0.804)	(1.221)	(0.868)	(0.767)	(1.651)
education=medium	2.340***	23.74***	-11.89***	1.310	6.373***	28.73***
	(0.654)	(0.773)	(1.170)	(0.851)	(0.768)	(1.600)
education=high	6.122***	26.31***	-2.352*	2.203**	6.297***	21.30***
	(0.570)	(0.689)	(1.033)	(0.744)	(0.665)	(1.379)
age=young	1.694***	2.298***	25.41***	2.172***	13.49***	23.26***
	(0.329)	(0.422)	(0.599)	(0.456)	(0.397)	(0.818)
age=old	1.723***	8.540***	-40.45***	-0.0169	4.050***	15.08***
	(0.352)	(0.448)	(0.696)	(0.507)	(0.410)	(0.914)
	(0.700)	(0.838)	(1.288)	(0.924)	(0.787)	(1.711)
constant	-9.274***	-5.960***	-38.62***	-9.303***	-13.81***	-15.26***
	(0.383)	(0.505)	(0.731)	(0.539)	(0.439)	(0.960)

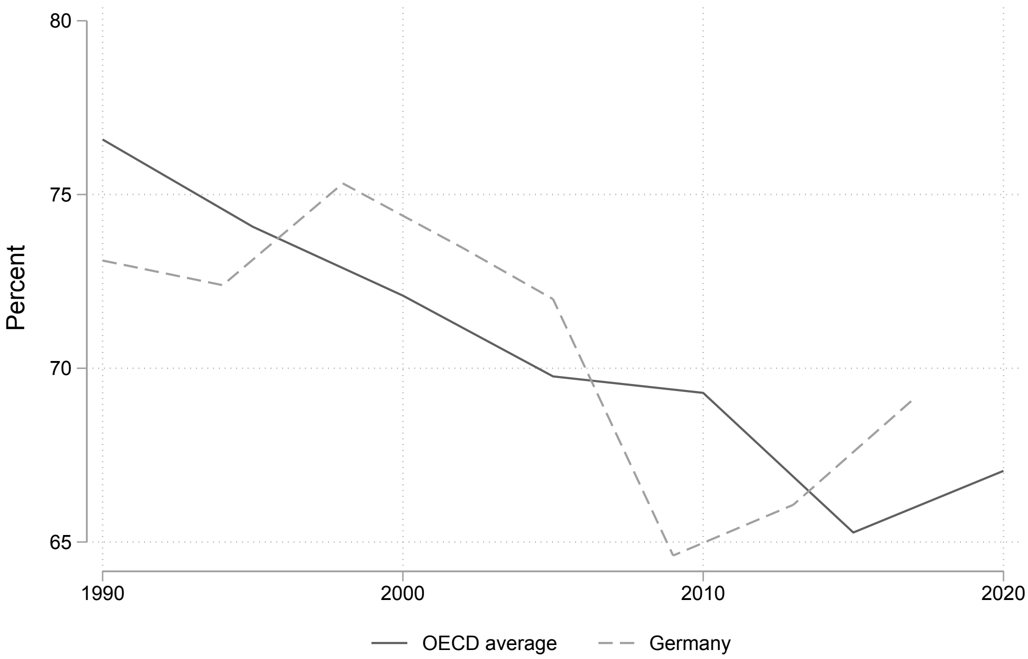
Notes: The regression shows the results for the reduced sample with four fixed-line and three mobile alternatives. The sample was reduced to test which have had full fixed-line coverage (at least 100 Mbps) in 2015 to make sure that all choices are eligible and that the results are comparable. It was control for whether the device, from which the test was performed, was connected via WLAN or LAN. Furthermore, the first digit of the postal code and the mobile upload coverage from the *Broadband Atlas* were applied as geographic covariates. Moreover, characteristics on zip code level were included. It was controlled for population (less than 10,000, less than 100,000, at least 100,000 in 2014, compared to less than 5,000), number of newborn, share of males, education level (compulsory, secondary and tertiary) and age distribution (young is defined as not older than 20 and old is defined as at least 60 years old). Standard errors are shown in parentheses. * p<0.05, ** p<0.01, *** p<0.001

Appendix C

Appendix to Chapter 3

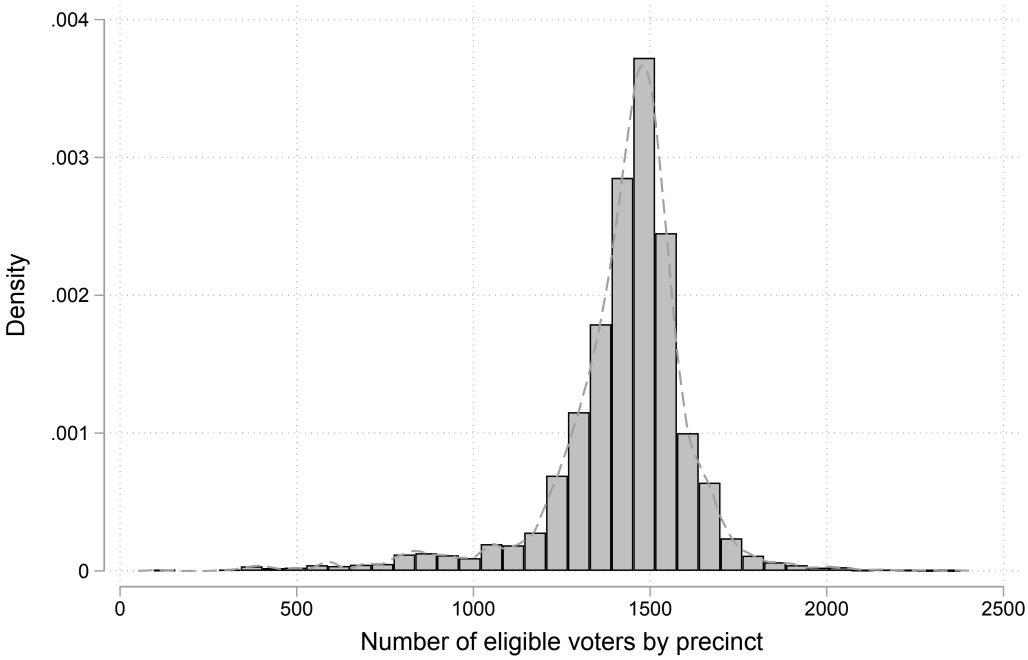
C.1 Additional Figures

Figure C.1.1: Voter turnout in the OECD and Germany



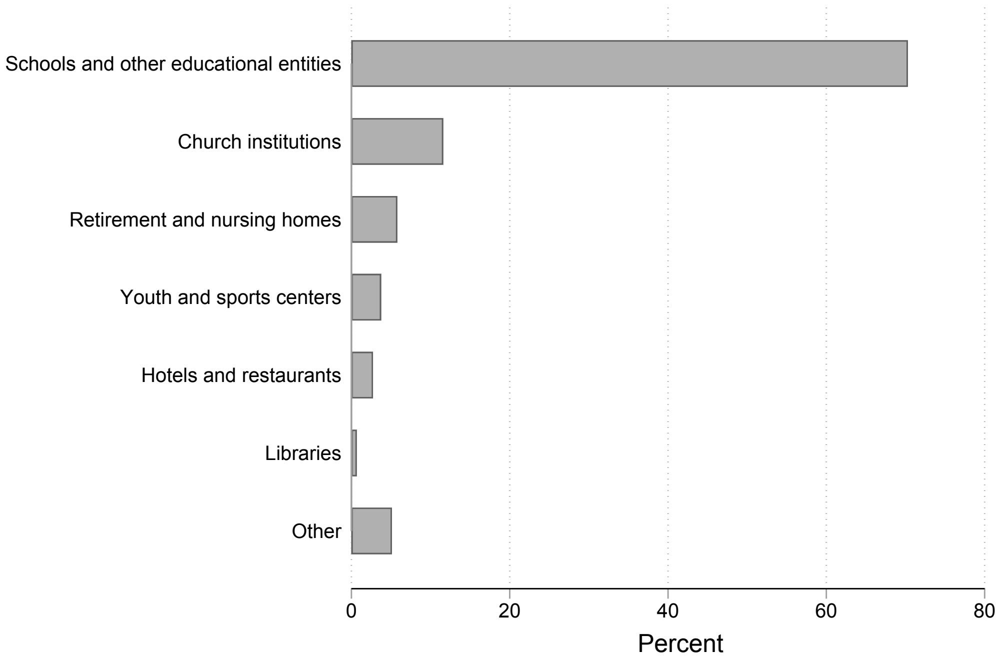
Notes: The figure plots voter turnout in Federal Elections in Germany and average voter turnout in national elections across OECD countries (5-year average). Data are from the International Institute for Democracy and Electoral Assistance.

Figure C.1.2: Distribution of precinct sizes



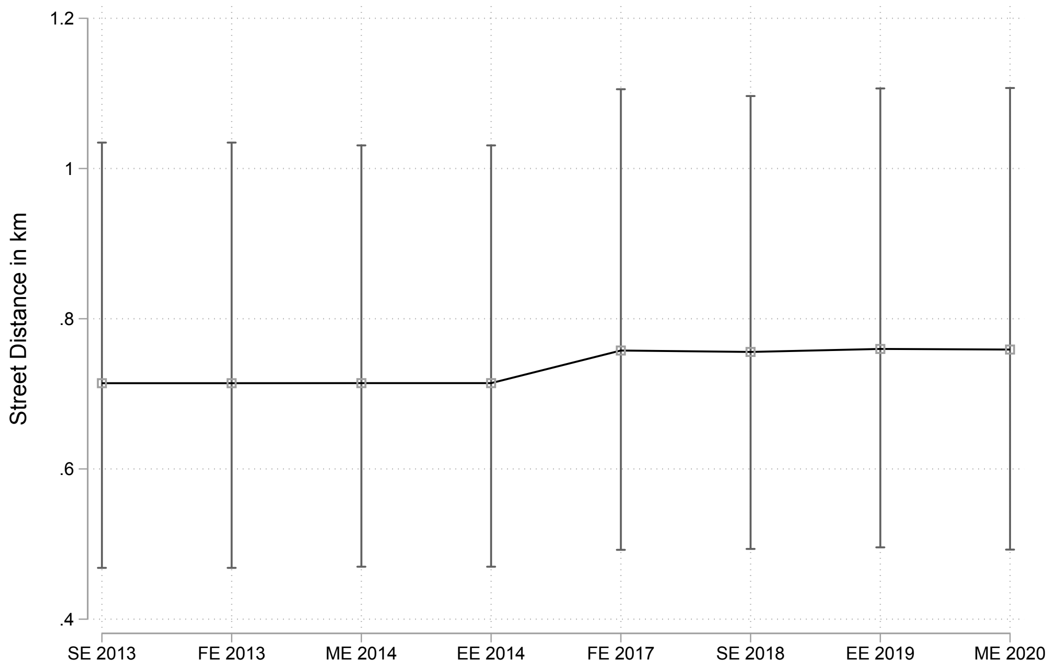
Notes: The figure plots the density precinct of sizes (number of eligible voters) over all elections. Precincts are delineated according to their election-specific boundaries (i.e., before harmonization of precinct borders).

Figure C.1.3: Types of polling venues



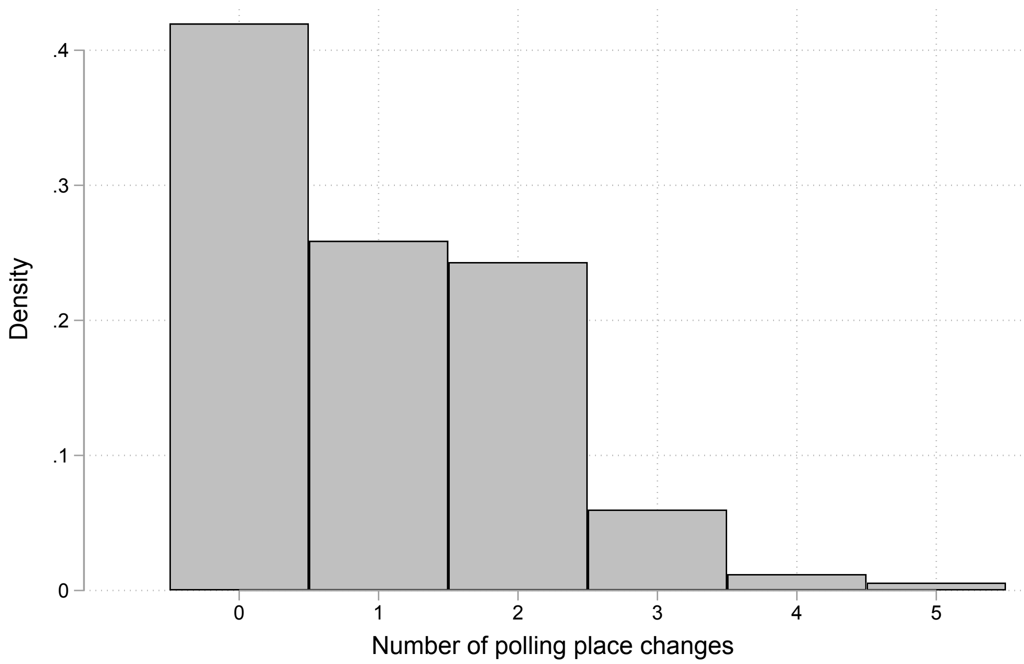
Notes: The figure depicts the frequency of types of polling places venues over the eight elections held in Munich between 2013 and 2020 (293 distinct venues in total).

Figure C.1.4: Median and interquartile range of distance to the polling place



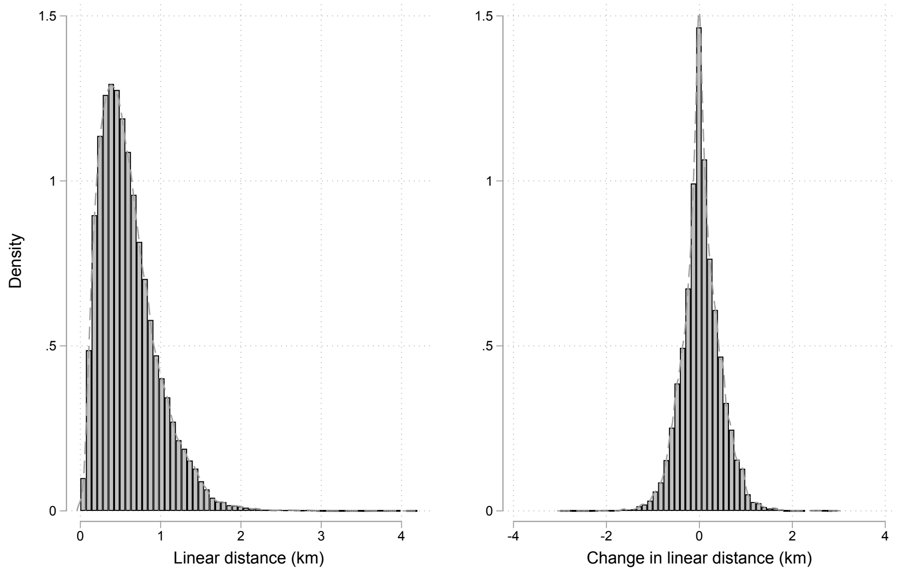
Notes: The figure plots the median and interquartile range (75th and 25th percentile) of the street distance between residential addresses of eligible voters and their designated polling place in each election between 2013 and 2020. SE = State Election, FE = Federal Election, ME = Municipal Election, EE = European Election.

Figure C.1.5: Distribution of polling place changes per eligible voter



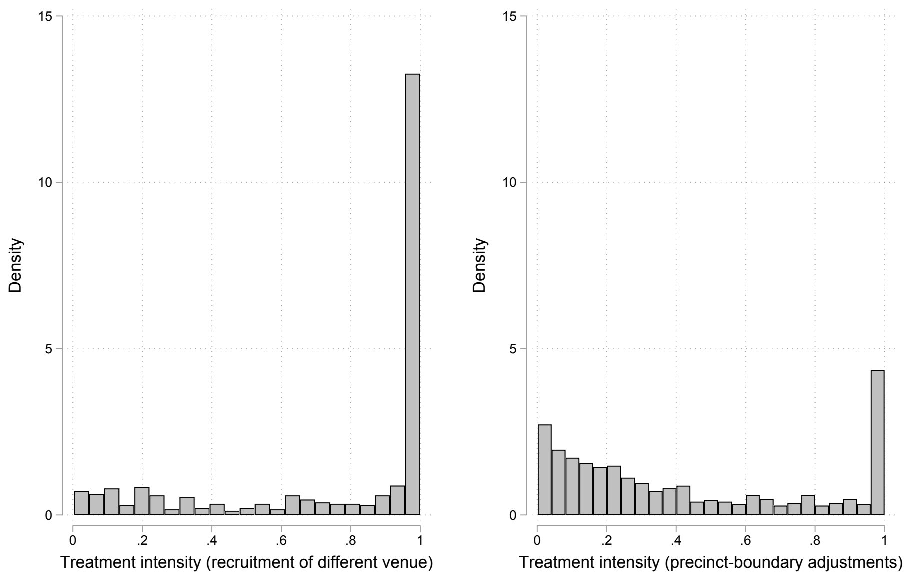
Notes: The figure plots the number of polling places changes per eligible voter.

Figure C.1.6: Density of straight-line distance and distance change to polling place



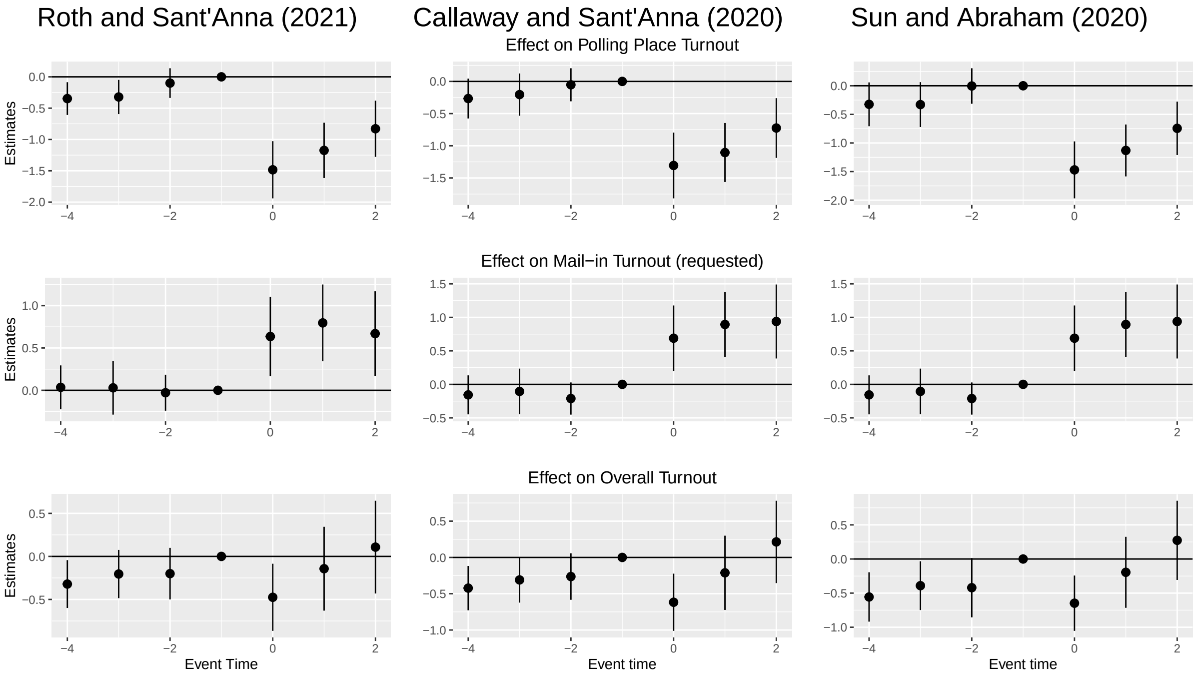
Notes: The figures present density plots for the straight-line (Euclidean) distance between residential addresses of eligible voters and their designated polling place (left plot) and the *change* in distance conditional on reassignment to a new polling place relative to the previous election (right plot) over the eight elections held between 2013 and 2020.

Figure C.1.7: Density of treatment intensity at the precinct level



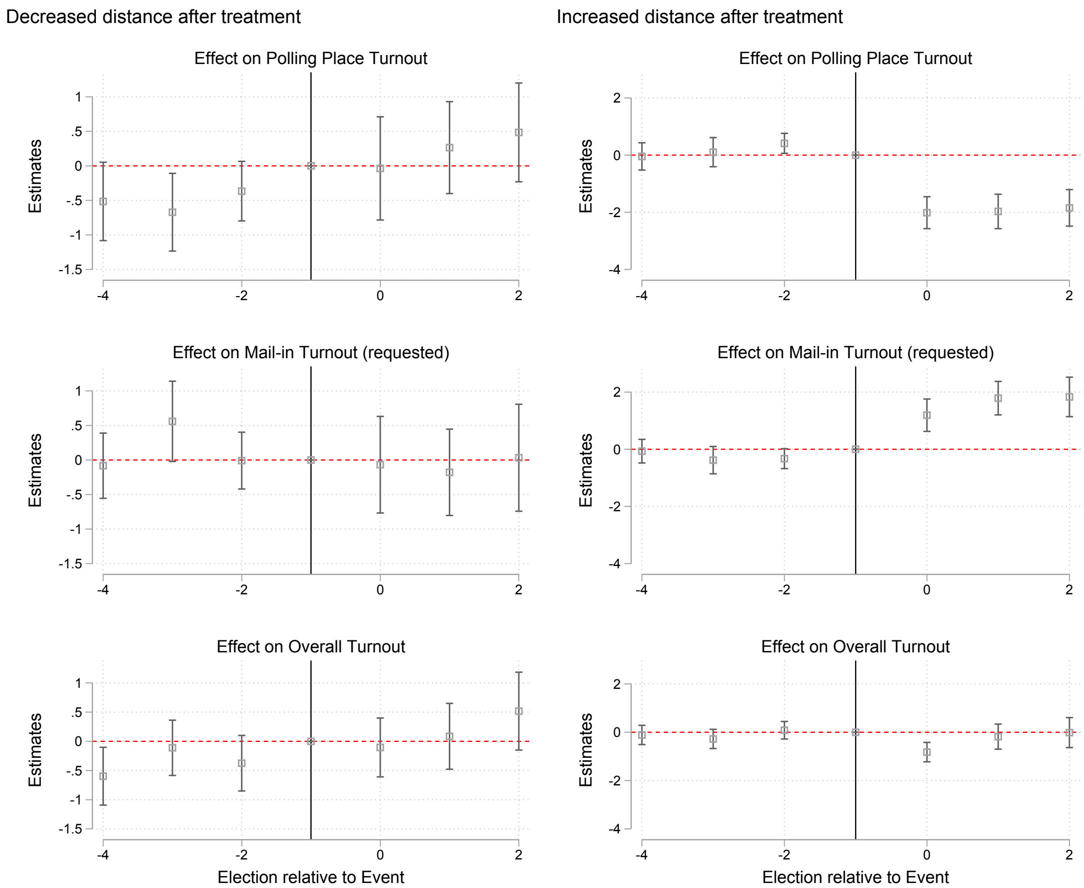
Notes: The figure shows the density of treatment intensity (share of residential addresses reassigned to different polling place) by reason of reassignment. The left panel reports the density for polling place changes due to recruitment of a different venue, the right panel reports the density for changes due to precinct boundary adjustments. Observations with zero reassignments are excluded.

Figure C.1.8: Robustness (event-study illustration)



Notes: The figure presents the event-study results estimate with the suggested procedures by Roth and Sant'Anna [2021*a*], Callaway and Sant'Anna [2021], and Sun and Abraham [2020]. Control variables are not included. Confidence intervals reported at the 95% level. Results replicate the specification of Column (4) in Table 3.3.

Figure C.1.9: Heterogeneity (event-study illustration)



Notes: The figure presents the event-study results from regressing turnout (at the polling place, via mail, and overall, respectively) on a election-date dummies around the event, which is defined as the first time the entire precinct is reassigned to a new polling place (Equation (3.2)). The treated group is further split into two subgroups: (i) precincts, where the new polling place is on average closer and (ii) precincts, where the new polling place is on average further away. Regressions are weighted with the number of eligible voters. Confidence intervals reported at the 95% level.

C.2 Additional Tables

Table C.2.1: Summary statistics of precinct characteristics

	mean	std. dev.	min	p25	median	p75	max
outcome variables							
polling place turnout	34.24	9.04	9.94	26.18	35.54	41.70	55.86
mail-in turnout (requested)	28.92	7.64	4.01	23.10	29.46	34.70	51.99
overall turnout (requested)	63.15	14.57	15.10	51.20	65.27	75.26	91.72
Variables of interest							
avg. linear distance to the polling place (km)	0.52	0.27	0.11	0.32	0.46	0.64	2.19
avg. street distance to the polling place (km)	0.71	0.34	0.16	0.47	0.63	0.87	2.57
share of reassigned residential addresses	0.14	0.32	0.00	0.00	0.00	0.00	1.00
reassigned (precinct boundary adjustments)	0.05	0.19	0.00	0.00	0.00	0.00	1.00
reassigned (recruitment of polling place venue)	0.09	0.27	0.00	0.00	0.00	0.00	1.00
other precinct characteristics							
number of residents	2,428	403	758	2,169	2,325	2,591	6,272
% residents eligible to vote	65.35	9.15	24.62	60.22	66.42	71.70	86.93
% non-native German residents	14.68	4.35	5.50	11.70	13.48	16.45	35.78
% native German residents	59.77	11.35	21.00	52.75	61.80	68.11	83.97
% EU foreigners	12.90	3.97	4.00	10.13	12.38	14.99	36.05
% non-EU foreigners	12.66	6.18	1.91	7.97	11.49	16.06	50.82
% single residents	49.73	7.34	35.28	43.72	48.84	55.02	80.20
% married residents	37.29	6.49	15.50	32.28	37.43	42.77	51.84
% electorate aged 18-24	8.74	2.87	2.41	7.20	8.25	9.64	49.07
% electorate aged 25-34	21.15	6.57	7.40	15.73	20.83	26.01	42.30
% electorate aged 35-44	17.92	4.00	6.30	15.23	17.37	20.08	34.70
% electorate aged 45-59	24.62	3.97	4.85	21.97	24.40	27.25	45.32
% electorate aged 60+	27.57	8.39	2.61	21.30	27.57	33.29	63.80
% Germans in the electorate	91.71	9.13	53.61	84.19	97.30	100	100
% EU-foreigners in the electorate	8.29	9.13	0.00	0.00	2.70	15.81	46.39
% households w/ children	17.53	6.08	5.31	13.35	16.69	20.43	58.75
average duration of residence	21.69	4.45	6.80	18.53	21.72	24.51	45.11
average quoted rent per sqm	17.42	4.54	6.69	13.67	16.45	20.30	43.92
area in sqkm	0.50	0.85	0.06	0.16	0.29	0.49	10.69

Notes: The table reports summary statistics based on 4,944 observations (618 precincts with harmonized boundaries observed over eight elections held between 2013 and 2020).

Table C.2.2: Robustness to alternative distance measures

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Turnout at the Polling Place						
reassigned	-0.46***	-0.46***	-0.45***	-0.50***	-0.51***	-0.49***
	(0.11)	(0.12)	(0.11)	(0.11)	(0.12)	(0.11)
log street distance	-3.44***					
	(0.23)					
street distance		-4.38***	-7.49***			
		(0.33)	(0.88)			
street distance (squared)			1.60***			
			(0.43)			
log linear distance				-3.15***		
				(0.21)		
linear distance					-5.39***	-8.88***
					(0.39)	(1.05)
linear distance (squared)						2.35***
						(0.66)
R^2	0.97	0.97	0.97	0.97	0.97	0.97
Panel B: Turnout Postal (requested)						
reassigned	0.08	0.07	0.07	0.11	0.12	0.11
	(0.12)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)
log street distance	2.41***					
	(0.24)					
street distance		3.27***	4.36***			
		(0.32)	(1.06)			
street distance (squared)			-0.56			
			(0.55)			
log linear distance				2.08***		
				(0.23)		
linear distance					3.74***	5.15***
					(0.40)	(1.37)
linear distance (squared)						-0.96
						(0.91)
R^2	0.95	0.95	0.95	0.95	0.95	0.95
Panel C: Overall Turnout						
reassigned	-0.38***	-0.39***	-0.39***	-0.38***	-0.39***	-0.38***
	(0.12)	(0.12)	(0.12)	(0.11)	(0.11)	(0.12)
log street distance	-1.03***					
	(0.20)					
street distance		-1.11***	-3.13***			
		(0.27)	(0.85)			
street distance (squared)			1.04**			
			(0.42)			
log linear distance				-1.07***		
				(0.18)		
linear distance					-1.66***	-3.73***
					(0.34)	(0.99)
linear distance (squared)						1.39**
						(0.65)
R^2	0.99	0.99	0.99	0.99	0.99	0.99
observations	4,944	4,944	4,944	4,944	4,944	4,944

Notes: Dependent variables are the percentage voter turnout at the polling place (Panel A), by mail (Panel B), and overall (Panel C). Mail-in voting is approximated by the number of requests for of polling cards (*Wahlscheine*). All specifications include the lag terms of *reassigned* and the respective distance variable and include the following controls: log of the number of residents, the share of residents eligible to vote, the share of eligible voters aged 18-24, 25-34, 35-44, 45-59, respectively, the share of EU-foreigners in the electorate, the share of native German residents, the share of non-native German resident, the share of single residents, the share of married residents, the average duration of residence (in years), the share of households with children, and the average quoted rent per square meter. Regressions are weighted with the number of eligible voters. Standard errors are clustered at the precinct level and reported in parentheses.

C.3 Elections in Munich

Federal Elections The German *Bundestag* is elected by German citizens over the age of eighteen for a four-year term. Elections are based on a mixed-member proportional representation system, in which half of the members of parliament are elected directly in 299 constituencies (*Wahlkreise*), four of which are located in Munich, and the other half is elected via (closed) party lists in the sixteen states. Accordingly, voters cast one vote for their local representative, who is elected by a plurality rule, and a second vote for a party list, drawn up by the respective party caucus. Each constituency is represented by one seat in the *Bundestag*, with the remaining seats being allocated based on the second votes to achieve proportionality.

Bavarian State Elections Similar to the federal parliament, the Bavarian *Landtag* is elected for a five-year term on the basis of to mixed-member proportional representation. German citizens of legal age elect the representatives of their constituencies (*Stimmkreise*) and vote for an (open) party list. In contrast to the federal parliament, the allocation of seats in the state parliament takes into account the parties' aggregate first (constituency) votes as well as their second (party-list) votes. The number of single-member constituencies in Munich increased from eight to nine in 2018 due to stronger population growth in Munich compared to the rest of the state.

Munich City Council Elections Municipal Elections in Munich comprise three distinct elections which are held on the same day every six years: the election of the local district committees (*Bezirksausschuss)*, charged with representing the interests of citizens living in 25 distinct city districts in Munich, the mayor's race, which is decided based on an absolute majority rule in a direct election, and the election of the city council (*Stadtrat*), which consists of 80 members elected based on (open) party lists and the mayor as the chairperson. In addition to German citizens of legal age, EU-foreigners are also eligible to vote in municipal elections.

European Elections The European Parliament is elected for a five-year term based on proportional representation. In Germany, each voter casts a single vote for a (closed) list of candidates nominated by a party. All Germans of legal age are eligible to vote in the European Election. It is also possible for non-German EU citizens living in Munich to vote in the city but they have to lodge a request for registration on the electoral roll before each election.

List of Figures

List of Tables

Bibliography

Agyeman, O. T. [2007], 'Survey of ict and education in africa: Benin country report'.

Aker, J. C. and Mbiti, I. M. [2010], 'Mobile Phones and Economic Development in Africa', *Journal of Economic Perspective* **24**(3), 207–232.

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

Ali, S. N. and Lin, C. [2013], 'Why People Vote: Ethical Motives and Social Incentives', *American Economic Journal: Microeconomics* **5**(2), 73–98.

Amos, B., Smith, D. A. and Ste. Claire, C. [2017], 'Reprecincting and Voting Behavior', *Political Behavior* **39**(1), 133–156.

Atasoy, H. [2013], 'The effects of broadband internet expansion on labor market outcomes', *Industrial and Labor Relations Review* **66**(2), 315–345.

Athey, S. and Imbens, G. W. [2021], 'Design-based analysis in Difference-In-Differences settings with staggered adoption', *Journal of Econometrics* .

Bai, Y. [2017], 'The faster, the better? the impact of internet speed on employment', *Information Economics and Policy* **40**, 21–25.

Banerjee, A., Duflo, E. and Qian, N. [2020], 'On the road: Access to transportation infrastructure and economic growth in China', *Journal of Development Economics* **145**(January), 102442.

Bechtel, M. M., Hangartner, D. and Schmid, L. [2018], 'Compulsory Voting, Habit Formation, and Political Participation', *The Review of Economics and Statistics* **100**(3), 467–476.

BenYishay, A., Rotberg, R., Wells, J., Lv, Z., Goodman, S., Kovacevic, L. and Runfola, D. [2017], 'Geocoding afrobarometer rounds 1-6: Methodology & data quality', *AidData. Available online at http://geo. aiddata. org* .

Bertschek, I., Cerquera, D. and Klein, G. J. [2013], 'More bits–more bucks? measuring the impact of broadband internet on firm performance', *Information Economics and Policy* **25**(3), 190–203.

Bhatti, Y. [2012], 'Distance and Voting: Evidence from Danish Municipalities: Distance and Voting', *Scandinavian Political Studies* **35**(2), 141–158.

Bloomberg [2020], 'Faster internet coming to africa with facebook's $1 billion cable'.

Bluhm, R. and Krause, M. [2018], 'Top lights - bright cities and their contribution to economic development', *CESifo Working Paper* **No. 7411**.

Borusyak, K., Jaravel, X. and Spiess, J. [2021], 'Revisiting Event Study Designs: Robust and Efficienct Estimation', *Work in Progress* pp. 1–48.

Brady, H. E. and McNulty, J. E. [2011], 'Turning Out to Vote: The Costs of Finding and Getting to the Polling Place', *American Political Science Review* **105**(1), 115–134.

Briglauer, W., Dürr, N. S., Falck, O. and Hüschelrath, K. [2019], 'Does state aid for broadband deployment in rural areas close the digital and economic divide?', *Information Economics and Policy* **46**, 68–85.

Brody, R. A. and Sniderman, P. M. [1977], 'From Life Space to Polling Place: The Relevance of Personal Concerns for Voting Behavior', *British Journal of Political Science* **7**(3), 337–360.

Bruederle, A. and Hodler, R. [2018], 'Nighttime lights as a proxy for human development at the local level', *PLoS ONE* **13**(9), 1–22.

Cairncross, F. [1997], *The death of distance: how the communications revolution will change our lives*, number C20-21, Harvard Business School.

Callaway, B. and Sant'Anna, P. H. [2021], 'Difference-in-differences with multiple time periods', *Journal of Econometrics* **225**(2), 200–230. Themed Issue: Treatment Effect 1.

Cantoni, E. [2020], 'A Precinct Too Far: Turnout and Voting Costs', *American Economic Journal: Applied Economics* **12**(1), 61–85.

Cardona, M., Schwarz, A., Yurtoglu, B. B. and Zulehner, C. [2009], 'Demand estimation and market definition for broadband Internet services', *Journal of Regulatory Economics* **35**(1), 70–95.

Cariolle, J. [2021], 'International connectivity and the digital divide in Sub-Saharan Africa', *Information Economics and Policy* **55**, 100901.

Chabossou, A. F. [2007], '2007 benin telecommunications sector performance review – a supply side analysis of policy outcomes'.

Chavula, J., Feamster, N., Bagula, A. and Suleman, H. [2015], Quantifying the effects of circuitous routes on the latency of intra-africa internet traffic: A study of research and education networks, *in* A. Nungu, B. Pehrson and J. Sansa-Otim, eds, 'e-Infrastructure and e-Services for Developing Countries', Springer International Publishing, Cham, pp. 64–73.

Chen, X. and Nordhaus, W. D. [2011], 'Using luminosity data as a proxy for economic statistics', *Proceedings of the National Academy of Sciences of the United States of America* **108**(21), 8589–8594.

Christiaensen, L. and Kanbur, R. [2017], 'Secondary Towns and Poverty Reduction: Refocusing the Urbanization Agenda', *Annual Review of Resource Economics* **9**(November), 405–

419.

Christiaensen, L. and Todo, Y. [2014], 'Poverty reduction during the rural-urban transformation - The role of the missing middle', *World Development* **63**, 43–58.

Colombo, M. G., Croce, A. and Grilli, L. [2013], 'Ict services and small businesses' productivity gains: An analysis of the adoption of broadband internet technology', *Information Economics and Policy* **25**(3), 171–189.

Conley, T. G. [1999], 'GMM estimation with cross sectional dependence', *Journal of Econometrics* **92**(1), 1–45.

Crandall, R. W., Sidak, G. J. and Singer, H. J. [2002], 'The Empirical Case Against Asymmetric Regulation of Broadband Internet Access', *Berkley Technology Law Journal* **17**(1), 953–987.

Czernich, N. [2014], 'Does broadband internet reduce the unemployment rate? Evidence for Germany', *Information Economics and Policy* **29**, 32–45.

Czernich, N., Falck, O., Kretschmer, T. and Woessmann, L. [2011], 'Broadband Infrastructure and Economic Growth', *Economic Journal* **121**(552), 505–532.

de Chaisemartin, C. and D'Haultfœuille, X. [2020], 'Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects', *American Economic Review* **110**(9), 2964–2996.

Dinkelman, T. [2011], 'The effects of rural electrification on employment: New evidence from South Africa', *American Economic Review* **101**(7), 3078–3108.

Donaldson, D. [2018], 'Railroads of the Raj: Estimating the impact of transportation infrastructure', *American Economic Review* **108**(4-5), 899–934.

Downs, A. [1957], 'An Economic Theory of Political Action in a Democracy', *Journal of Political Economy* **65**(2), 135–150.

Dyck, J. J. and Gimpel, J. G. [2005], 'Distance, Turnout, and the Convenience of Voting', *Social Science Quarterly* **86**(3), 531–548.

Faber, B. [2014], 'Trade integration, market size, and industrialization: Evidence from China's national trunk highway system', *Review of Economic Studies* **81**(3), 1046–1070.

Fauvelle-Aymar, C. and François, A. [2018], 'Place of registration and place of residence: The non-linear detrimental impact of transportation cost on electoral participation', *Public Choice* **176**(3), 405–440.

Feddersen, T. J. [2004], 'Rational Choice Theory and the Paradox of Not Voting', *Journal of Economic Perspectives* **18**(1), 99–112.

Fetzer, T., Henderson, V., Nigmatulina, D. and Shanghavi, A. [2016], 'What happens to cities when countries become democratic?', (8005077).

Flamm, K. and Chaudhuri, A. [2007], 'An analysis of the determinants of broadband access', *Telecommunications Policy* **31**(6-7), 312–326.

Ford, G. S. [2018], 'Is faster better? quantifying the relationship between broadband speed and economic growth', *Telecommunications Policy* **42**(9), 766–777.

Fujiwara, T., Meng, K. and Vogl, T. [2016], 'Habit Formation in Voting: Evidence from Rainy Elections', *American Economic Journal: Applied Economics* **8**(4), 160–188.

Funk, P. [2010], 'Social Incentives and Voter Turnout: Evidence from the Swiss Mail Ballot System', *Journal of the European Economic Association* **8**(5), 1077–1103.

Gaebler, S., Potrafke, N. and Roesel, F. [2020], 'Compulsory voting and political participation: Empirical evidence from Austria', *Regional Science and Urban Economics* **81**, 103499.

Gerber, A. S., Green, D. P. and Shachar, R. [2003], 'Voting May Be Habit-Forming: Evidence from a Randomized Field Experiment', *American Journal of Political Science* **47**(3), 540–550.

Ghani, E., Goswami, A. G. and Kerr, W. R. [2016], 'Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing', *Economic Journal* **126**(591), 317–357.

Gibson, J., Kim, B., Stillman, S. and Boe-Gibson, G. [2013], 'Time to vote?', *Public Choice* **156**(3-4), 517–536.

Gimpel, J. and Schuknecht, J. [2003], 'Political participation and the accessibility of the ballot box', *Political Geography* **22**(5), 471–488.

Goldfarb, A. and Tucker, C. [2019], 'Digital economics', *Journal of Economic Literature* **57**(1), 3–43.

Goodman-Bacon, A. [2021], 'Difference-in-differences with variation in treatment timing', *Journal of Econometrics* **225**(2), 254–277. Themed Issue: Treatment Effect 1.

Green, D. P. and Shachar, R. [2000], 'Habit Formation and Political Behaviour: Evidence of Consuetude in Voter Turnout', *British Journal of Political Science* **30**(4), 561–573.

Grimes, A., Ren, C. and Stevens, P. [2012], 'The need for speed: Impacts of internet connectivity on firm productivity', *Journal of Productivity Analysis* **37**(2), 187–201.

Grogan, L. and Sadanand, A. [2013], 'Rural Electrification and Employment in Poor Countries: Evidence from Nicaragua', *World Development* **43**, 252–265.

Grzybowski, L., Nitsche, R., Verboven, F. and Wiethaus, L. [2014], 'Market definition for broadband internet in Slovakia - Are fixed and mobile technologies in the same market?', *Information Economics and Policy* **28**(1), 39–56.

Hamilton Research [2020], 'Africa bandwidth maps'.

Haspel, M. and Knotts, H. G. [2005], 'Location, Location, Location: Precinct Placement and the Costs of Voting', *The Journal of Politics* **67**(2), 560–573.

Henderson, J. V., Storeygard, A. and Weil, D. N. [2012], 'Measuring economic growth from outer space', *American Economic Review* **102**(2), 994–1028.

Henderson, V., Storeygard, A. and Weil, D. N. [2011], 'A bright idea for measuring economic growth', *American Economic Review* **101**(3), 194–199.

Hjort, J. and Poulsen, J. [2019], 'The arrival of fast internet and employment in Africa', *American Economic Review* **109**(3), 1032–1079.

Hjort, J. and Tian, L. [2021], 'The Economic Impact of Internet Connectivity in Developing Countries', *Working paper* .

Huber, S. and Rust, C. [2016], 'Calculate Travel Time and Distance with Openstreetmap Data Using the Open Source Routing Machine (OSRM)', *The Stata Journal: Promoting communications on statistics and Stata* **16**(2), 416–423.

Invesco [2019], 'China embarks on digital silk road'.

Jensen, M. [2006], 'Open access. lowering the costs of international bandwidth in africa', *APC Issue Papers Series* **October**.

Jensen, R. [2007], 'The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector', *The Quarterly Journal of Economics* **3**(August), 879–924.

Kende, M. and Rose, K. [2015], Promoting Local Content Hosting to Develop the Internet Ecosystem, Technical report, Internet Society.

Kolko, J. [2012], 'Broadband and local growth', *Journal of Urban Economics* **71**(1), 100–113.

Li, X., Zhou, Y., Zhao, M. and Zhao, X. [2020], 'A harmonized global nighttime light dataset 1992â2018', *Scientific Data* **7**(1), 1–9.

McFadden, D. and Train, K. [2000], 'Mixed Mnl Models for Discrete Response', *Journal of Applied Econometrics* **470**(May), 447–470.

McNulty, J. E., Dowling, C. M. and Ariotti, M. H. [2009], 'Driving Saints to Sin: How Increasing the Difficulty of Voting Dissuades Even the Most Motivated Voters', *Political Analysis* **17**(4), 435–455.

Meredith, M. [2009], 'Persistence in Political Participation', *Quarterly Journal of Political Science* **4**(3), 187–209.

Nunn, N. and Puga, D. [2012], 'Ruggedness: The blessing of bad geography in Africa', *Review of Economics and Statistics* **94**(1), 20–36.

Pereira, P. and Ribeiro, T. [2011], 'The impact on broadband access to the Internet of the dual ownership of telephone and cable networks', *International Journal of Industrial Organization* **29**(2), 283–293.

Picard, R. [2019], 'GEODIST: Stata module to compute geographical distances', Boston College Department of Economics.

Plutzer, E. [2002], 'Becoming a Habitual Voter: Inertia, Resources, and Growth in Young Adulthood', *American Political Science Review* **96**(1), 41–56.

Poppe, Y. [2009], 'Subsea cables: Jugular veins for africa's global communications'.

Potrafke, N. and Roesel, F. [2020], 'Opening hours of polling stations and voter turnout: Evidence from a natural experiment', *The Review of International Organizations* **15**(1), 133–163.

Rappoport, P., Kridel, D. J., Taylor, L. D., Duffy-Deno, K. T. and Alleman, J. [2003], Residential demand for access to the Internet, *in* G. Madden, ed., 'International Handbook of Telecommunications Economics', Vol. 2, Edward Elgar, Cheltenham.

Riker, W. H. and Ordeshook, P. C. [1968], 'A Theory of the Calculus of Voting', *American Political Science Review* **62**(1), 25–42.

Roth, J. and Sant'Anna, P. H. [2021*a*], 'Efficient estimation for staggered rollout designs', *arXiv preprint arXiv:2102.01291* .

Roth, J. and Sant'Anna, P. H. C. [2021*b*], 'Staggered: R module for implementation of DiD estimators'.

Rud, J. P. [2012], 'Electricity provision and industrial development: Evidence from India', *Journal of Development Economics* **97**(2), 352–367.

Srinuan, P., Srinuan, C. and Bohlin, E. [2012], 'Fixed and mobile broadband substitution in Sweden', *Telecommunications Policy* **36**(3), 237–251.

Storeygard, A. [2016], 'Farther on down the road: Transport costs, trade and urban growth in sub-Saharan Africa', *Review of Economic Studies* **83**(3), 1263–1295.

Sun, L. and Abraham, S. [2020], 'Estimating dynamic treatment effects in event studies with heterogeneous treatment effects', *Journal of Econometrics* .

Whitacre, B. E., Alam, M. R. and Lobo, B. J. [2018], 'Econometric error nullifies finding of the impact of broadband speed on county-level employment', *Information Economics and Policy* **44**, 58–60.

World Bank [2016], World Development Report 2016: Digital Dividends, Technical report.