Four Empirical Essays on the Effects of New Technologies in Shaping Markets

Andreas Mazat

ifo INSTITUT
Leibniz-Institut für Wirtschaftsforschung
an der Universität München e.V.

München 2018

Four Empirical Essays on the Effects of New Technologies in Shaping Markets

Inaugural-Dissertation
zur Erlangung des Grades Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2018

vorgelegt von
Andreas Daniel Mazat

Referent: Prof. Dr. Oliver Falck
Korreferent: Prof. Dr. Ingo Vogelsang
Promotionsabschlussberatung: 30. Januar 2019

Erstgutachter:	Prof. Dr. Oliver Falck
Zweitgutachter:	Prof. Dr. Ingo Vogelsang
Drittgutachterin:	Prof. Dr. Monika Schnitzer
Datum der mündlichen Prüfung:	14. Januar 2019

Acknowledgments

First and foremost, I would like to thank my supervisor Oliver Falck. His advice, guidance and outstanding support over the past years have been very important both for the completion of my thesis and in my decision to pursue a PhD. I have learned and benefited in many distinct ways through our joint work. Further, I would like to thank Ingo Vogelsang, who has supported my research with ideas and valuable comments. His highest academic standards and his ability to translate research into relevant policy advice has been and keeps inspiring and motivating me. I would also like to thank Monika Schnitzer for completing my dissertation committee.

I am very grateful to Tobias Lohse, who contributed significantly to the completion of my dissertation not only as a coauthor. As my office colleague and friend he made some difficult times less painful and the good times even better. I am indebted to Johannes Koenen. His guidance and his example inspired me to push harder, learn more, and think deeper about economics. I also thank Simone Schüller, whose thoughtful advice and sincere support were instrumental in many moments during my studies.

I am also deeply grateful to my remaining coauthors Alexandra Heimisch and Bastian Stockinger. Working jointly with them has been and continues to be a great source of mutual support and motivation.

Further, I thank my colleagues at ifo Institute and at the University of Munich (LMU) for their support. They were always helpful and made the years of graduate school much nicer.

Throughout my thesis I have benefited from through financial support by Deutsche Telekom AG, which is gratefully acknowledged.

Without my good friends, I would not have been able to achieve this. And more importantly, life would have been much less enjoyable. I am very grateful for having you.

Most of all, I thank my family and my girlfriend for their love and endless support.

Contents

List of Figures

List of Tables

Preface

“Creative Destruction is the essential fact about capitalism“
Joseph A. Schumpeter

Creative Destruction describes a process in which – through recombination of production factors – innovations are replacing existing products, processes, and market structures. According to Joseph A. Schumpeter it is the essential fact about capitalism that ultimately enables technological change and economic growth. According to this theory, policy makers should therefore create an environment that is conducive to the process of creative destruction. Yet, while it may be beneficial for the economy overall in the long term, the gains from technological change and disruption of existing structures are likely heterogeneously distributed and may produce losers, at least in the medium term. This illustrates that policy makers face the challenge to promote change and innovation, and at the same time compensate potential losers to ensure societal unity. Consequently, designing adequate policies requires a strong understanding of how technological and structural change affects the various actors involved. This dissertation aims at contributing to this understanding, analyzing the role of new technologies in shaping markets, competition between existing and new technologies, and in producing winners and losers.

Chapter 1 focuses on the market for voice telephony services, which has been subject to rapid technological change. Not only the rise of mobile telephony but also internet telephony, as well as voice- and video messaging are increasing the set of options available for telephony. This is the starting point of Chapter 1 analyzing the extent of fixed-to-mobile voice telephony substitution across Austria between 2011 and 2014 and its implications for market definition. Chapter 2 turns to the market for broadband internet access. Similar to voice telephony, internet technologies have changed substantially over time, providing ever increasing bandwidth to surf the internet at higher speed. In pursuit of understanding how the emergence of these new technologies – coming along with the entry of new operators – changes the existing structure of the market, Chapter 2 analyzes the role of bandwidth in shaping competition in the fixed-line broadband market between 2014 and 2016. As a general purpose technology and by providing a platform, the internet has generated a myriad of services competing with existing companies and business models. Chapter 3 focuses on one such service that is commonly associated with the decline of an existing industry. It examines the effect of online shopping on the survival of and employment in traditional brick-and-mortar bookstores across Germany between 1999 and 2013. Finally, Chapter 4 returns to an essential of Schumpeter’s theory, the entrepreneur, who by striving for his own economic prosperity creates innovations – such as the internet or mobile phones – and ultimately social welfare.

In particular, Chapter 4 examines whether broadband internet provision might be a suitable policy to promote entrepreneurship and thus the economic gains associated with it.

Throughout all four chapters of this dissertation, I exploit regional differences in the availability, quality, and adoption of new technologies to determine their effect on market change. In such regional studies, an essential complication with regards to identification of causal effects is the fact that regional differences in the availability and take-up of new technologies are determined endogenously. Throughout the dissertation I address this issue in two ways: First, I try to isolate the effect of interest by thoroughly elaborating on and accounting for structural differences between regions that affect both, the availability or adoption of new technologies, as well as market change. Second, I make use of instrumental variables such as environmental factors, pre-existing municipality structures, historic differences in the demographic composition of the population across regions, and technological peculiarities to derive exogenous variation in the provision and adoption of the technologies considered. Since the validity of instrumental variables is generally not testable, I provide placebo estimates whenever possible to substantiate the validity of the effects found. In what follows, 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.

Chapter 1 examines how the emergence of new technologies affects the boundaries of existing markets. The standard textbook approach to define these boundaries in economics is to look at cross-price elasticities. However, determining cross-price elasticities is often difficult, depending on the products or services considered. For example, in telecommunications determining cross-price elasticities is often fraught with difficulties arising from non-linear price schedules or a high degree of price differentiation. This paper aims to address these difficulties by proposing a new approach to quantify substitution between two goods or services. The underlying idea of the approach is that consumers may switch products depending on differences in the quality of the product. We apply the approach of estimating – what we term – a “cross-quality elasticity” to the telecommunications market and focus on quantifying fixed-mobile voice telephony substitution in Austria between 2011 and 2014. Using novel data sets, including user-generated network tests and fine-grained satellite data allows us to estimate causal effects. In particular, we pursue an IV approach that exploits exogenous variation in mobile network quality derived from the local density of trees as well as the proximity of a postal code area to railway tracks. These instruments rely on the fact that a small enough distance and direct visual contact between mobile handset and base station antenna are decisive in determining mobile signal quality. Our results indicate substantial fixed-mobile substitution across Austria during the time period considered. For example, an increase in mobile bandwidth by one standard deviation (approx. 8 Mbps) increases the incidence of incumbent fixed-line customers switching away from fixed-line usage by 1.09 percentage points. This corresponds to 6.81% of the decline in the incumbent fixed-line market penetration between 2011 and 2014. Our analysis further suggests substantial heterogeneous effects across different Austrian regions, with effects being stronger in regions with a low share of an elderly population, a low share of tertiary educated, or a high share of unemployed and foreign individuals. Overall, our findings indicate that differences in the characteristics of regions substantially matter for substitution patterns and thus competition across different local fixed-line voice telephony markets in Austria. We therefore conclude that any optimal regulation would have to take regional characteristics and particularities into account.

Since the liberalization of telecommunications, the fixed-line broadband internet market in Germany has been changing substantially. On the one hand, alternative network operators are expanding their networks and gaining increasing shares of the market. On the other hand, demand for bandwidth is rising. Since alternative cable and fiber network operators can typically provide bandwidth beyond what is achievable via (A)DSL, the German incumbent operator is facing competition in quality that may be increasingly relevant in view of rising demand for bandwidth. Reacting to changing demand, the German incumbent operator has launched a fixed-mobile hybrid product combining bandwidth from its fixed-line DSL and mobile (LTE) networks to surf the internet at higher speed. In Chapter 2, I exploit the introduction of this product to examine the role of bandwidth in shaping competition. I combine comprehensive data on fixed-line market characteristics that includes information on network deployment cots, market size, competition intensity and network quality with information on the local composition of the population across 5200 local exchange areas across Germany between 2011 and 2016. This allows me to thoroughly control for potential confounding factors in estimating the effect of hybrid product coverage on changes in the incumbent fixed-line market penetration. The results suggest that hybrid product introduction increased the annual rate of change in incumbent fixed-line market penetration by 0.86 percentage points, corresponding to about 33.99% of the average annual decline in fixed-line market penetration between 2011 and 2014. Besides, the results suggests that effects are substantially larger among younger population groups than among older population groups. Further, hybrid product introduction seems to have a comparably small impact on market penetration, where infrastructure-based competition is small. A placebo test finding zero effect suggests that the effects estimated reflect the causal effect of hybrid product coverage on incumbent fixed-line market penetration. Overall, the results of the analysis contribute to the debate on how to incentivize private investment in high-capacity broadband networks by showing who (in terms of population groups) and where (in terms of particular regions) the provision of higher bandwidth results in increasing market penetration.

The internet not only provides easy access to information, new forms of communication via e-mail, instant voice- and video-messaging or social media platforms. It gave also rise to e-commerce, which is frequently associated with the demise of brick-and-mortar retail. It even goes as far as policy makers prophesying deserted inner cities. These nightmare visions motivate our analysis in Chapter 3, examining the effect of online shopping on the stock of and employment in traditional brick-and-mortar books retailers across Germany between 1999 and 2013. We focus on the German book sector rather than stationary retail overall since the institutional setting allows us to identify how the availability of a convenient distribution channel affects traditional retailers. In particular, the German book price law fixing the prices of books sold across Germany allows us to abstract from price competition effects. Besides, a negligible market share of self-publishing outlets suggests that the character and variety of books sold is similar offline and online. We combine administrative data on all book selling establishments across Germany with a novel data set providing information on the online-shopping affinity across all German municipalities. To identify causal effects we implement a Bartik-type IV strategy that exploits historic differences in the composition of the population in terms of its age structure to instrument regional differences in online shopping affinity. Our results suggest that regional exposure to online shopping has a robust negative effect on brick-and-mortar bookstores. For example, an increase in online shopping affinity by one standard deviation leads to a reduction in the local stock of bookstores by approximately 0.11 establishments per 10,000 residents. This effect is economically sizable, corresponding to a decline of 14% compared to baseline levels in 1999 (when online shopping was not yet popular across Germany). Further, our results indicate that the effects of online shopping affect bookstores in different type of markets differently. For example, in regions with a high initial density of bookstores an increase in online shopping affinity tends to reduce total employment. In contrast, in regions with a rather low density of bookstores, overall employment remains unchanged but full time-employment is reduced in favor of part-time employment. This likely corresponds to a reduction of hours worked.

Many local policy makers aim at attracting entrepreneurs since they expect them to be innovative, to create jobs, and local economic growth. In Chapter 4, we examine whether and to what extent broadband internet provision may be a suitable policy to attract local entrepreneurial activity. We make use of a comprehensive data set capturing all new establishments that started up their business within a Western German municipality between 1992 and 2009 and combine it with data on local broadband coverage between 2005 and 2009. Since broadband roll-out is generally determined endogenously, simply comparing changes in establishment startup rates over time with changes in broadband coverage across municipalities is likely complicated by endogeneity concerns. To avoid such endogeneity concerns we exploit a technological peculiarity in broadband internet provision via DSL as an instrument to derive causal variation in broadband coverage across German municipalities. In particular the instrument makes use of the fact that after DSL was introduced in 1999, households in some municipalities could not readily adopt the new technology because off the local layout of the pre-existing public switched telephone network. The results of our analysis therefore suggest that an increase in broadband coverage by ten percentage points increases the joint number of new establishments per 100 employees across manufacturing, services, and trade by 0.28. With regards to the average startup rate between 1994 and 1996, this effect corresponds to an increase of 18.42%. We find that the effects differ between sectors and show that the positive overall effect is driven by the trade sector while the startup rate in manufacturing even declines in response to higher broadband coverage. Since placebo estimates generally support the validity of our results, the paper contributes to the provision of causal evidence on the relationship between broadband provision and entrepreneurial activity. We argue that this evidence may be used by policy makers that are interested in contrasting the cost and potential economic benefits of broadband deployment projects.

1 Regulating an Industry Undergoing Technological Change - The Case of Telecom in Austria*

1.1 Introduction

Many regulated industries have undergone or are currently undergoing considerable technological change, often with new types of competitors and new channels of competition arising. The most prominent examples are the telecommunications and the electricity sectors. In such dynamic environments, when should regulators start taking alternative technologies into account as potential substitutes? This question of market definition is essential both for evaluating mergers as well as fundamentally analyzing the case for regulation – and more often than not it is hotly contested. To determine relevant substitutes and competitors, econometricians typically rely on (national level) cross-price elasticities, but this is often fraught with substantial difficulties. Consider the telecommunications sector, where fixed-mobile substitution for voice telephony has received interest both from policy-makers and academics. Here, even the first step – determining the relevant prices – poses multiple challenges: Non-linear price schedules, such as a combination of base price, included volume of minutes plus a per-minute price if that volume is exceeded, make assumptions regarding (the distribution of) usage profiles in the population necessary. Relatedly, a high degree of price differentiation makes price comparisons difficult both for consumers, and for researchers trying to assess effective prices. Any definition of "mobile prices" therefore involves considerable assumptions or simplifications.

In this paper, we address this issue by developing a new approach to assess the substitutability of different technologies that does not rely on observed prices, and apply it to the market for voice telephony. The underlying intuition of our approach is simple: The willingness of an individual to switch away from a given technology should depend on the availability and quality of relevant alternatives or substitutes. We translate this idea into an econometric model that exploits quality variation at the regional level to estimate a "cross-quality" elasticity. If the quality of a given alternative technology positively affects switching behavior, then we have identified it as a substitute. Our approach has a number of important advantages compared to estimating national-level cross-price elasticities. Most importantly, it contributes to detecting and explaining differences in substitution behavior across regions. Without observing cross-sectional price differences on the individual or regional level, researchers are constrained to rely on aggregate price-variation over time to assess substitution behavior. As a consequence, it is impossible to identify potential differences in substitution behavior between regions within a country. Not surprisingly, telecommunication market boundaries

* This chapter is based on joint work with Oliver Falck, Alexandra Heimisch, and Johannes Koenen

are, to date, predominantly drawn along country boundaries rather than at regional levels or for population groups.

In addition to introducing a new estimation approach, we generate a complex, novel data set on the Austrian telecommunication market from different sources. We combine data on fixed-line voice telephony usage provided by the Austrian incumbent operator with different sources of information on mobile network quality as well as data on the composition of the population of more than 2200 postcode regions across Austria. Two things, in particular, deserve to be emphasized: First, for the measures of mobile network quality, we are able to utilize fine-grained regional user-generated data on reliability and bandwidth. We can test for their accuracy due to the fact that we can additionally also rely on market research of the incumbent operator, thereby making this valuable source of data usable for research. Second, for our instrumental variable (IV) approach we gather novel, high-resolution satellite data allowing us the precisely characterize Austrian land cover at the level of grid-cells no larger than one hundred square meters.

Our results indicate a substitutive relationship between fixed-line voice and mobile voice telephony access. An increase in mobile quality in terms of bandwidth by one standard deviation (approx. 8 Mbps) reduces the incumbent's fixed-line market penetration by additional 1.09 percentage points over the time period from 2011 to 2014. This is equivalent to about 6.81% of its overall decline in market penetration between 2011 and 2014. The effects are larger for latency. A one standard deviation increase in latency (approx. 63 ms) leads to an additional decline in the incumbent's fixed-line market penetration by 2 percentage points (12.5%) for the same time period. Our results further suggest that fixed-mobile voice telephony substitution differs substantially across population groups and regions. For example, in postcode areas with higher shares of elderly people, substitution is substantially weaker. In contrast, fixed-mobile voice telephony substitution increases with the share of unemployed, and the share of foreigners within a postcode area, which may be related to price-sensitivity. These results suggest that the different market and competition environments across Austrian regions matter for substitution-patterns. This is an important finding for regulatory authorities across Europe, given the expressed aim of limiting interventions to markets, where they are essential to safeguard competition.

Identifying the causal effect of mobile quality on fixed-line usage is likely complicated by endogeneity concerns since the roll-out of mobile telephony infrastructure is driven by operators' profit considerations and therefore most likely related to the characteristics of the respective regions; there may be a stronger focus on regions with more (or more affluent) potential customers, for example. We address this matter in two different ways. First, we try to reduce endogeneity concerns by thoroughly controlling for regional differences in market size, network deployment cost, fixed-line market competition, the presence of amenities such as parks, lakes and rivers, and the composition of the population with regards to age, gender, education and income. Second, we pursue an IV approach. The quality of mobile connections depends on the distance (negatively) and unimpaired visual contact (positively) between handsets and base station antenna. Based on this observation, we derive two instrumental variables: i) the density of trees, and ii) the presence of railway tracks within a postcode area. Trees tend to conceal direct visual contact between mobile handset and base station antenna lowering actual mobile quality. Railway customers, on the other hand, are a distinct customer group on which mobile operators focus. Therefore, the presence of railway tracks likely increases mobile quality, because a relatively high base station antenna density is required to provide reliable mobile services to railway passengers. Houses located in the vicinity to railway tracks therefore incidentally benefit. We argue that in a sample of rural postcode areas without an own train station our instrumental variable approach is valid after taking differences in market size, household density and the presence of amenities into account.

Our findings complement the prior literature dealing with fixed-mobile substitution. A discussion of the topic and a comprehensive survey of early studies is provided by Vogelsang (2010). Among more recent studies, Briglauer et al. (2011) estimate cross-price elasticities between fixed-line voice-telephony access and mobile services in Austria, using balance-sheet data such as firm turnover, number of subscriptions and amount of call minutes to approximate mobile prices over time. They find fixed-line access to be inelastic while calls are elastic. In contrast, Ward and Zheng (2012) find that fixed-line and mobile subscriptions are rather strong substitutes in China during a similar time period. Similarly, Barth and Heimeshoff (2014a), Barth and Heimeshoff (2014b) and Lange and Saric (2016) conclude that fixed-line and mobile services are substitutes, using panel data on (a selection of) European Union member states up until 2010/2011. The existence of fixed-mobile substitution is also established in studies by Grzybowski (2014) and Grzybowski and Verboven (2016), who build their analysis on telecommunication choices of households across 27 EU countries. They find that substitution substantially differs across households and EU regions which is in line with our findings for Austria.

The remainder of this paper proceeds as follows. Section 1.2 presents our modeling with regards to “cross-quality elasticities”. Section 1.3 introduces the data we use and provides descriptive statistics. Section 1.4 introduces our estimation approach. Section 1.5 presents our results on our baseline as well as instrumental variable results for fixed-mobile substitution. Section 1.6 concludes by highlighting the implications of our findings for market definition.

1.2 Conceptual Framework

To briefly motivate our empirical modeling choices, consider the decision of a budget-restrained decision maker choosing a telephony product for calls placed from or to her home location. There are two technical options available for voice telephony at home: Fixed-line telephony (FL) and an alternative technology (AL), which is typically mobile telephony.

Fixed-line telephony (FL) is a mature technology with constant quality and availability. The network of the fixed-line operator that we look at reaches every household in Austria. The quality of the alternative technology (AT), mobile, on the other hand, may vary substantially. Define $u_{\text{j}}(k)$ as the utility that consumer j derives from using a given technology k for a time-period in monetary terms and p_{k} as technology k's per period cost to the consumer. If the decision maker picks either one or the other (i.e., if they were perfect substitutes), for any given prices there is a critical quality level of the alternative technology q^*_{AT} such that the decision maker is exactly indifferent between FL and AT, as $u_{\text{j}}(FL) - p_{\text{FL}} = u_{\text{j}}(AT, q^*_{\text{AT}}) - p_{\text{AT}}$, assuming that the utility of the decision maker is increasing in the quality of AT. Therefore, similarly to the cross-price elasticity between the two technologies which is based on the fact that individuals will switch from one technology to the other as the relative prices change, if the technologies are substitutes, we should find cross-quality elasticity between technologies; i.e. as the quality of mobile (AT) increases, the likelihood of choosing FL decreases.

In the case of Austria, a mature telecommunication market with close to complete mobile penetration[1], we have to slightly adjust the underlying thought experiment. In fact, the decision maker is almost certain to own and employ AT. Then, the decision problem is rather, whether the quality differential for voice telephony at home between AT and FL is large enough to justify the additional expenses for the fixed line. The prediction remains unchanged: The higher the quality of the alternative technology, the more likely the decision maker is to cancel her fixed-line contract, if the two are, in fact, substitutes. In the aggregate, we therefore predict that a higher quality of the mobile network should induce more customers to cancel their fixed line contracts.

Further note that the size of this effect should depend on the economic background of the decision maker. We would expect tighter budget constraints to be associated with more frequent re-evaluations of the necessity of multiple contracts, for example. Therefore it is necessary to control for sociodemographic characteristics of the populations of the regions under consideration.

1.3 Data

Our data comprises highly disaggregated regional information on all 2,216 4-digit postcode areas across Austria, which stretch over an area of 37 square kilometers on average. Data on fixed-line usage, mobile network quality and local geographical characteristics is available at the postcode level. In contrast to this, data on the local composition of the population is available only at the municipality level. Converting municipality level data to the postcode level is not straight forward since postcode areas and municipalities are not nested within each other. To recode municipality level information, we calculate area weights using geographic information software[2]. We have full information on all relevant variables for 2,169 postcode

[1] According to market research by Statista, there were 1.7 active SIMs per inhabitant in Austria in 2017. https://de.statista.com/themen/2930/mobiltelefonie-in-oesterreich/

[2] More precisely, we calculate the share of a municipality's total building ground area that is located in a particular postcode area. We use this to assign e.g. the absolute number of people at a specific age or with a

Table 1.1 : Descriptive Statistics

	Mean	SD	Min.	Max.
Panel A: Fixed Line Usage				
Growth Rate FL p.hh. (2014-2011)	-0.16	0.07	-0.45	0.03
Panel B: Mobile Network Quality				
HQ Mobile Coverage	0.75	0.16	0.00	1.00
Mobile Bandwidth (in Mbps)	14.58	7.97	0.09	73.02
Mobile Latency (in ms)	95.82	63.47	22.17	796.29
Panel C: Competition & Demographic Controls				
Cov. Cable (Yes/No)	0.42	0.49	0.00	1.00
Vol. Download p.c.	10.35	5.16	0.84	92.09
Frac. Female 2011	0.50	0.01	0.41	0.65
Frac. Unempl. 2011	0.05	0.03	0.00	0.34
Frac. Foreig. 2011	0.08	0.06	0.01	0.69
Frac. Sec. Educ. 2011	0.63	0.04	0.43	0.74
Frac. Tert. Educ. 2011	0.08	0.04	0.00	0.39
Frac. >65 2011	0.18	0.03	0.09	0.40
Frac. Students 2011	0.04	0.01	0.01	0.10
Panel D: Austrian Land Cover				
Dens.Trees	0.51	0.47	0.00	4.05
Rail(=1)	0.51	0.50	0.00	1
Station(=1)	0.31	0.46	0.00	1
Area Parks p.hh. (in 1000m2)	51.63	147.32	0.07	4997
Area Lakes p.hh. (in 1000m2)	0.38	1.81	0.00	30.97
Avg. Dist. River (in 1000m)	0.41	0.51	0.005	8.33
Avg. Dist. Station (in 1000m)	4,88	4.17	0.065	28.08

Note: Table shows 4-digit postcode averages. Data: A1 Telekom Austria, Statistik Austria, OpenStreetMap contributors (2018a), CadasterENV, RTR-NetTest (CC BY 4.0), Gfk Austria GmbH.

areas. To avoid outliers affecting our analysis, we drop the top and bottom percentile of postcode areas with regards to the log difference of fixed-line market penetration between 2011 and 2014. Our final data set therefore contains 2,127 postcode areas covering more than 98% of all Austrian households in 2011. All summary statistics shown in this chapter reflect the respective numbers for these postcode areas.

specific education of a municipality to a postcode area by multiplying the number of a these people within a specific municipality with the building ground share. This procedure assumes that overall population as well as population characteristics are distributed uniformly across building ground areas of municipalities.

1.3.1 Fixed-Line Usage & Mobile Network Quality

Our data on fixed-line usage is provided by the Austrian incumbent telecommunications operator for the time between 2011 and 2014. It comprises monthly information on the number of private fixed-line voice telephony accesses that are in use. We consider fixed-line accesses being in use if at least one call is made or received within every three months. Our analysis thus abstracts from fixed-line accesses that are held by customers with product bundles that are mainly booked for internet use or receiving television service. We aggregate monthly information to the year level, dropping postcode areas with missing information on the number of existing fixed-line accesses in more than 6 months of either 2011 or 2014[3]. We use this data to calculate the incumbent operator's fixed-line market penetration. It is measured as the number of fixed-lines used for voice telephony (but not exclusively, it may be additionally used for internet access or receiving TV services) on a regular basis relative to the number of households within a postcode area. Panel A in Table 1.1 illustrates how the incumbent fixed-line market penetration changed over time. In the average postcode area, it dropped by approximately 16%[4] between 2011 and 2014.

We derive information on mobile network quality from two different sources. First, the incumbent operator's market intelligence unit provides us with information on the overall, operator-independent 2G, 3G, and 4G mobile coverage for each individual Austrian postcode area in 2015. We use this information to calculate a measure of "higher quality mobile coverage". It is defined as the average household coverage of 2G, 3G, and 4G networks within a particular postcode area. Since 2G coverage is generally high throughout Austria, regional variation in higher quality mobile coverage mainly arises from regional differences in 3G and 4G coverage. Second, we retrieve information on regional differences in mobile quality from data provided by the Austrian Regulatory Authority for Broadcasting and Telecommunications (RTR). In 2013, RTR has established a network test, providing internet users around the world an opportunity to test their fixed-line as well as mobile connections. With regards to mobile connections, users performing a test generate, among other things, information on the data downstream transfer rate (henceforth bandwidth) as well as the latency of their current connection. We exploit information gathered by network tests between January 2013 and December 2014.

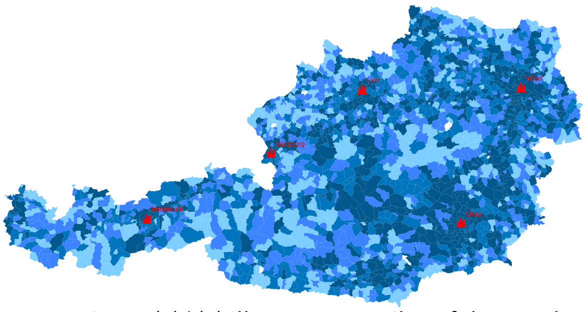
Figure 1.1 illustrates the regional distribution of these approx. 234,000 individual tests in total. It illustrates the absolute number of tests as well as the number of tests per household. A darker color implies a higher number of tests or tests per household. Overall, Figure 1.1 illustrates that the absolute number of tests as well as the number of tests per household tends to be higher in postcode areas close to major Austrian cities with more than 100,000 inhabitants. For each postcode area, we calculate the average bandwidth as well as latency from all available tests for this particular postcode area. Both bandwidth and latency affect the quality of voice

[3] This reduces the number of observations from 2,216 to 2,215.

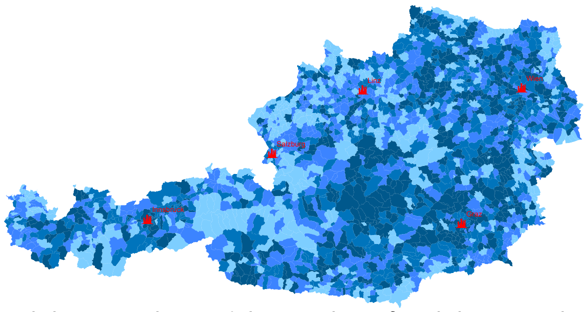
[4] This value is measured as the difference in natural logarithms of fixed-line market penetration in 2011 and 2014, respectively.

Figure 1.1 : Distribution of Mobile Network Tests across Austria

(1) Number of Tests



(2) Number of Tests per Households



Note: Figure (1)/ (2) illustrate quartiles of the number of mobile network tests/ the number of mobile network tests per households in the respective postcode area during 2013 and 2014. The darker the color, the more tests/ tests per households had been performed. The red marks illustrate Austrian cities with more than 100,000 inhabitants. Own illustrations. Data: RTR-NetTest (CC BY 4.0), Statistik Austria, Gfk Austria GmbH. Map: GfK Austria GmbH.

and video messaging. For example, low bandwidth may only allow for voice messaging, while higher bandwidth makes HD video streaming possible. Latency is important for Voice-over-Internet services working smoothly. If latency is too high, conversations are delayed which massively impairs conversations. In summary, both measures of mobile quality, as well as higher quality mobile coverage, mainly indicate regional differences in telecommunication customers' ability to use text, voice, and video messaging services. To a lesser extent, they are informative with regards to differences in the quality of basic voice telephony connections.

Figures 1.2 (1) and (2) illustrate mobile network quality across Austria. While Figure (1) depicts our measure of higher quality mobile coverage, Figure (2) illustrates latency. Comparing the two figures reveals that higher quality mobile coverage and latency across regions are highly correlated. This is a first indication with regards to the representativeness of the user generated network test data[5]. Figures 1.2 further shows that network quality varies substantially across Austria not only between densely and sparsely populated regions, but also between neighboring postcode areas. Table 1.1 illustrates that higher quality mobile coverage in the average Austrian postcode area of our sample is 75%. While average download speed is 14.6 Mbps, the latency attained during the average test is 95.8 ms.

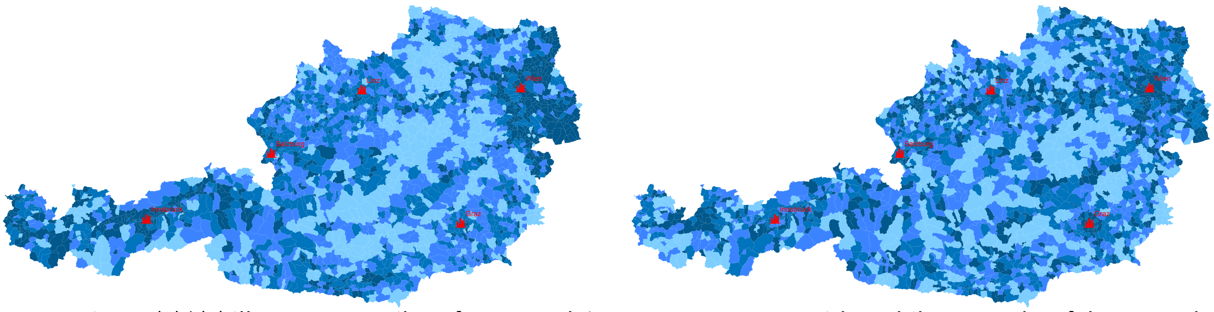
1.3.2 Geographic and Demographic Data

We gather data on the composition of the population of Austrian postcode areas and household numbers from Statistik Austria. The data allows us to identify differences in the composition of the population across postcode areas with regards to age, education, origin, and employment, among others.

[5] Note that in our regression analysis, we address remaining measurement issues by instrumenting the measures with each other, see the discussion below.

Figure 1.2 : Mobile Network Quality across Austria

(1) Average 2G, 3G, & 4G Coverage (2) Latency



Note: Figure (1)/ (2) illustrate quartiles of a postcode's average coverage with mobile networks of the second (2G), third (3G), and forth (4G) generation/ latency in 2015/ 2013 and 2014. The darker the color, the higher the quality (i.e. the higher coverage, the lower latency) of the mobile networks. The red marks illustrate Austrian cities with more than 100,000 inhabitants. Own illustrations. Data: A1 Telekom Austria, RTR-NetTest (CC BY 4.0), GfK Austria GmbH. Map: Gfk Austria GmbH.

Finally, we derive information on land cover (e.g. buildings, vegetation, water, etc.) at the postcode level from OpenStreetMap and CadasterENV Austria. OpenStreetMap provides the exact location of railway tracks, lakes, parks, rivers, the location of train stations, and the location of buildings and their shape across Austria. We use this information to identify all built-up areas across Austrian postcode regions to calculate the water area of lakes per household, the area of parks per household, the average distance of a postcode's household to the closest river, and a dummy variable that is equal to one if railway tracks are running through a postcode area (train station is located in a postcode area) and zero otherwise. CadasterENV Austria provides an Austrian wide land cover map funded by the European Space Agency (ESA) using Sentinel satellite imagery from August 2015 until December 2017. CadasterENV Austria characterizes Austrian land cover in great detail, providing information at the grid-level with a cell size of 100 square meters. We use this data to calculate tree density, measured as the number of cells that are mainly covered by trees within a radius of 30 meters from buildings relative to a postcode's built-up area. Panel D in Table 1.1 illustrates that railway tracks run through 51% of all Austrian postcode areas in our sample. Further, it indicates that substantial variation exists in tree density across Austrian postcode areas.

1.4 Empirical Approach

Our conceptual model suggests that the probability of individual i choosing alternative j is:

$$P_{ij} = Pr\big(U_{ij} > U_{ik}\big) = G\big(x_{ij}, x_i, \beta\big), \cap_{j \neq k} \tag{1.4.1}$$

where x_{ij} are attributes such as quality or price of alternative j that are faced by individual i, while x_i captures characteristics of the individual, choosing an alternative. β is a set of parameters to be estimated statistically. To determine to probability of individual i choosing a particular technology, a standard approach is to estimate a discrete choice model. Yet, this requires data on individual technology choices. Instead of observing individual decisions, we observe choices on a more aggregate regional level. Besides, we are not looking at a market in equilibrium but at a market that is facing technological change and individuals that (periodically) re-evaluate their technology choice. Intuitively, we thus look at a shift from one equilibrium, in which the incumbent fixed-line technology was the overwhelming standard, to a new equilibrium with more technological diversity. Accordingly, to capture the substitution relationship between technologies, we have to focus on the switching behavior of decision makers. To that end, we estimate the following empirical model:

$$\Delta(\ln FL_{rt=1}, \ln FL_{rt=0}) = \alpha + \beta_1 \ln FL_{rt=0} + \beta_2 MQ_{rt=0} + \mathbf{X}_{rt=0}\beta_3 + \varepsilon_{rt} \tag{1.4.2}$$

The dependent variable is the percentage change in the incumbent fixed-line market penetration in postcode area r between $t = 0$ (2011) and $t = 1$ (2014). We control for the initial share of fixed-line subscribers in 2011 in the postcode area ($\log FL_{\mathrm{rt=0}}$) to account for the fact that switching behavior of consumers in a given time period likely not only depends on the current availability of alternative technologies, but also on previous developments. If, for example, the mobile technology was introduced earlier in area A, but only recently in area B, then the pool of potential switchers and the probability of any given individual to switch, could be higher in area B than in A, where many potential switchers may have switched already.

Our main coefficient of interest is β_2. It measures the relationship between mobile network quality and the percentage change in incumbent fixed-line market penetration between 2011 and 2014. Our measure of higher quality mobile coverage is available for 2015, data on our measures for bandwidth and latency stem from 2013/ 2014 rather than 2011. Our measures of mobile quality thus rather proxy than precisely measure mobile coverage in 2011.

Estimating this equation by OLS is potentially problematic regarding the identification of causal effects for two reasons. On the one hand, mobile networks are rolled-out by profit maximizing network operators, who may deploy or increase the quality of local networks in response to local demand and willingness to pay. On the other hand, mobile network coverage and quality likely differ across regions because of heterogeneous deployment cost, not only affecting mobile network quality but also the quality of fixed-line networks[6] and competition in the fixed-line market. Consequently, β_2 may capture the effect of customers switching fixed-line operators rather than technologies. We address these concerns by accounting for regional differences in market size, network deployment cost and fixed-line competition in 2011. We measure market size as the natural logarithm of households within a postcode area. To account for network deployment cost, we include two different measures for household density. First, we calculate the number of households as the fraction of a postcode's built-up area. Second, we include a set of dummy variables that indicate the minimum number of separated built-up areas covering 50% of the postcode's total built-up area. We account for differences in fixed-line competition across postcode areas by including a dummy variable that is equal to one if

[6] Note that the quality of fixed-line voice services is constant across all Austrian postcode areas. However, differences in the fixed-line bandwidth achievable may affect voice telephony provider choice indirectly, in particular because of product bundles.

cable operators offer broadband services in a particular postcode area. Besides, we control for the average download volume of incumbent customers per month. Finally, we include a set of demographic variables to take into account differences in the composition of the population with regards to age, education and gender. As a proxy for differences in income, we further include the share of unemployed and control for the presence of amenities, namely the area of parks per household, the area of lakes per household, and the average distance of all buildings within postcode area to the closest river (all measured in natural logarithms).

Apart from endogeneity concerns, results from a simple OLS regression may suffer attenuation bias because of measurement error. Measurement error is likely an issue for our measures of mobile quality, bandwidth and latency. Even though they result from real-use information, such that they likely reflect actual mobile quality, our measures of bandwidth and latency may be relatively noisy in postcode areas where we only observe few tests.

We address potential attenuation bias by instrumenting bandwidth and latency with higher quality mobile coverage. Even though all three measures – including higher quality mobile coverage – tend to be measured with error, instrumenting bandwidth and latency with higher quality mobile coverage allows us to remove attenuation bias as long as the errors are not perfectly correlated. Since the data used to calculate average bandwidth and latency arise from a different source than the data used to calculate higher quality mobile coverage, i.e. the RTR network test on the one hand and the incumbent operator on the other, this is likely true.

1.5 Results

1.5.1 Baseline Results & Heterogeneous Effects

This section presents results on the relationship between mobile quality (measured according to the different approaches described above) and the percentage change in incumbent fixed-line market penetration. Columns (1) to (3) of Table 1.2 present results for the data downstream rate (bandwidth), while Columns (4) to (6) present results on latency. Data downstream rate and latency are both measured in natural logarithms. To make effect sizes comparable between the two measures, we normalize their standard deviation to one. Hence, effect sizes illustrate the impact of an increase in mobile quality by one standard deviation on the percentage change in incumbent fixed-line market penetration between 2011 and 2014.

We have argued that our measures of mobile quality are subject to attenuation bias. Throughout all specifications in Table 1.2, we therefore instrument bandwidth and latency with higher quality mobile coverage, i.e. the average share of a postcode's area that is covered by networks of the second (2G), third (3G), and fourth (4G) generation. We provide the corresponding non-instrumented estimates in Table B.2 in the Appendix. To account for path dependencies in switching behavior, we control for initial fixed-line diffusion throughout all specifications. Besides, we control for postcode size in 2011, household density in 2011, and fixed-line compe-

Table 1.2 : Ordinary Least Squares Results

	lnBandwidth			lnLatency		
	(1)	(2)	(3)	(4)	(5)	(6)
Mobile Quality	-0.0121**	-0.00902*	-0.0109**	0.0167**	0.0159*	0.0200**
	(0.00501)	(0.00476)	(0.00490)	(0.00685)	(0.00847)	(0.00919)
Log of FL p.hh. '11	0.0490***	0.0471***	0.0433***	0.0454***	0.0447***	0.0412***
	(0.00583)	(0.00592)	(0.00595)	(0.00601)	(0.00598)	(0.00596)
Log HH. in '11	-0.00143	0.00189	0.00440**	-0.000967	0.000895	0.00290
	(0.00194)	(0.00206)	(0.00215)	(0.00202)	(0.00193)	(0.00203)
HH Dens. '11	-0.0334***	-0.0312***	-0.0266***	-0.0313***	-0.0302***	-0.0255***
	(0.00421)	(0.00424)	(0.00456)	(0.00420)	(0.00418)	(0.00455)
Cov. Cable (Yes/No)	-0.0408***	-0.0389***	-0.0396***	-0.0389***	-0.0376***	-0.0380***
	(0.00322)	(0.00320)	(0.00319)	(0.00317)	(0.00316)	(0.00317)
lnDown. p.hh	0.00175	0.00469	0.00652*	0.00538	0.00715*	0.00918**
	(0.00391)	(0.00389)	(0.00392)	(0.00413)	(0.00410)	(0.00413)
lnArea Parks p.hh		0.00183	0.00332*		-0.000396	0.000661
		(0.00157)	(0.00172)		(0.00203)	(0.00217)
lnArea Lakes p.hh		-0.000867	-0.000523		-0.00118*	-0.000923
		(0.000675)	(0.000682)		(0.000695)	(0.000712)
lnDist. River		0.000605	-0.000647		0.000502	-0.000578
		(0.00160)	(0.00160)		(0.00154)	(0.00155)
lnDist. Station		0.0111***	0.00920***		0.0101***	0.00819***
		(0.00156)	(0.00164)		(0.00160)	(0.00165)
Frac. >65 '11			0.0258			-0.0189
			(0.0471)			(0.0565)
Frac. Students '11			0.283			0.233
			(0.182)			(0.184)
Frac. Sec. '11			-0.163***			-0.131***
			(0.0372)			(0.0392)
Frac. Tert. '11			0.000591			0.00686
			(0.0460)			(0.0459)
Frac. Female '11			-0.389***			-0.288*
			(0.141)			(0.161)
Frac. Unempl. '11			-0.0292			-0.0390
			(0.0569)			(0.0574)
Frac. Foreig. '11			-0.116***			-0.0980***
			(0.0393)			(0.0380)
Frag. FE	Yes	Yes	Yes	Yes	Yes	Yes
N	2127	2127	2127	2127	2127	2127

Note: Dependent variable is the percentage change in incumbent fixed-line market penetration between 2011 and 2014. Mobile Quality as indicated in the Column head. Frag. FE refers to a set of dummy variables indicating the minimum number of a postcode's separated built-up areas to cover ≥ 50% of its total built-up area. Throughout all specification *Mobile Quality* is purged from measurement error by instrumenting it with higher quality mobile coverage. Huber-White standard errors applied. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

tition in 2011 throughout Columns (1) to (6). We additionally account for regional differences in environmental amenities such as parks, lakes and rivers in Column (2) and (4). Finally, we control for differences in the composition of the population across postcode areas in Columns (3) and (6).

For the interpretation of results, we focus on the most comprehensive specifications in Column (3) and Column (6) of Table 1.2. Column (3) suggests that a rise in achievable bandwidth by one standard deviation induces an additional decline in incumbent fixed-line market penetration by 1.09 percentage points. This corresponds to about 6.8% of the actual decline in market penetration (-16%) faced by the incumbent operator in the average Austrian postcode area between 2011 and 2014. Column (6) illustrates that effects are larger with regards to latency. An increase in latency by one standard deviation leads to an additional decline in incumbent fixed-line market penetration by 2 percentage points or 12.5% of the average observed decline in incumbent market penetration.

Throughout all specifications, we find a positive relationship between initial market penetration in 2011 and the percentage change in incumbent fixed-line market penetration between 2011 and 2014. This implies that differences in fixed-line usage across postcode areas in 2011 do not vanish over time. On the contrary, postcode areas with a higher share of incumbent fixed-line market-penetration in 2011 depict a higher share of incumbent fixed-line market penetration in 2014 as well. This suggests that the market environment and competition intensity in the telecommunications market persistently differ across postcode areas in Austria. This motivates the following analysis, more closely investigating potentially heterogeneous effects in fixed-mobile voice substitution by population group and across Austrian regions. To that end, we estimate the following model:

$$\Delta(\ln FL_{rt=1}, \ln FL_{rt=0}) = \alpha + \beta_1 \ln FL_{rt=0} + \beta_2 MQ_{rt=0} + \beta_3 (MQ \times X^j)_{rt=0} + \beta_4 X^j_{rt=0} + \mathbf{X}^{k \neq j}_{rt=0} \beta_5 + \varepsilon_{rt} \tag{1.5.1}$$

where $MQ \times X^j$ is an interaction of two continuous variables, namely our measure of mobile network quality and the share of a population subgroup of interest. A positive estimate on β_3 would imply that the greater the share of the population group, X^j, the stronger the effect of a given level of mobile network quality on the rate of change of incumbent fixed-line market penetration between 2011 and 2014. In a model with interacted continuous variables, coefficient estimates on the main effects, β_2 and β_4 reflect conditional relationships. For example, the estimate on β_2 is the effect of mobile network quality, MQ, on the rate of change in incumbent fixed-line penetration, conditional on X^j being zero. This is illustrated in Equation 1.5.2. In what follows, we skip control variables for the ease of exposition:

$$\begin{aligned} \Delta(\ln FL_{rt=1}, \ln FL_{rt=0}) &= \beta_2 MQ_{rt=0} + \beta_3 (MQ \times 0)_{rt=0} + \beta_4 0_{rt=0} + \mathbf{X}^{k \neq j}_{rt=0} \beta_5 + \varepsilon_{rt} \\ &= \beta_2 MQ_{rt=0} + \mathbf{X}^{k \neq j}_{rt=0} \beta_5 + \varepsilon_{rt} \end{aligned} \tag{1.5.2}$$

while a share of 1 for a specific population group would result in:

$$\begin{aligned}\Delta(\ln FL_{rt=1}, \ln FL_{rt=0}) &= \beta_2 MQ_{rt=0} + \beta_3 (MQ \times 1)_{rt=0} + \beta_4 1_{rt=0} + \mathbf{X}^{k \neq j}_{rt=0}\beta_5 + \varepsilon_{rt} \\ &= (\beta_2 + \beta_3) MQ_{rt=0} + 1\beta_4 + \mathbf{X}^{k \neq j}_{rt=0}\beta_5 + \varepsilon_{rt}\end{aligned} \tag{1.5.3}$$

In case of a completely homogeneous population within a postcode area, the effect of mobile network quality on the rate of change in incumbent fixed-line market penetration is $\beta_2 + \beta_3$.

Estimating this model, we allow for heterogeneous effects of mobile coverage across postcode areas depending on the composition of their population. At the same time, we assume that the effects of all other variables included in the regression are constant across postcode areas. This is a major difference with regards to splitting the sample, which would be equivalent to a fully-interacted model. We pursue this approach for two reasons: First, it allows us to estimate effects on all observations in our sample, which increases precision substantially. Second, we argue that it is reasonable to assume that the effects of our control variables are constant across postcode areas. For example, younger people may be inclined to adopt new technologies, or switch their telecommunication operator more frequently, independent of the place they live in. In contrast, mobile phone adoption is likely to differ across regions because the composition of their population is different and these population groups differ in their likelihood to adopt new technologies. Table 1.3 presents results for the interacted model considering different population groups and the effect of bandwidth. Corresponding results for latency are provided in Table B.3 in the Appendix. Panel A investigates the role of high shares of people beyond the age of 65, Panel B focuses on the share of people with tertiary education. Finally, Panel C and D investigate the role of the share of unemployed individuals and the share of foreigners within a postcode area. Comparing coefficients on the main effects of mobile quality across Columns (1) to (3), we observe that the effect of an increase in bandwidth systematically differs along the distribution of demographic characteristics. For example, moving along the percentiles of elderly, we observe that the effect becomes smaller in absolute terms. This implies that the substitutive relationship between fixed-line voice and mobile voice telephony access is smaller in regions with a high share of elderly in the population. It is in line with the belief that older people are on average less likely to give up their habits and more reluctant to adapt to changes. Similarly, the substitutive relationship between fixed-line voice and mobile voice telephony tends to be smaller in regions with high shares of tertiary educated. Our results further suggest that fixed-mobile substitution increases with the share of unemployed and with the share of foreigners within a postcode area. One potential explanation could be that unemployed are more likely to be financially constrained, giving them additional incentive to re-evaluate and cancel their fixed line contracts. With regards to differential effects across postcode areas depending on their share of foreigners, an explanation could be that it is easier to use a pre-paid mobile tariff compared to getting fixed-line access. Overall, this suggests that the competition from mobile operators faced by fixed-line operators differs across Austrian postcode areas.

Table 1.3 : The Role of Population Composition in Fixed-Mobile Voice Substitution

	(1) 10th percentile	(2) 50th percentile	(3) 90th percentile
Panel A: Fraction Population > 65 in 2011			
Mobile Quality	-0.0161**	-0.0115**	-0.00540
	(0.00706)	(0.00517)	(0.00541)
Mobile Quality x Frac. > 65	0.00427	0.00427	0.00427
	(0.00306)	(0.00306)	(0.00306)
Panel B: Fraction Population Max. Tertiary Educated in 2011			
Mobile Quality	-0.0123*	-0.0107**	-0.00738
	(0.00676)	(0.00508)	(0.00688)
Mobile Coverage x Frac. tert. educ.	0.00259	0.00259	0.00259
	(0.00511)	(0.00511)	(0.00511)
Panel C: Fraction Population Unemployed in 2011			
Mobile Quality	-0.00200	-0.00900*	-0.0220***
	(0.00542)	(0.00487)	(0.00852)
Mobile Coverage x Frac. unemployed	-0.0108**	-0.0108**	-0.0108**
	(0.00514)	(0.00514)	(0.00514)
Panel D: Fraction Foreigners in 2011			
Mobile Quality	0.00526	-0.00664	-0.0322**
	(0.00871)	(0.00503)	(0.0155)
Mobile Coverage x Frac. foreigners	-0.0178*	-0.0178*	-0.0178*
	(0.0104)	(0.0104)	(0.0104)
Frag. FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
N	2127	2127	2127

Note: Dependent variable is the percentage change in incumbent fixed-line market penetration between 2011 and 2014. Mobile Quality indicates the average tested data downstream rate (bandwidth). Frag. FE refers to a set of dummy variables indicating the minimum number of a postcode's separated built-up areas to cover ≥ 50% of its total built-up area. Controls include all controls of Column (6) in Table 1.2. Throughout all specification *Mobile Quality* is purged from measurement error by instrumenting it with higher quality mobile coverage. In Column (1), (2), and (3), the panels' respective demographic variables are standardized to have a value of zero at the 10th, 50th, 90th percentile. Huber-White standard errors applied. Significance levels: *p<0.10, **p<0.05, ***p<0.01.

1.5.2 Validity

To substantiate the validity of our results, we pursue an instrumental variable strategy. It exploits the fact that the quality of mobile connections decreases i) with distance and ii) to the extent that the direct visual contact between handset and base station antenna is impaired. Apart from buildings, likely obstacles located between mobile handsets and base station antenna are trees. As a first instrument, we thus calculate the “tree density” within postcode areas. However, simply calculating tree density as the share of a postcode’s area that is covered by trees is potentially problematic with regards to instrument validity. This is because postcode areas in rural regions comprise relatively large, unpopulated areas that may be covered by trees, which do not though intercept signals from antennae to homes. There are two potential issues that we want to take into account with regard to this instrument: first, the measure should disregard trees that are not close to buildings (such as wooded, unpopulated areas in rural regions), since they do not affect mobile signal quality. Second, the measure should not be responsive to public parks that may contribute amenity value to the postcode area. Relatedly, measuring tree density simply as the share of a postcode’s area that is covered by trees would likely capture other general differences between urban and rural areas. We address these issues in two ways in our calculation of tree density from high-resolution satellite data: the measure is composed only of the number of trees that are located within a radius of 30 meters around a building, relative to the postcode’s total built-up area (thereby addressing the issue of trees in unpopulated areas). Second, we control for the area of parks per household (amenity value). Consequently, variation in our measure of tree density across postcode areas only results from differences in the number of trees in (non-communal) gardens – close to buildings – and along streets.

Our second instrument exploits the fact that travelers are a particularly important group of customers for mobile network operators. As a result, mobile infrastructure is typically especially developed along major roads as well as the railway network, in particular. For our purposes, railways are of special interest: Physically, trains are Faraday cages, i.e. metal cases that shield their contents from electromagnetic waves, which makes it difficult to provide railway passengers with a high-quality mobile signal. (Long-distance) trains in Austria usually carry special equipment to improve signal quality, which further increases demand for these services by travelers. As a result, network operators deploy a dense network of base station antennas along train tracks to provide reliable service. Consequently, households close to railway stations and train tracks (incidentally) benefit from increased network density. We exploit this to derive a second instrument, a dummy variable that is equal to one, if train tracks run through a particular postcode area and zero otherwise. Again, this instrumental variable likely also captures differences in urban development of postcode areas, which we have to take into account. Especially the existence of a train stations within a postcode area is generally not random and will also reflect other economic differences. Again, we address these issues in two ways: First, we restrict our sample to postcode areas without a train station. In addition to controlling for the average distance of all buildings in a postcode area to the closest train station, this isolates the possibly positive direct economic effects of a train station. These effects might influence both switching behavior of customers (wealth or income effects) as well as, relatedly, the incentives of operators to invest. Second, we also directly proxy for differences in income across postcode areas by accounting, among other things, for the presence of amenities and other factors related to income. In particular, we control for the area of lakes per household, the average distance of all buildings in a postcode area to the closest river, and the share of unemployed within a postcode area. In summary, our second instrumental variable exploits differences in mobile quality arising from the fact that households in some postcode areas - even though not connected to the train network via railway stations themselves – but not others profit from tracks leading through their area.

We use both instruments to estimate the following first-stage equation:

$$MQ_{rt=0} = \theta + \gamma_1 \ln FL_{rt=0} + \gamma_2 Dens.Trees_{rt=0} + \gamma_3 D(Rail)_{rt=0} + \mathbf{X}_{rt=0}\gamma_4 + \varepsilon_{rt} \qquad (1.5.4)$$

where Dens.Trees and D(Rail) are our instrumental variables. Dens.Trees measures the number of 10×10 meter grid cells that are mainly covered by trees within a radius of 30 meters from a building relative to the postcode's overall built-up area. D(Rail) is a dummy variable that is equal to one if railway tracks run through the postcode area and zero otherwise. Table 1.4 presents the results of our instrumental variable estimation for both our measures of mobile quality. While Column (3) depicts results for bandwidth, Column (4) depicts results for latency. Columns (1) and (2) contain the OLS results using the model from Columns (3) and (6) in Table 1.2, but applied to the restricted sample that we use in our instrumental variable estimations. A comparison of Columns (3) and (6) in Table 1.2 with Columns (1) and (2) in Table 1.4 reveals that restricting the sample barely changes the results. Moreover, the estimated coefficients between the two samples are not significantly different from each other. This suggests that the results for the restricted sample are informative with regards to the results that can be expected for the average Austrian postcode area.

Columns (3) and (4) of Table 1.4 show the results for our instrumental variables estimations. The IV coefficients are similar for both of our measures. Column (3) indicates that an increase in the natural logarithm of bandwidth by one standard deviation increases the decline in incumbent market penetration between 2011 and 2014 by an additional 6.08 percentage points. The estimate for latency in Column (4) suggests that an increase in latency (reduction in mobile quality) by one standard deviation reduces the decline in incumbent market penetration by 5.13 percentage points.

Both in terms of tests and signs, the first stage results are as expected. The results of a test for over-identifying restrictions (Hansen's J) support the validity of our instrumental variable approach. In the first stage, a higher density of trees reduces, while the presence of rail tracks increases mobile quality. The results for latency reflect the same findings, though the signs are switched for the same reason (higher latency implies lower quality). Depending on the measure of mobile quality, different instruments are particularly relevant in terms of first stage coefficient size and statistical significance: For the data downstream rate, both first

Table 1.4 : IV Results

	OLS		IV	
	(1) lnBandwidth	(2) lnLatency	(3) lnBandwidth	(4) lnLatency
Mobile Quality	-0.00978*	0.0190*	-0.0608**	0.0513**
	(0.00514)	(0.0103)	(0.0248)	(0.0211)
First stage				
Dens.Trees			-0.156	0.235***
			(0.099)	(0.073)
Rail(=1)			0.184***	-0.072
			(0.059)	(0.063)
Kleib.-Paap F-stat.			6.53	5.91
Hansen's J (p-value)			0.898	0.155
Frag. FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
N	1461	1461	1461	1461

Note: Dependent variable is the percentage change in incumbent fixed-line market penetration between 2011 and 2014. Mobile Quality as indicated in the Column head. Columns (1) and (2) show OLS results, where *Mobile Quality* is purged from measurement error by instrumenting it with higher quality mobile coverage. Columns (3) and (4) show instrumental variable estimations. Frag. FE refers to a set of dummy variables indicating the minimum number of a postcode's separated built-up areas to cover ≥ 50% of its total built-up area. Controls include all controls of Column (6) in Table 1.2. Huber-White standard errors applied. Significance levels: *p<0.10, **p<0.05, ***p<0.01.

stage coefficients are sizable. Yet, the coefficient on tree density is not statistically significant. For the latency variable, the effect of tree density is sizable and statistically significant. The physical mechanism behind this effect of tree density can be explained as follows: A signal being reflected at an obstacle causes the signal to occur multiple times. As a result, the mobile handset has to test constantly which of the signals is the best and most recent, leading to a rise in latency. In contrast, the coefficient on D(Rail) regarding latency is both rather small and not statistically significant, when simultaneously controlling for tree density and the mechanism associated with it.

1.6 Conclusion

This paper proposes a new method to quantify substitution between two goods or services, based on switching behavior of consumers depending on the quality of alternatives in a given region – this can be interpreted as a cross-quality elasticity. We apply our approach to the telecommunications market and focus on quantifying fixed-mobile voice telephony substitution in Austria. Cross-quality elasticities complement the standard cross-price elasticity approach in important dimensions. First, if prices do not vary across regions (there is no regional price discrimination in Austria), identification does not require any price information, or assumptions to derive prices for mobile telephony, in particular. Second, our approach allows us to focus on differences in substitution behavior across regions and distinct groups of the population. This helps to identify regional differences in competition intensity and the drivers thereof, e.g. the composition of the population or the availability of alternative infrastructure. These findings are highly informative for regulators and would allow them to tailor interventions to (local) markets, where they are necessary to safeguard competition.

Using novel data sets, including user-generated network tests and fine-grained satellite data, we apply our model to quantify the extent of fixed-mobile voice substitution in Austria between 2011 and 2014. To estimate causal effects, we pursue an IV approach that exploits exogenous variation in mobile network quality derived from the local density of trees as well as the proximity of a postal code area to railway tracks. The underlying idea of these instrumental variables is to exploit the fact that a small enough distance and direct visual contact between mobile handset and base station antenna are decisive in determining mobile signal quality. Trees impair direct visual contact, while along railway tracks base station antenna density is substantially higher on average.

Our results indicate substantial fixed-mobile substitution across Austria between 2011 and 2014. An increase in mobile network quality – the alternative technology to fixed line telephony – substantially increases the incidence of switching away from fixed-line telephone contracts during that period. In particular, an increase in bandwidth (latency) by one standard deviation reduces (increases) the percentage change in incumbent fixed-line market penetration by 1.09 (2.0) percentage points or 6.81% (12.5%) of its average overall decline from 2011 to 2014. An analysis of potential heterogeneous effects further suggests that fixed-mobile substitution is stronger among highly-educated, foreign and unemployed individuals, while it is smaller among elderly population groups. This is derived from regional differences in substitution behavior. For example, in postcode areas with a relative high share of an elderly population, fixed-mobile substitution is significantly less pronounced compared to a postcode area with a low share of elderly (and analogous results derive for a region with low share of foreigners or unemployed individuals).

Overall, our findings indicate that market characteristics at the level of (sub-national) regions matter substantially for competition and substitution patterns across different telecoms services. Any optimal regulation would have to take regional characteristics and particularities into account.

2 Is it all about Speed? Competition in the Fixed-Line Broadband Market

2.1 Introduction

For more than a decade, incumbent operators dominated telecommunications markets in many countries across Europe; only regulatory intervention limited their market power. However, surging digitization changes the market increasingly. On the one hand, demand for higher bandwidth is rising. For example, the number of people surfing at a data downstream rate of up to 30 Mbps nearly quadrupled in only five years from 7.2% in 2012 to 27% in 2017. On the other hand, alternative operators have been investing continuously, covering an increasing number of households. For example, cable networks (DOCSIS 3.0) and fiber networks (FTTP) jointly cover 57.8% of all households across Europe. (European Commission, 2018)

Cable networks and fiber networks are typically providing bandwidth capacity beyond what can be achieved via traditional copper-based networks. Incumbent operators are thus facing increasing competition in quality. To increase achievable bandwidth, they typically replace (parts of) the copper wires with fiber wires. However, replacing these wires is costly and time-consuming, especially in countries like Germany, where wires have to be rolled-out subsurface. To bridge the time until its fixed-line network is modernized, the German incumbent operator has launched a *fixed-mobile hybrid product* combining bandwidth from its fixed-line DSL and mobile Long Term Evolution (LTE) networks to surf the internet at home at higher speed.

In this paper, I examine the role of this fixed-mobile hybrid product in shaping local competition. To that and, I analyze the effect of differences in fixed-mobile hybrid coverage across German regions on the incumbent operator's fixed-line market penetration. I exploit the fact that households across different regions gained access to the fixed-mobile hybrid product depending on the quality of the local mobile LTE network. Even though mobile networks are unlikely deployed in response to fixed-line demand and competition – issues of reverse causality should thus not matter – regional differences in mobile network quality are not occurring at random. To take structural differences in mobile quality across regions into account I control for differences in market size, deployment cost, fixed-line infrastructure quality, competition intensity, and the composition of the population, using data on more than 5200 German local exchange areas[1] (LE) across Germany.

I find that hybrid product introduction increased the annual rate of change in incumbent fixed-line market penetration between 2011 and 2014 by 0.86 percentage points. I further find

[1] A local exchange area is a geographically restricted section of the telephony access network. Historically, a LE in Germany comprised all subscribers with the same local area code

that the effects differ substantially by age of customers, with effects being more than three times larger among customers that are between 18-34 years old than among customers that are 65 or older. Effects also vary across different types of regional markets. For example, the fixed-mobile hybrid product does not statistically significantly affect the change in incumbent fixed-line market penetration in LEs with weak infrastructure-based competition. However, it significantly affects incumbent market penetration in LEs with relatively high infrastructure-based competition. The results presented throughout this chapter are robust to variations in the control variables included in the estimations. Besides, a placebo test finds zero effect, evaluating how hybrid product coverage affects incumbent fixed-line market penetration prior to the actual launch of the hybrid product. This is reassuring with regards to the validity of the results presented.

The results of this paper contribute to the current debate on how to incentivize market driven network expansion to overcome existing imbalances in high-capacity broadband coverage and broadband usage, in particular between rural and urban areas. My findings make two contributions. First, they suggest that absent of substantial investment incentives or state-funding private investments in high-capacity broadband networks in rural, sparsely populated regions, where infrastructure-based competition is low are not to be expected. Even regulatory exemptions may not provide sufficient incentives. Second, they suggest that quality leadership in terms of bandwidth is unlikely to lead to market monopolization, even in regions where infrastructure based competition is small. Yet, given that younger customers are more responsive to increases in bandwidth, this may change in future.

The paper contributes to a well-developed literature investigating the effect of regulation on investment, broadband penetration, and quality from a theoretical (Bourreau and Dogan, 2006; Klumpp and Su, 2010; Vogelsang, 2003) as well as empirical perspective (Economides et al., 2008; Grajek and Röller, 2012; Nardotto et al., 2015; Xiao and Orazem, 2011). In particular, this paper relates to a set of recently emerging papers that examine how regulatory policy should be designed to incentivize investment in high-capacity, next generation access (NGA) networks. For example, Briglauer (2015) examines the role of regulation and competition on investment dynamics. His results indicate that existing access regulation, infrastructure-based competition from mobile operators, and existing incumbent infrastructure negatively affect investment incentives in fiber networks. In a similar vein, Vogelsang (2017) argues for a reduction of legacy regulations in the telecommunications sector that conflict with the dynamics of the industry and product innovations. Further, Briglauer and Vogelsang (2017) propose that regulatory policies should be oriented on geographic differences in the competitive environment. In particular, they distinguish competitive black areas from non-competitive gray areas with only one infrastructure provider and white areas, where investing would not even be profitable for a monopolist. Since it is particular expensive to deploy NGA networks in such gray and white areas, high valuation of bandwidth is essential in generating investment incentives. There exist few papers analyzing consumer valuation for higher bandwidth. For example, Grzybowski et al. (2018) find that valuation of FTTH connections providing 100 Mbps relative to DSL connections providing between 1 Mbps and 8 Mbps has been rapidly increasing during January and December 2014. Besides, using data from a household survey, Liu et al. (2018) show that valuation of an increase in bandwidth decreases with the level of bandwidth already available.

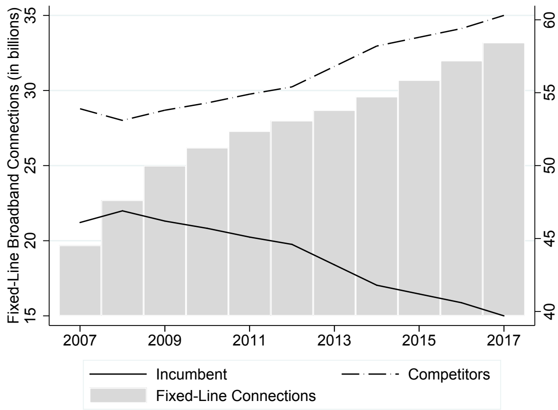
The rest of this chapter proceeds as follows. Section 2.2 provides an overview on the German fixed-line broadband market, outlining the competitive environment as well the regulatory framework in place. Section 2.3 presents the fixed-mobile hybrid product. Apart from explaining the technical realization of the product and discussing the contracts offered, it also demonstrates structural differences with regards to its availability across Germany. Section 2.4 turns to analyzing the role of the fixed-mobile hybrid product in shaping competition in the fixed-line broadband market. Prior to outlining the estimation approach in Section 2.4.2, I present the data used in Section 2.4.1. Section 2.4.3 presents the results of the analysis and Section 2.5 concludes.

2.2 The German Fixed-Line Broadband Market

The German fixed-line broadband market has substantially grown during the last decade from 19.7 billion connections in 2007 to 33.2 billion connections in 2017. Figure 2.1 illustrates that about 77% of these broadband connections are realized by Digital Subscriber Line Technology (DSL) via the public switched telephone network (PSTN). While cable modem connections via Hybrid-Fiber-Coax (HFC) networks make up for 21.5%, fiber lines ranging as far as to the building or even customer premises (FTTH/P) account for 1.3%. Alternative technologies, such as broadband wireless access (BWA), power-line or satellite contribute 0.7% of all broadband connections to the market. Historical developments and path dependencies are reasons why market penetration of HFC and FTTH/P connections is relatively low compared to DSL. (Bundesnetzagentur, 2017).

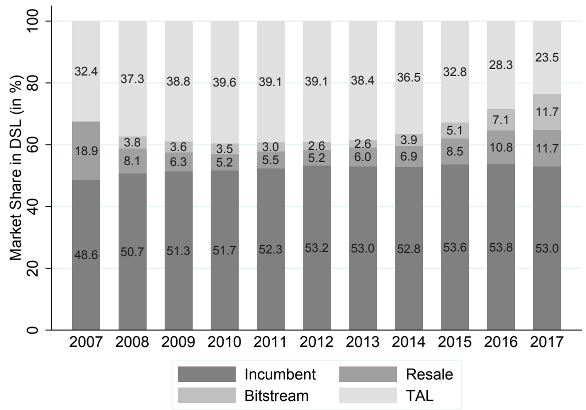
Until the liberalization of telecommunications markets across Europe during the 1990s, posts and telecommunications were administered publicly. The authority in charge, Deutsche Bundespost, started to develop the PSTN as well as the cable television network in the late 1950s and 1970s, respectively. Since constitutional law stipulated universal voice telephony coverage, i.e. an obligation to provide direct distance dialing telephony services to all German households, Deutsche Bundespost had already established the PSTN throughout the entire country, when broadband internet was commercialized and becoming increasingly popular during the 1990s. In contrast, only about half of all households were connected to the cable television network at that time. It’s roll-out concentrated in metropolitan areas where the per household cost of connection were lower. Yet, even within metropolitan areas, the cable television network was fragmented because landlords had to authorize and could refuse the necessary construction works. Besides, providing broadband service via existing cable television networks required substantial investment since wires had to be enabled to transmit signals bidirectionally. While this is essential for the provision of broadband services, television

Figure 2.1 : FL Connections in Germany



Note: Vertical bars illustrate the number of fixed-line connections in Germany between 2007-2017. The solid and dashed lines depict the fixed-line broadband market share of the German incumbent operator and its competitors between 2007-2017,respectively. Data: Bundesnetzagentur (2017)

Figure 2.2 : DSL Connections in Germany



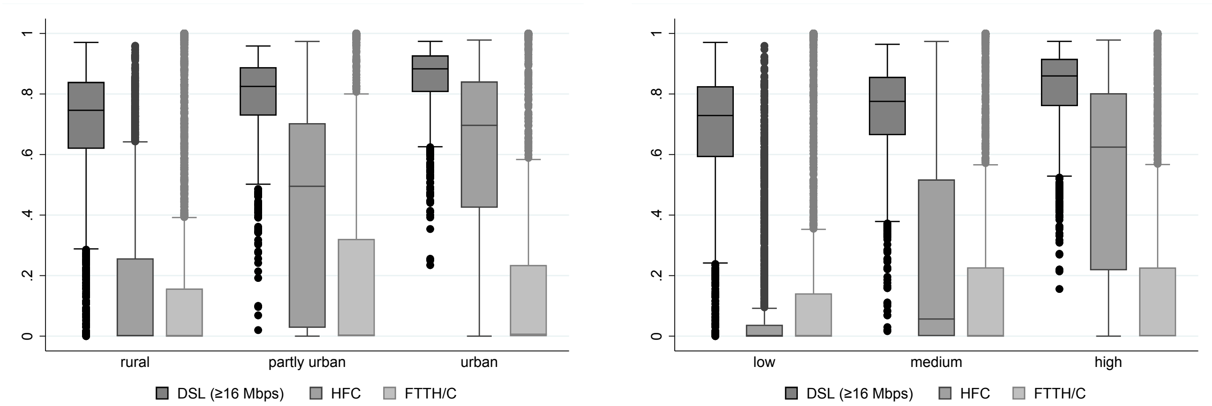
Note: Figure illustrates the share of DSL connections that are provided by the incumbent operator itself, and the share of DSL connections that are provided by competitors via bitstream access (Bitstream), physical access to subscriber lines (TAL), and resale (Resale) between 2007-2017. Data: Bundesnetzagentur (2017)

networks only need to transmit signals one-directionally from the program provider to the customer. Bidirectional network upgrading only started in the early 2000s, when Deutsche Telekom, who took over Deutsche Bundespost's telecommunications business during the liberalization and privatization of the telecommunications market during the 1990s, sold its cable television networks. (Bundesrechnungshof Deutschland, 2009)

During liberalization and privatization of the telecommunications market in Germany, the control over the PSTN was passed on to Deutsche Telekom (henceforth incumbent operator). At the same time a regulatory agency, the Federal Network Agency (BNetzA), was founded which in return obliged the incumbent operator to provide potential competitors access to its network. Until today BNetzA considers the incumbent operator as a company with significant market power (SMP). As such, it is subject to sector-specific regulation and obligated to provide competing service providers i) physical access to subscriber lines of the PSTN via local loop unbundling (LLU), and ii) bit-stream access. Both access options are provided against a fee which is subject to ex-post charge control by BNetzA. In 2017, 47% of all broadband connections via DSL area realized by competing internet service providers rather than the incumbent operator and network owner itself. While 50% of these rely on LLU, access via bitstream accounts for 25%. The remaining 25% are accounted for by resale, i.e. the sale of an incumbent's product by competitors under their own name. Figure 2.1 and Figure 2.2 illustrate that even though the incumbent's overall market share is in decline, its market share in DSL is rather stable over time. This suggests that its declining market share is mainly driven by inter-platform competition arising from cable (HFC) as well as FTTH/B networks. (Bundesnetzagentur, 2017)

Figure 2.3 : Inter-Platform Competition

(1) By Structural Type (2) By Population Density



Note: Figure illustrates box-plots of DSL, HFC, and non-incumbent FTTH/C coverage across (1) rural, partly urban and urban LEs and (2) LEs with low, medium and high population density in 2015. Data: Deutsche Telekom.

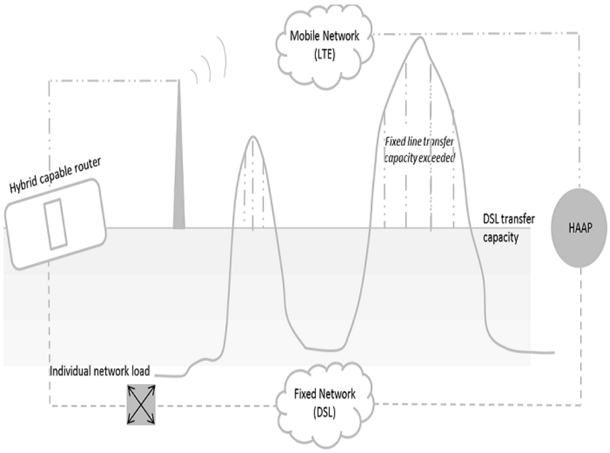
The extend of inter-platform competition varies substantially across Germany. Figure 2.3 illustrates the distribution of DSL, HFC and non-incumbent FTTH/C coverage across rural, partly urban, and urban LEs as well as LEs where population density is low, medium, and high. I cannot distinguish HFC and non-incumbent FTTH/C coverage by achievable bandwidth. Since both HFC and non-incumbent FTTH/C connections typically allow for a data downstream bandwidth beyond 16 Mbps, I focus on analyzing regional differences in inter-platform competition considering DSL coverage at a downstream bandwidth of at least 16 Mbps. Figure 2.3 (1) and (2) illustrate that inter-platform competition increases in urbanity and population density of LEs. For example, the share of households covered by HFC networks in the median LE is just above zero in rural LEs while it is nearly 70% in urban LEs. To cope with competition in urban and densely population regions, the incumbent operator invests in the roll-out of fiber wires to replace existing copper wires. However, replacing these wires is costly and time-consuming since wires have to be rolled-out subsurface. To temporary increase its competitiveness in the fixed-line market, the incumbent operator has launched a fixed-mobile hybrid product. It combines bandwidth from fixed-line DSL and mobile LTE to surf the internet at higher speed. In what follows, I explain this fixed-mobile hybrid product in more detail.

2.3 The Fixed-Mobile Hybrid Product

In November 2014[2], the German incumbent telecommunications operator launched fixed-line contracts for hybrid products, which allow its customers to combine bandwidths from fixed-line and mobile LTE networks to surf the internet *at home*. Technically, this is implemented by

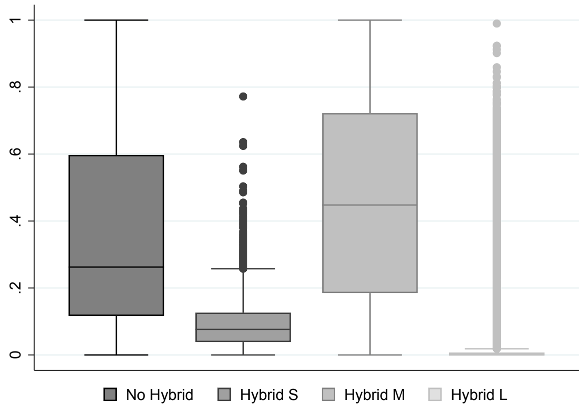
2 Hybrid product contracts were offered in some German regions in November 2014, prior to marketing them throughout Germany in March 2015.

Figure 2.4 : Hybrid Network Structure



Note: Figure illustrates how the provision of the hybrid product is organized. HAAP stands for Hybrid Access Aggregation Point. Illustration: Deutsche Telekom.

Figure 2.5 : Hybrid Coverage



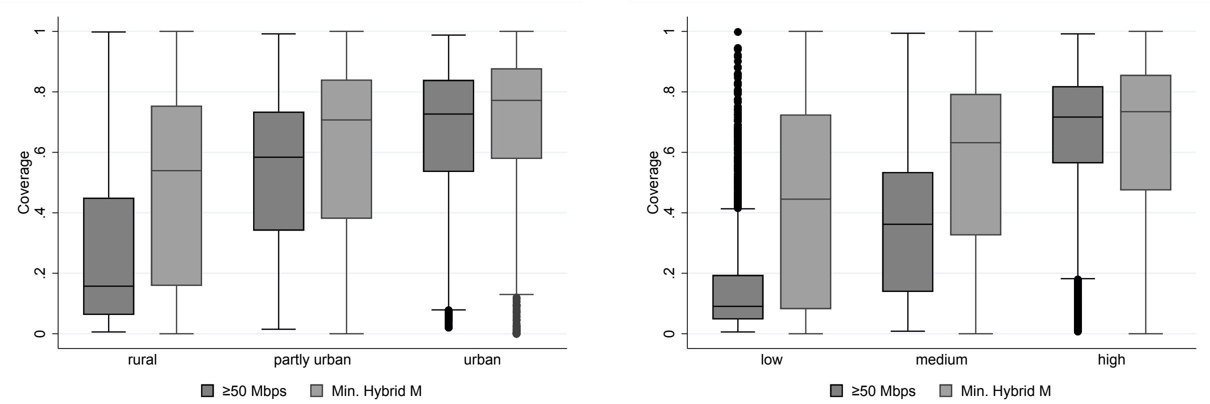
Note: Figure illustrates hybrid product coverage in Germany. Additional mobile bandwidth (in Mbps) for product type S: 16-≤50, M: 50-≤100, L: ≥100. Data: Deutsche Telekom.

installing a Hybrid Access Aggregation Point (HAAP) on the network side. Figure 2.4 illustrates that in the down-link, this HAAP unbundles the incoming data stream and distributes it to the two transmission paths. On the end-user side, a hybrid-capable router recombines the fixed-line and mobile data streams. Surfing via the fixed-line connection is default; bandwidth provided via LTE is not subject to volume restrictions but switched on only if i) the fixed-line connection has reached 95% of its maximum capacity, and ii) available LTE capacities are not required by mobile LTE users whose demand is prioritized at any time. Rather than permanently increasing available bandwidths, the additional LTE capacities thus primarily support the fixed-line connection during traffic peaks and in the event of malfunction. The fixed-mobile hybrid product allows customers reaching very low bandwidths via the fixed-line network to watch, for example, online video streams without infinite buffering times due to the additional mobile bandwidth provided.

The incumbent operator offers three different dual- and triple-play fixed-mobile hybrid contracts, differing only in terms of achievable bandwidth. While option S provides a downstream bandwidth of up to 16 Mbps via the fixed-line network and an additional 16 Mbps via the LTE network, option M (L) provides up to 50 Mbps (100 Mbps) via the fixed-line network and an additional 50 Mbps (100 Mbps) via the LTE network at most. Customers using product type S may thus use the internet at download speeds of up to 32 Mbps, while Hybrid M (L) customers may even use it at 100 Mbps (200 Mbps). Apart from these standard contracts, non-symmetric contracts are available, depending on the quality of the local LTE network and the technical parameters of the subscriber line. For example, customers who can receive at most 16 Mbps via their fixed-line access, may still receive an additional bandwidth of 50 Mbps (Speedoption M) or even 100 Mbps (Speedoption L) via LTE instead of only 16 Mbps in case of the standard type S hybrid contract. Figure 2.5 illustrates the distribution of the coverages of

Figure 2.6 : Hybrid Product Coverage

(1) By Structural Type (2) By Inter-Platform Competition Intensity



Note: Figures illustrate box-plots of the share of households within particular types of LEs that are covered by fixed-line broadband infrastructure at a data downstream rate of ≥50 Mbps. Min. Hybrid M indicates the share of households within a LE that can access at most 16 Mbps via the fixed-line network of the hybrid product offering operator and at least up to an additional bandwidth of 50 Mbps or even up to 100 Mbps via its LTE network, by booking a hybrid product contract. LEs divided into three groups according to local inter-platform competition intensity: low (≤33th percentile), medium(>33th & ≤66th), high (>66th percentile). Data: Deutsche Telekom, TÜV Rheinland (Breitbandatlas).

the different hybrid products across Germany. It shows that the option to combine fixed-line and mobile bandwidth is not available to 27% of all households in the median LE. About 8% of all households in the median LE are covered by hybrid products of type S and even 45% can top up their fixed-line connections with mobile bandwidth of up to 50 Mbps, respectively.

During the period under investigation, fixed-mobile hybrid contracts are generally offered at the same price as the respective standard fixed-line contracts which include the same services except the additional LTE bandwidth[3]. Differences in cost arise from the fact that a special hybrid-capable router is required. This router can be either bought at a price of 399.99 Euros or rented at a monthly fee of 9.95 Euros. Compared to customers with a standard fixed line contract, who may buy a standard router at a price of 159.99 or pay a monthly rental-fee of 4.94 Euros, hybrid customers thus face an additional monthly rental cost of 5.01 Euros.

Figure 2.6 illustrates the share of households that are covered by fixed-line networks with a bandwidth capacity of 50 Mbps or more and the share of households covered by fixed-mobile hybrid infrastructure such that hybrid products of type M or L are available. It shows that in

[3] Example: A customers with the type S hybrid contract who may reach a download speed of up to 16 Mbps via the fixed-line network and up to an additional 16 Mbps via the LTE network pays the same monthly fee as a customer with a standard fixed-line type S contract that only provides 16 Mbps via the fixed-line network but no additional bandwidths via LTE. Hybrid customers with a contract on the basis of 16 Mbps via the fixed-line network who want to get an additional bandwidth of 50 Mbps (100 Mbps) via the LTE network need to pay an additional 4.94 Euros (9.95 Euros). Accordingly, customers with a hybrid contract on the basis of 50 Mbps via DSL pay an additional 4.95 Euros to receive 100 Mbps via LTE rather than 50 Mbps only

the median rural LE, 54% of all households are covered by hybrid products of at least type M, compared to 71% in the medium LE in partly urban areas and 77% in the medium LE across urban areas. In absolute terms, hybrid products thus profit densely populated areas more than rural areas. Yet, considering how the introduction of the hybrid product has changed the number of households that can surf at download speeds of up to 50 Mbps, the situation is different. While in in the median rural LE, 54% of all households may surf with up to 50 Mbps, using the respective hybrid product, only 16% can surf at such speed via any kind of fixed-line network. In partly urban and urban areas, this difference is much smaller. Similarly, Figure 2.6 (2) illustrates that the percentage of households that can surf at up to 50 Mbps only due to hybrid product introduction is higher in LEs, where inter-platform competition intensity in the fixed-line market is low.

In the following, I examine whether and how the introduction of the fixed-mobile hybrid product has altered competition in the fixed-line market. This is a relevant issue not only for network operators. In view of the concern raised that network operators without own infrastructure may be no longer be able to compete with the incumbent operator if they cannot replicate the fixed-mobile hybrid product, it is also an issue that is practically relevant for regulatory authorities.

2.4 Competition Effects

To analyze competition effects of the introduction of the fixed-mobile hybrid product I exploit highly disaggregated data on 5200 German LEs, which I will present in more detail prior to outlining the estimation approach.

2.4.1 Data

The data at hand comprise highly geographically disaggregated information on i) incumbent and non-incumbent fixed-line and mobile coverage, ii) the number and demographic structure of the German incumbent operator's customers, and iii) demographic population structure and economic characteristics of all 5200 German LEs.

I derive detailed information on fixed-line as well as mobile broadband coverage across German LEs in 2016 from the incumbent's administrative database. I use this data to calculate the share of households that are covered by incumbent fixed-line networks that are providing download speeds between 6-≤16 Mbps, 16-≤50 Mbps, or ≥50 Mbps, respectively. The data provided by the incumbent's market intelligence unit further allows me to calculate the share of households within a LE that is connected to HFC networks, non-incumbent FTTH/C networks, and non-incumbent mobile LTE networks in 2016. Provider and technology independent information on broadband coverage stems from the German broadband atlas (Breitbandatlas Deutschland) which is provided by TÜV Rheinland on behalf of the Federal Ministry of Transport and Digital Infrastructure. It summarizes information on fixed-line as

Table 2.1 : Descriptive Statistics

	Mean	SD	Min.	Max.
Panel A: Incumbent Fixed-Line Coverage in 2016				
Cov. 6-≤16 Mbps (in %)	2.85	4.46	0	51.13
Cov. 16-≤50 Mbps (in %)	41.49	26.46	0	94.31
Cov. ≥50 Mbps (in %)	42.7	32.06	0	95.49
Min. Hybrid M (2015)	70.49	25.61	0	100
Panel B: Non-Incumbent Broadband Coverage in 2016				
Cov. HFC (in %)	58.65	32.56	0	97.81
Cov. FTTX (in %)	19.6	30.62	0	100
Cov. LTE op. A (in %)	29.98	31.89	0	100
Cov. LTE op. B (in %)	42.35	29.79	0	100
Cov. LTE 2 Mbps (2015)	92.31	14.10	0	100
Panel C: Demographic LE Characteristics in 2014				
Population* (in 1000)	12.47	37.99	0.09	1827.36
HH. Density	1237.13	701.17	123.32	3359.23
Fr. Pop. 35-64	42.96	2.35	33.73	60
Fr. Pop. ≥65	21.37	3.01	9.92	40.06
Fr. Pop. Female	50.89	.88	38.69	61.11

Note: Table shows household weighted local exchange area (LE) averages for LEs in the sample. The number of households in 2015 is used as weights. *not weighted. Data: Deutsche Telekom, TÜV Rheinland (Breitbandatlas), Regionalstatistik.

well as mobile broadband coverage from about 350 network operators across Germany. For the years between 2011 and 2016 the contributing operators report – on a non-mandatory basis – the number of private households that are connected to their network(s). This data on broadband coverage indicates the share of households within a municipality that has access to download speeds of at least 1 Mbps, 16Mbps, or 50 Mbps via fixed-line and 2 Mbps via mobile LTE. I use build-up area weights to convert this municipality level data to the LE level[4].

[4] In detail, I calculate the share of the build-up area of a municipality that contributes to the build-up area of a particular LE and multiply it with the number of households within a municipality that are covered by networks of a particular bandwidth. Collapsing these numbers on LE-level, I retrieve the number of households that are covered by networks of the same particular bandwidth within a LE. It is important to note that applying this procedure, I implicitly assume that broadband coverage is equally distributed across the populated space within a municipality.

Annual information on the total number of its customers[5] between 2011-2016 stems from the incumbent operator's administrative database. Importantly, my measure of overall customers comprises all fixed-line customers, i.e. customers that only use voice telephony services (single-play), voice telephony and broadband services (dual-play), or voice telephony, broadband and broadcasting services (triple-play). In 2016, 60% of the incumbent's fixed-line customers use dual- and triple-play plans compared to 40% with single-play plans. I can further distinguish fixed-line customers by age and calculate the number of young (18-34), middle aged (35-64), and older (≥65) customers.

I derive the population share of young (18-35), middle aged (35-64), and older (above 64) people from the regional database Germany (Regionaldatenbank Deutschland) for each of the 5200 LEs across Germany. Since population data is originally provided on municipality-level, I use built-up area weights to convert it to the level of LEs. Finally, I distinguish LEs according to their settlement structure into rural, partly urban and urban areas, following a definition by the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR).

Table 2.1 presents household weighted summary statistics of the data at hand. In view of the identification approach pursued later on, I exclude all LEs that coincide with county borders. As a result, the sample comprises 5142 out of 5200 LEs. Panel A illustrates that both incumbent fixed-line broadband coverage at a bandwidth between 16-≤50 Mbps and a bandwidth ≥50 Mbps is approximately 42% in the average household's LE. Instead, 3% are covered by incumbent broadband networks at a maximum data downstream rate between 6-≤16 Mbps while 13% can surf at a speed <6 Mbps. Household weighted average HFC coverage is nearly 59%, while non-incumbent FTTH/C networks cover nearly 20% of all households. Finally, Panel B illustrates that more than 70% of all households may use hybrid products of type M, providing additional bandwidth of up to 50 Mbps via LTE or even up to 100 Mbps via LTE to top up their fixed-line bandwidth in 2015.

2.4.2 Estimation Approach

I am interested in the relationship between fixed-mobile hybrid coverage and incumbent fixed-line market penetration. Therefore, I estimate:

$$\Delta(lnMP_{rt=1}, lnMP_{rt=0}) = \beta_0 lnMP_{rt=0} + \beta_1 hybrid_{rt=0} + \mathbf{X}'_{rt=0}\beta_2 + \gamma_c + \varepsilon_{rt} \tag{2.4.1}$$

where $\Delta(lnMP_{rt=1}, lnMP_{rt=0})$ is the percentage change in incumbent fixed-line market penetration, i.e. the share of incumbent fixed-line customers per households in LE r between t=0 (2014) and t=1 (2016). $lnMP_{rt=0}$ is the natural logarithm of market penetration in the initial period 2014. It accounts for mean reverting dynamics, i.e. the fact that switching between operators is subject to path dependence. The main explanatory variable of interest is *hybrid*. It measures the share of households within a LE that could use the fixed-mobile hybrid product. In addition to the bandwidth provided via DSL, this product provides up to 50 Mbps or even

[5] Only customers whose contract information includes age.

up to 100 Mbps per LTE to surf the internet at home. The fixed-mobile hybrid product was introduced throughout the entire country. Variation in the share of households that could use it thus mainly results from regional differences in the quality and load of the incumbent's LTE network.

Fixed-line broadband services are provided at a particular location, for example at a customer's network termination point. Investment decision of fixed-line operators are therefore driven, among others, by local demand, local competition and the local cost of network deployment. In contrast, investment decisions of mobile operators, who aim at providing their services nationwide are less likely to depend on local market characteristics. In particular, mobile operators are unlikely to invest in their network quality in response the competition being strong in the fixed-line market. Therefore, investigating the effect of the fixed-mobile hybrid product on market penetration arguably allows me to abstract from potential issues of reverse causality that are inherent in a regression of fixed-line broadband coverage on fixed-line market penetration.

However, like fixed-line networks mobile networks are deployed by profit-oriented network operators. Regional differences in mobile network quality thus unlikely occur at random. For example, mobile quality is likely to differ across regions depending on population size and household density. One the one hand, larger markets promise greater profits. On the other hand, higher household density tends to reduce the cost of network deployment per customer. Since this not only applies to mobile networks but also to fixed-line networks, mobile network quality is likely to be positively correlated with the quality of fixed-line networks and with competition in the fixed-line market. As a result, the estimate of interest would be biased.

I take such differences in to account by including into the regression: the natural logarithm of a LE's population and household density, measured as the number of households per built-up area in a LE. Besides, I control for fixed-line competition intensity by including HFC and non-incumbent FTTH/C network coverage. I further account for regional differences in the quality of the fixed-line network of the incumbent operator. In detail, I account for the share of households within a LE that are covered by incumbent fixed-line networks providing a data downstream rate of 6-≤16 Mbps, 16-≤50 Mbps, and 50 Mbps or more, respectively.

As suggested by Chapter 1, the market for broadband internet access may include both fixed-line technologies as well as mobile technologies. This implies that a reduction in incumbent fixed-line market penetration could result from fixed-to-mobile substitution, i.e. incumbent fixed-line customers abandoning their fixed-line access in favour of using incumbent or non-incumbent mobile services. This is problematic with regards to the validity of our estimate of interest, if fixed-to-mobile substitution depends on the quality of mobile networks. Controlling for LTE coverage of all German non-incumbent mobile operators, I account for cross-channel competition between incumbent fixed-line and non-incumbent mobile services. Yet, I cannot control for cross-channel competition between incumbent fixed-line and incumbent mobile services since regional differences in incumbent LTE coverage coincide with regional differences in our main explanatory variable of interest. Consequently, if incumbent fixed-line customers switch to incumbent mobile services depending on regional differences in the quality of the incumbent mobile network, $\hat{\beta}_1$ is downward biased. However, concerns in this regard are mitigated by the Federal Network Agency (BNetzA) stating that mobile broadband services are generally not perceived as a substitute for fixed-line broadband services in Germany (Bundesnetzagentur, 2015).

Finally, I include county fixed-effects, γ_c, to account for unobservable aggregate differences in population characteristics across regions, and for shocks that are shared by LEs within the same county. I also account for the composition of the population with respect to age, education and gender to account for demographic differences across LEs more specifically. Eventually, β_1 is identified from differences in fixed-mobile hybrid product coverage across LEs that are comparable otherwise, in particular in terms of fixed-line network and non-incumbent mobile network quality.

The provision of additional bandwidth via LTE to hybrid product customers is not guaranteed. It is only available if mobile customers (rather than hybrid customers) do not use the entire bandwidth that is available in a particular cell. Even though my data on hybrid coverage takes mobile network load into account, actually achievable bandwidth of hybrid product users across regions is likely measured with error. This is also due to the fact that the quality of mobile connections hinges on the direct visual contact between hybrid capable router and base station antenna and thus depends on the positioning of the router within a particular household as well as the local surrounding. For example, trees, building characteristics, and the positioning of surrounding buildings, among others affect the quality of the mobile connection and thus achievable bandwidth. While some of these factors are in principal observable by the incumbent operator, others are not. This makes it difficult for the incumbent operator to assess actual coverage rates. Classical measurement error is known to attenuate point estimates, i.e. bias them towards zero. In case of a positive effect, this would imply that estimates reflect a lower bound effect. To derive the true effect, I pursue an instrumental variables strategy where the first stage is as follows:

$$hybrid_{rt=0} = \theta_0 lnMP_{rt=0} + \theta_1 mobileLTE_{rt=0} + \mathbf{X}'_{rt=0}\theta_2 + \gamma_c + \varepsilon_{rt} \tag{2.4.2}$$

As instrument for hybrid product coverage, I use joint mobile LTE coverage of all mobile network operators in Germany. Even though this measure of mobile LTE coverage may itself be measured with error, estimating Equation 2.4.2 allows me to reduce or even remove attenuation bias as long as measurement error in hybrid product coverage is not perfectly correlated or even uncorrelated with the measurement error that is inherent in the instrumental variable. In this regard it is important to note that the data on the instrumental variable is derived from the German broadband atlas, while information on hybrid product coverage is gathered from the incumbent operator's administrative database. This suggests that the sources of measurement error are different. Using this measure of mobile LTE coverage may thus serve as an instrument to reduce measurement error. Importantly, pursuing this instrumental variables approach, I only reduce concerns regarding measurement error. However, it does not provide exogenous variation in a sense that would make it obsolete to thoroughly control for local market characteristics, in particular competition intensity.

2.4.3 Results

Baseline Results

Table 2.2 illustrates the baseline results from OLS and IV estimation[6]. The sample comprises all 5142 German LEs, whose borders do not coincide with county borders. I weight all regressions with the number of households in 2015. Therefore, the estimates reflect the average household's probability to give-up, keep or conclude a fixed-line contract with the incumbent operator in response to the introduction of the hybrid product. All specifications include county fixed-effects and standard errors are clustered at the same level. To account for potential mean-reverting dynamics, all specifications include the natural logarithm of the lag dependent variable. In Columns (2)-(4), I additionally control for market size and households density. To account for competition intensity, I include non-incumbent HFC coverage, non-incumbent FTTH/C coverage and non-incumbent mobile LTE coverage throughout Columns (3)-(4). Since the relevant data is available only for the year 2016, these controls are a proxy for actual competition intensity in 2014. Throughout Columns (3)-(4) I also include the share of households that are covered by the incumbent's fixed-line network at a maximum data downstream rate of 6-≤16 Mbps, 16-≤50 Mbps, and ≥50 Mbps. Again I rely on data from 2016 as a proxy for coverage rates in 2014. In Column (4), I further control for differences in the composition of the population across LEs, including the share of female, the share of young (18-34), middle aged (35-64), and older (≥65) as control variables. Panel A presents results from the OLS estimation. Panel B presents IV results taking potential measurement error in our main explanatory variable into account.

Estimates are positive and statistically significant at conventional levels in the more comprehensive specifications of Column (3) and Column (4). Accounting for regional differences in market size, population density, competition intensity, and the composition of the population across LEs increases the point estimate on the main explanatory variable of interest. This is in line with a downward bias emerging from the fact that incumbent mobile network quality is correlated with fixed-line network quality and competition in the fixed-line market. Throughout all specifications in Table 2.2, IV estimates are larger than the corresponding OLS estimates. Even though differences are not statistically significant, this suggests that the main explanatory variable of interest, Min. Hybrid M (2015), is measured with error. OLS estimates are thus subject to small attenuation bias. Therefore, I concentrate on IV estimates to calculate and quantitatively interpret the results.

I first consider the most comprehensive specification in Column (4) of Panel B in Table 2.2. It indicates that the percentage change in incumbent market penetration between 2014 and 2016

[6] Coefficients on all control variables are provided in Tables B.4 and B.5 in the Appendix

Table 2.2 : Baseline Results

	(1)	(2)	(3)	(4)
Panel A: OLS Estimation				
Min. Hybrid M (2015)	0.0120***	0.0128***	0.0152***	0.0158***
	(0.0043)	(0.0045)	(0.0044)	(0.0044)
Log FL p.hh (2014)	-0.0065	-0.0085	-0.0438***	-0.0468***
	(0.0079)	(0.0089)	(0.0104)	(0.0106)
Panel B: IV Estimation, accounting for Measurement Error				
Min. Hybrid M (2015)	0.0150	0.0169	0.0227**	0.0243**
	(0.010)	(0.0114)	(0.0108)	(0.0108)
Log FL p.hh (2014)	-0.00523	-0.00820	-0.0437 ***	-0.0468***
	(0.00808)	(0.00859)	(0.0100)	(0.0103)
County FE	Yes	Yes	Yes	Yes
Lag. Dep.	Yes	Yes	Yes	Yes
Market Charac.	No	Yes	Yes	Yes
Broadband Infr.	No	No	Yes	Yes
Socio-Econ. Charac	No	No	No	Yes
LEs	5142	5142	5142	5142

Note: Dependent variable is the percentage change in incumbent fixed-line market penetration between 2014-2016. Min. Hybrid M is the share of households that receive up to an additional 50 Mbps or even up to an additional 100 Mbps via LTE. Lag. Dep. is the natural logarithm of customer p.hh. in 2015. Market Charac. include the natural logarithm of population in 2014, and household density in 2014. Broadband Infr. includes HFC-, non-incumbent FTTH/C, non-incumbent LTE-coverage in 2016, and incumbent fixed-line coverage at a data downstream rate between 6-≤16 Mbps, 16-≤50 Mbps, ≥50 Mbps in 2016. Socio-Econ. Charac. include share female, and share of population between 18-34, 35-64, ≥65 in 2014. All regressions are weighted with the number of households in 2014. Standard errors clustered at the county level. Significance levels: $^{*}p < 0.1,^{**} 0.05,^{***} p < 0.01$.

is 2.43 percentage points larger in a LE, where all rather than none of the households in a LE can connect to the internet, using the fixed-mobile hybrid product. Household-weighted hybrid coverage in the average LE amounts to 70.49%. Consequently, the introduction of the hybrid product increased the rate of change in incumbent fixed-line market penetration between 2014 and 2016 by 1.71 percentage points in the average household's LE. This corresponds to an annual effect of 0.86 percentage points. To get a better grasp of the economic meaning of the effect size, I compare it to the average annual percentage change in incumbent fixed-line market penetration between 2011 and 2014. During this time, annual decline in incumbent fixed-line market penetration amounted to approximately 2.53%. Consequently, the annual effect of hybrid product introduction (0.86 percentage points) corresponds to 33.99% of the annual rate of decline between 2011 and 2014.

The Role of Bandwidth among Different Population Groups

Among others, Lengsfeld (2011), Michailidis et al. (2011), Mills and Whitacre (2003), and van Dijk and Hacker (2003) have found age to be a substantial determinant of information and communication technology (ICT) usage. This suggests that age may also determine adoption of broadband connections providing higher bandwidth. In the following, I examine whether the effect of the introduction of the fixed-mobile hybrid product differs across age-groups. To that end, I distinguish incumbent fixed-line customers by age into three groups of young (18-34), middle aged (35-64), and older (≥65) customers. The dependent variables measures the age-group specific percentage change in incumbent fixed-line market penetration between 2014 and 2016. I calculate the age-group specific market penetration as the share of incumbent fixed-line customers in an age group relative to the population in that age group. I also adapt weighting, using the population in the respective age group in 2014 as weight rather than households. Panel A and Panel B of Table 2.3 present OLS as well as IV results for all LEs with at least one customer in each age group in 2014. This reduces the number of LEs in the sample from 5142 to 5129. Column (1) presents results from the most comprehensive specification of Column (4) in Table 2.2 for this restricted sample. Columns (2) to (4) present results for the age group indicated in the Column heads.

Table 2.3 illustrates that the effect of fixed-mobile hybrid coverage is particularly strong among the younger population group. Among the young, the rate of change in incumbent fixed-line market penetration is 5.06 percentage points higher in a LE where all households are covered by the fixed-mobile hybrid product compared to a LE were not a single one is covered. Among middle aged and older population groups instead, the difference is 1.77 percentage points and 1.74 percentage points, respectively. While the results for young and older customers are both relevant in terms of their size, and in terms of statistical significance, I do not find a statistically significant result with regards to the coefficient for the medium aged. Hybrid coverage in the average LE of young, middle aged and older incumbent fixed-line customers amounts to 71%, 69.08% and 68.47%, respectively. Consequently, hybrid product introduction increases the annual rate of change in incumbent market penetration between 2014 and 2016 by 1.80 percentage points in the average LE of younger incumbent fixed-line customers, by 0.61 percentage points in the average LE of middle aged incumbent fixed-line customers, and by 0.60 percentage points in the average LE of older incumbent fixed-line customers. Overall, this suggests that bandwidth is relatively important for younger population groups in their decision for or against a particular fixed-line contract (and operator). The results further indicate that in order for fixed-line operators to remain competitive in future, investments in the quality of their networks in terms of bandwidth are essential.

Table 2.3 : Effect Heterogeneity by Age

	(1) ≥18	(2) 18-34	(3) 35-64	(4) ≥65
Panel A: OLS Estimation				
Min. Hybrid M (2015)	0.0149***	0.0388***	0.0169***	0.0122***
	(0.0047)	(0.0095)	(0.0061)	(0.0037)
Log FL p.P. (2014)	-0.0425***	-0.115***	-0.0333***	-0.0812***
	(-5.25)	(-10.41)	(-3.46)	(-12.75)
Panel B: IV Estimation, accounting for Measurement Error				
Min. Hybrid M (2015)	0.0183	0.0506***	0.0177	0.0174**
	(0.0112)	(0.0185)	(0.0149)	(0.0075)
Log FL p.P. (2014)	-0.0428***	-0.116***	-0.0334***	-0.0814***
	(-5.22)	(-10.47)	(-3.43)	(-12.70)
County FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
LEs	5129	5129	5129	5129

Note: Dependent variable is the percentage change in incumbent fixed-line market penetration in the respective age group between 2014-2016. Min. Hybrid M is the share of households that receive up to an additional 50 Mbps or even up to an additional 100 Mbps via LTE. *Controls* include: natural logarithm of customers in the respective age group relative to the population in the respective age group in 2014, natural logarithm of population in 2014 and household density in 2014, HFC-, non-incumbent FTTH/C, and non-incumbent LTE-coverage in 2016, incumbent fixed-line coverage at a data downstream rate of 6-≤16Mbps, 16-≤50Mbps, and ≥50Mbps in 2016, share female, and share of population between 18-34, 35-64, ≥65 in 2014. All regressions are weighted with population in the respective age group in 2014. Standard errors clustered at the county level. Significance levels: $^{*}p < 0.1,^{**}0.05,^{***}p < 0.01$.

Regional Differences in the Role of Bandwidth in Shaping Competition

Regional imbalances in broadband coverage as well as broadband usage have attracted substantial interest both from policy makers as well as academics. Among others, Salemink et al. (2017), Townsend et al. (2013), and Ellershaw et al. (2009) have discussed the role of deployment cost and socio-economic factors in shaping differential broadband access and usage across rural and urban areas. Besides, overcoming the rural-urban digital gap is one of the main targets of the Digital Agenda for Europe that aims at providing all households across Europe access to connectivity offering at least 100 Mbps by 2025. To provide sufficient incentives for infrastructure investments into high-speed broadband networks, the European Commission has published a proposal of an European Electronic Communications Code, revising the existing regulatory framework in the telecommunications sector. In summary, it suggests a set

Table 2.4 : Effect Heterogeneity by Inter-Platform Competition Intensity

		Inter-Platform Competition Intensity		
	(1)	(2)	(3)	(4)
	Overall	Low	Medium	High
Panel A: OLS Estimation				
Min. Hybrid M (2015)	0.0158***	0.0140***	0.0093*	0.0171**
	(0.0044)	(0.0043)	(0.0048)	(0.0075)
Log FL p.hh. (2014)	-0.0468***	-0.0519***	-0.0417***	-0.0543***
	(0.0103)	(0.0155)	(0.0132)	(0.0155)
Panel B: IV Estimation, accounting for Measurement Error				
Min. Hybrid M (2015)	0.0243**	0.0148*	0.0175*	0.0352**
	(0.011)	(0.009)	(0.010)	(0.018)
Log FL p.hh. (2014)	-0.0468***	-0.0520***	-0.0414***	-0.0532***
	(0.0103)	(0.0155)	(0.0131)	(0.0157)
County FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
LEs	5142	1696	1698	1748

Note: Dependent variable is the percentage change in incumbent fixed-line market penetration between 2014-2016. Min. Hybrid M is the share of households that receive up to an additional 50 Mbps or even up to an additional 100 Mbps via LTE. *Controls* include: natural logarithm of customers p.hh. in 2014, natural logarithm of population in 2014, household density in 2014, HFC-, non-incumbent FTTH/C-,and non-incumbent LTE-coverage in 2016, incumbent fixed-line coverage providing data downstream rates of 6-≤16 Mbps, 16-≤50 Mbps, and ≥ 50Mbps in 2016, share female, and share of population between 18-34, 35-64, ≥65 in 2014. LEs dividied into three groups according to local inter-platform competition intensity: low (≤33th percentile), medium(>33th & ≤66th), high (>66th percentile). All regressions are weighted with the number of households in 2014. Standard errors clustered at the county level. Significance levels: $^{*}p < 0.1,^{**} 0.05,^{***} p < 0.01$.

of measures to soften regulatory intervention and reduce regulatory uncertainties, particularly with regards to high-capacity broadband network deployment. Briglauer et al. (2017) provide a discussion of the specific measures proposed. Investigating potential differences in the role of bandwidth in shaping competition in different types of regional markets, I contribute to this discussion in two important dimensions. First, my analysis provides insights on how improvements in bandwidth affect market penetration. This is important since gaining larger shares of the market is an important incentive for network operators to invest. Second, it contributes to the question of whether quality leadership in terms of bandwidth might lead to market monopolization in regions where infrastructure based competition is low. This is important for regulatory authorities , when deciding about how to specifically design potential regulatory policies.

In what follows, I investigate how the introduction of the fixed-mobile hybrid product affects the rate of change in incumbent market penetration in different types of regional markets. Table 2.4 presents OLS as well as IV results in Panel A and Panel B, respectively. While Column (1) illustrates our baseline results for all LEs across Germany, Columns (2), (3) and (4) indicate the effect of hybrid product coverage in LEs, where infrastructure-based competition intensity is low, medium, and high, respectively. Unfortunately, I do not have information on the regional distribution of broadband service providers without own infrastructure to the end customer. In some LEs, where infrastructure based-competition is very low, also service-based competition may be low. Yet, three broadband operators, providing broadband services via own infrastructure, local loop unbundling or bit-stream access claim to provide their services nearly all over Germany". Consequently, extreme cases, where the incumbent operator is the only broadband provider in a LE, and hybrid product introduction thus can not affect provider choice by construction should be rare.

Overall, results in Panel B of Table 2.4 indicate that the effect of the introduction of the fixed-mobile hybrid product is increasing in local competition intensity. For example, Column (2) indicates that in LEs where competition intensity is low, incumbent fixed-line market penetration increased by 1.48 percentage points more when all households compared to any households are covered by fixed-mobile hybrid infrastructure. In a LE with medium and high competition intensity instead, the effect is 1.75 percentage points and 3.52 percentage points, respectively. The fixed-mobile hybrid product covers 50.82%, 62.94%, and 76.05% in the average household's LE with low, medium and high competition intensity, respectively. Consequently, the introduction of the fixed-mobile hybrid product increases the annual rate of change in incumbent fixed-line market penetration between 2014 and 2016 by 0.38 percentage points in LEs, where competition intensity is low, and by 0.55 (1.34) percentage points in LEs, where competition intensity is medium (high).

Broadband service providers without own infrastructure raised concerns that they would not be able to compete with the incumbent operator if the latter is not obliged to provide an access product that allows them to replicate the characteristics of the fixed-mobile hybrid product. However, the results of this paper suggest that the ability to provide higher bandwidth than broadband service providers without own infrastructure increases the rate of change in incumbent fixed-line market penetration to a limited extend. In contrast to LEs where inter-platform competition is strong, demand in LEs with low infrastructure based competition seems to be less responsive to a rise in bandwidth. On the one hand this suggests that presently, unilateral improvements in achievable bandwidth unlikely lead to market monopolization in markets without infrastructure-based competition. On the other, the limited effect of bandwidth on market penetration points towards the willingness to pay for higher bandwidth being small. This reduces incentives to invest, such that private investments in broadband infrastructure are likely to concentrate in more urban areas, where the ability to offer high-capacity internet connections is essential to remain competitive.

Table 2.5 : Robustness Tests & Placebo Estimate

	Robustness			Placebo
	(1) Baseline	(2) Struc. Cust.	(3) Outlier	(4)
Panel A: OLS Estimation				
Min. Hybrid M (2015)	0.0158***	0.0155***	0.0111***	0.0038
	(0.0044)	(0.0044)	(0.0030)	(0.0028)
Log FL p.hh. (2014)	-0.0468***	-0.0746***	-0.0282***	-0.0018
	(0.0103)	(0.0109)	(0.0076)	(0.0088)
Panel B: IV Estimation, accounting for Measurement Error				
Min. Hybrid M (2015)	0.0243**	0.0232**	0.0161***	-0.0007
	(0.011)	(0.011)	(0.006)	(0.005)
Log FL p.hh. (2014)	-0.0468***	-0.0745***	-0.0282***	-0.0019
	(0.0103)	(0.0109)	(0.0076)	(0.0088)
County FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Cust. Struc.	No	Yes	No	No
LEs	5142	5142	5038	5142

Note: In Columns (1)-(3) (Column (4)), dependent variable is the percentage change in incumbent fixed-line market penetration between 2014-2016 (2011-2013). Min. Hybrid M is the share of households that receive up to an additional 50 Mbps or even up to an additional 100 Mbps via LTE. *Controls* throughout Column (1)-(3) (Column (4)) include: natural logarithm of customer p.hh. in 2014 (2011), natural logarithm of population in 2014 (2011), household density in 2014 (2011), HFC- and non-incumbent FTTH/C, and non-incumbent LTE-coverage in 2016 (2016), incumbent fixed-line coverage at a data downstream rate of 6-≤16 Mbps, 16-≤50 Mbps, ≥50 Mbps in 2016 (2016), share female, and share of population between 18-34, 35-64, or ≥65 in 2014 (2011). All regressions are weighted with the number of households in 2014 (2011). Standard errors clustered at the county level. Significance levels: $^{*}p < 0.1,^{**} 0.05,^{***} p < 0.01$.

2.4.4 Robustness Tests & Placebo Estimates

I argue that the results derived represent the causal effect of bandwidth on market penetration, once controlling for local market environment and population characteristics. To substantiate this claim, I provide further robustness checks and placebo estimates in Table 2.5. Column (1) repeats our most comprehensive specification from Table 2.2. Column (2) additionally controls for the incumbent operators customer structure in terms of age in 2014. Including these additional controls should not alter the effect significantly, if hybrid product roll-out has taken place independent of local fixed-line market conditions. Column (3), drops the top and bottom percentile of LEs with regards to the percentage change in market penetration between 2014 and 2016. I find that both robustness tests do not qualitatively alter the results.

To further elaborate on the validity of the results, I pursue a placebo test. The idea is to estimate the same model as in Equation 2.4.1 but for a time period closely related to but before hybrid products were introduced to the market. In detail, I regress the percentage change in market penetration between 2011 and 2013 on hybrid product coverage in 2015. Similar to the main model, I include baseline control variables, such as the lag of the dependent variable, market size and household density into the regression. Since I have data on PSTN, HFC and non-incumbent FTTH/C coverage at hand only for 2016, I include these variables as a proxy for competition intensity and overall fixed-line coverage in 2011. Column (4) of of Table 2.5 presents results of the placebo test. It illustrates that fixed-mobile hybrid product coverage does not statistically significantly effect the rate of change in incumbent market penetration between 2011 and 2013. With a value of -0.0007 the estimate is quantitatively close to zero and significantly smaller than our baseline estimate. Overall, this is reassuring with regards to the validity of the results presented.

2.5 Conclusion

During the last decade, telecommunications markets have changed substantially. Among others, mobile phones have penetrated the market, new services have spurred competition among voice telephony service providers, and fixed-line market demand is increasingly focusing on the available bandwidth to connect to the internet. Besides, new (fixed-line) operators are entering the market, gaining an increasing share of the market.

In response to the increasing competition, the German incumbent operator has launched a fixed-mobile hybrid product that combines bandwidth from its fixed-line DSL and mobile LTE networks to surf the internet at higher speed. This paper analyzes how the introduction of this product is shaping competition in the German fixed-line broadband market.

The results suggest that the introduction of the fixed-mobile hybrid product reduces annual decline in incumbent fixed-line market penetration by 0.86 percentage points. This corresponds to 33.99% of the average annual decline in market penetration (-2.53%) faced by the incumbent operator between 2011 and 2014. Moreover, the results indicate that effects are substantially stronger among younger population groups and in regions where infrastructure-based competition is strong. Overall the results point towards the importance of bandwidth in future, with younger population groups growing older. Besides, they also suggest that quality is essential to remain competitive in regions where infrastructure-based competition is strong.

The results of this paper make two important contributions with regards to the discussion on how to incentivize private investments in high-capacity broadband networks. First, comparably small effects in regions, where infrastructure-based competition is low indicate that private investments in high-capacity broadband networks in rural areas are unlikely to happen in the absence of state-funding. Even regulatory exemptions may not provide sufficient incentives. Second, the results indicate that quality leadership in terms of bandwidth is unlikely to lead to market monopolization, even in regions where infrastructure based competition is small. Yet, given substantially larger effects for younger customer groups this may change in future

3 Online Shopping and Retail Employment: Evidence from Brick-and-Mortar Bookstores*

3.1 Introduction

The retail industry has undergone substantial change over the past two decades. While ever prospering malls have been the symbol of successful retailing in the United States for years, they now often fail to lease out their vacant spaces or even shut down entirely. In various European regions, the decline of brick-and-mortar retailers has alerted some local politicians and real estate developers so much that they are already prophesying deserted city centers and high streets.

Although this situation is often blamed on the rise of online shopping, rigorous empirical evaluations of this relationship are surprisingly scant. To contribute towards closing this gap in the literature, we investigate the causal effect of regional exposure to online shopping on brick-and-mortar bookstores in Germany between 1999 and 2013. Our analysis is based on a novel data set that combines detailed geomarketing information on regional shopping behavior with administrative records from social insurance. We find that regional exposure to online shopping has substantial negative effects on both the stock of local bookstores and their employees. For example, an increase in regional exposure to online shopping by one standard deviation reduces the stock of brick-and-mortar bookstores by about 0.11 bookstores per 10,000 residents (i.e., 14% relative to 1999). With regard to employment, we find effects of similar magnitude.

In a regional context, identifying the causal effects of exposure to online shopping on brick-and-mortar bookstores gives rise to various endogeneity concerns. For example, availability and usage of the internet is likely influenced by the same individual and regional characteristics that also drive online shopping and may themselves have a direct effect on reading. More generally, any unobserved regional shock or differences in the composition of the local population, as well as agglomeration structures may affect both brick-and-mortar retailers and the uptake of online shopping. Additionally, reverse causality may be a problem if an insufficient regional shopping infrastructure forces consumers to buy online.

To address these endogeneity concerns, we estimate a long difference model in combination with an instrumental variables (IV) strategy. The central idea of our IV approach is to construct a Bartik-type instrument that exploits differences in local age structures before the internet was introduced to explain regional variation in the uptake of online shopping. Within our

* This chapter is based on joint work with Tobias Lohse, and Bastian Stockinger

long difference model, this approach controls for the effects of any time-invariant factors, unobserved shocks that are uncorrelated with (pre-internet) local age structures, and measurement error. While other endogeneity concerns may persist, we argue that our IV approach is valid, conditional on controlling for agglomeration dynamics and changes in online usage over time. Robustness checks, including a large set of socio-demographic control variables, confirm the validity of our results.

This paper relates to the empirical literature that investigates how supermarket entry affects local retail employment, as well as the survival and size of incumbent retailers. For example, Neumark et al. (2008), Jia (2008), and Basker (2005) use U.S.-county-level data to investigate local retail employment effects of Walmart entry. Even though Basker (2005) finds positive employment effects in the short run, all three papers show that long-run effects on employment as well as the survival of other (incumbent) retail stores are negative. A similar study by Zhang and Lei (2015) uses city-level data. Borraz et al. (2014) and Igami (2011) instead define very small geographic markets within cities to study the effect of supermarket entry on existing stores. They mostly confirm the previously discovered negative effects on employment and establishment survival. Igami (2011) documents that, unlike large and medium-sized supermarkets, small ones actually profit from entry of large supermarkets as their product portfolios are sufficiently different or can even serve as complements.

Online shopping likely distorts local equilibria even more than supermarket entry since online retailers are not as locally attached as supermarkets. While supermarkets create local employment, online retailers may generate employment in an area far away from the markets served. This possibility is investigated by Goolsbee (2000), Ellison and Ellison (2009), Goolsbee et al. (2010), and Einav et al. (2014), who find that online demand is less dependent on geographic proximity of sellers. Factors such as price may divert demand from the place of residence to other regions. Specifically, they show that customers tend to shop more often in U.S. states where sales taxes are lower.

We also contribute to the literature that analyzes local labor market effects of technological change and digitization. For example, Acemoglu and Restrepo (Acemoglu and Restrepo) investigate the effect of robot usage on local industrial employment, using a Bartik-type instrument that combines variation in historic industry shares across U.S. commuting zones and aggregate trends in robot adoption by industry. Similarly, Graetz and Michaels (2018) and Dauth et al. (2017) exploit a Bartik-type instrument to investigate labor market effects of robot adoption in a panel of more than 17 countries as well as in regional labor markets across Germany. Similar to other studies, for example, Akerman et al. (2015) or Atasoy (2013), they investigate how (different groups of) workers in manufacturing are affected by technological change. In contrast, we focus on employment effects in retail, a sector rarely studied. We are mainly interested in the effect of online retail on retail markets, rather than particular groups of employees. Therefore, our results extend the existing literature substantially, which will be particularly useful for urban planners, policy-makers, brick-and-mortar retailers, and real estate developers, among others, who are interested in understanding how high streets will evolve in future.

The remainder of the paper proceeds as follows. Section 3.2 reviews the German book market between 1999 and 2013. Section 3.3 describes the data we use in this study. Section 3.4 outlines our empirical approach and discusses the validity of our instrument. Section 3.5 presents the results of our analysis and Section 3.6 concludes.

3.2 The German Book Market Between 1999 and 2013

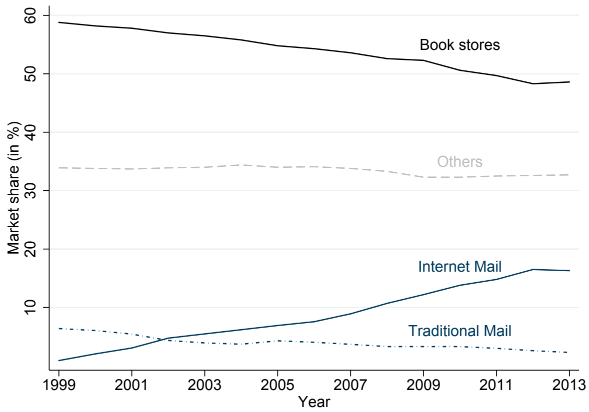
Figure 3.1 illustrates the importance of different book distribution channels between 1999 and 2013. Over the whole period, brick-and-mortar retail was the primary distribution channel through which private customers obtained their books. In 1999, sales in traditional bookstores accounted for 59% of total market turnover, whereas not even 1% of German book sales took place online.[1] At the same time, roughly 6% of all books were sold via telephone or catalogs. All other books were sold by warehouses that do not specialize in selling books (9%), other selling points like kiosks or gas stations (5%), book clubs (4%), and publishers that sell directly to institutional customers or private companies (16%).

Due to the rise of online retailers, neither brick-and-mortar bookstores nor companies selling books via telephone or catalogs were able to maintain their initial market shares. The revenues of brick-and-mortar retailers decreased by more than 10 percentage points between 1999 and 2013; sales made via telephone or catalogs saw even sharper decreases and only accounted for 2.3% of the market in 2013. In contrast, online retailers gained substantially over the same period. The first online bookseller to enter the German market was ABC Bücherdienst, which began its online services in 1996. Over the next few years, others followed. However, the online market soon consolidated as many of the initial entrants had already disappeared by 2003. Among those surviving was Amazon, which grew to be the most important player in the German online market for books during the following decade. In 2013, Amazon accounted for almost three-quarter of the 19% market share for online retailers (Schrape, 2011).

Regarding competition between online and offline retail, the economic literature so far mostly focuses on price differences and product differentiation. Investigating price differences across channels, Brynjolfsson and Smith (2000) and Cavallo (2017) show that, depending on the product considered, online prices tend to be lower. For the computer industry, these price differences are particularly important as Goolsbee (2001) finds that online computer demand is highly elastic to offline price changes. In our context, however, differences between online and offline play a negligible role for cross-channel competition in the German book market as resale prices of books are regulated by the fixed book price law, the "Buchpreisbindungsgesetz" (BPBG).

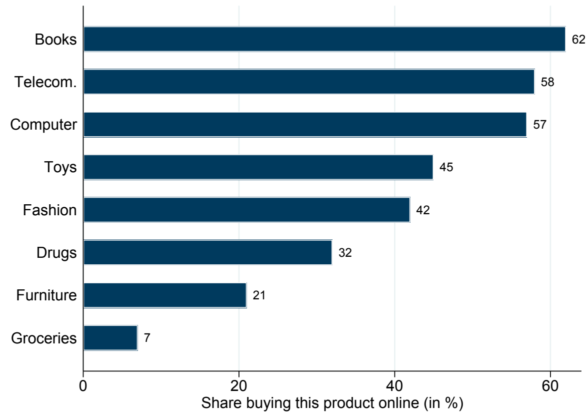
[1] This also includes the books that brick-and-mortar retailers sold online.

Figure 3.1 : Book sales by Retail Channel



Note: This figure illustrates shares of sales for four different book distribution channels between 1999 and 2013. Data: Langendorf (2003), Cronau (2006, 2011, 2013, 2014).

Figure 3.2 : Online shopping Preferences



Note: This figure presents the results of a survey among 1,003 Germans aged 16–69 who were asked whether they prefer to buy the respective products online or offline. Data: Bundesverband E-Commerce und Versandhandel Deutschland (2017), Bundesverband E-Commerce und Versandhandel and Boniversum (2017).

The BPBG requires anyone who publishes or imports non-foreign-language books to set a common retail price to be paid by the final customers. These prices have to be made publicly available and cannot be changed by the outlets eventually selling the books (BPBG 2002, §5). Following a series of book price agreements since 1888, the BPBG was enacted in 2002 to safeguard books as cultural goods of particular value that should be ubiquitously available to the general public. To this end, the BPBG also makes explicit that it is in favor of a large number of retail outlets (BPBG 2002, §1). While this law does not affect the resale of used books, it applies to all other books, including cartographic products, music sheets, and foreign-language books that are primarily sold in Germany (i.e., dictionaries).[2] As a result, online and offline prices for new books are identical in Germany.

Publishers are allowed to set different prices for different types of books. For example, they may charge different prices for hard-cover copies, soft-cover copies, or the e-book version of a book. If online retailers primarily sold e-books, there would be price differences across online and offline retail channels (if e-books are considered substitutes for printed books). However, as e-books accounted for only 0.5% (3.9%) of the overall consumer market for books in 2010 (2013), such price differences should have limited impact on cross-channel competition in the German book market (Börsenverein des Deutschen Buchhandels, 2018).

In terms of product differentiation, it is important to note that books are standardized goods that typically come in a single form. Beyond the design of different book types, most publishers do not face quality competition from other publishers as they are the only ones distributing a particular book. Similarly, retailers have little opportunity to alter the quality of a book (e.g.,

[2] Until a reform of the BPBG in 2016, cross-border book sales were also exempt from the law unless books were exported solely for the purpose of re-import. As cross-border sales of foreign-language books are a negligible share of the overall book market, we do not further discuss this exception.

by bundling the book with other products or services). As a result, the quality of a particular book does not differ across different distribution channels and is thus easily verifiable. These properties may also contribute to the fact that books were among the first goods to be frequently bought online (Figure 3.2).

The economic literature also discusses whether online and offline product portfolios differ and how this affects customer demand. For example, Brynjolfsson et al. (2009) show that online retail is virtually immune from brick-and-mortar competition when selling niche products. This finding is the so-called long tail phenomenon: online retailers can sell a wider range of products than brick-and-mortar retailers due to lower distribution and inventory costs (Brynjolfsson et al., 2009). As Zentner et al. (2013) show for the case of video rental patterns, a greater product portfolio may also affect customers' choices in that people renting videos online rather than in brick-and-mortar stores are more likely to watch niche films.

In our context, however, there should be no meaningful differences in the book portfolios offered online as compared to offline. Even if a bookstore does not showcase a specific book, customers can typically order any available book on the market at a bookstore at no or very low cost. If the local bookstore can achieve economies of scale in ordering books or has better access to preferable delivery services, books ordered by local stores may even arrive earlier at the customer than ones privately bought online. Another reason why the product supply between online and offline channels could differ involves self-published books. As self-published books are mainly sold online, local stores might not be able to order them. Yet, self-published books accounted for less than 2% of the general book market in 2014 (Behrens, 2017). Consequently, there is no substantial choice difference between the online and offline sales channels.

In summary, our empirical analysis of the relationship between online shopping and brick-and-mortar bookstores in Germany should neither be driven by differences between online and offline prices nor by product differentiation. Instead, we are primarily identifying the joint effects of having a new, convenient *distribution channel* available as well as the complementary services associated with this channel (e.g., better information based on customer reviews). As these joint effects are increasingly important for the retail of other products as well, our results provide important insights into understanding how these retail industries may develop in the future.

3.3 Data

3.3.1 Geomarketing Information

Our empirical analysis is based on a data set that combines novel information on regional shopping behavior with administrative records. To measure the regional uptake of online shopping, we obtain information from Nexiga GmbH (Nexiga). As a geomarketing company, Nexiga gathers micro-geographical data for more than 22 million addresses in Germany and offers marketing services to businesses that want to target specific groups. Nexiga's extensive database contains information on more than 500 characteristics, ranging from the structure of a building and the quality of the surrounding infrastructure to detailed sociodemographic attributes of the residents.

Importantly for us, Nexiga calculates a building-level index of the *affinity to shop online* (*OSA*). Its OSA index is based on an extensive consumer survey conducted by Schober Information Group Deutschland GmbH (Schober) between 1997 and 2012. With more than 5 million respondents, this survey was the largest written consumer survey over that time period and covered more than 10% of all German households. Since responses to this survey can be geocoded via the postal addresses of the participants, Nexiga is able to match the responses with additional address- and building-level data. Using these additional data, Nexiga applies multivariate regression techniques to extrapolate survey responses across buildings and time. Eventually, it assigns each building b an ordinal value between 1 (low) and 9 (high) that denotes the residents' affinity to shop online (OSA_b).

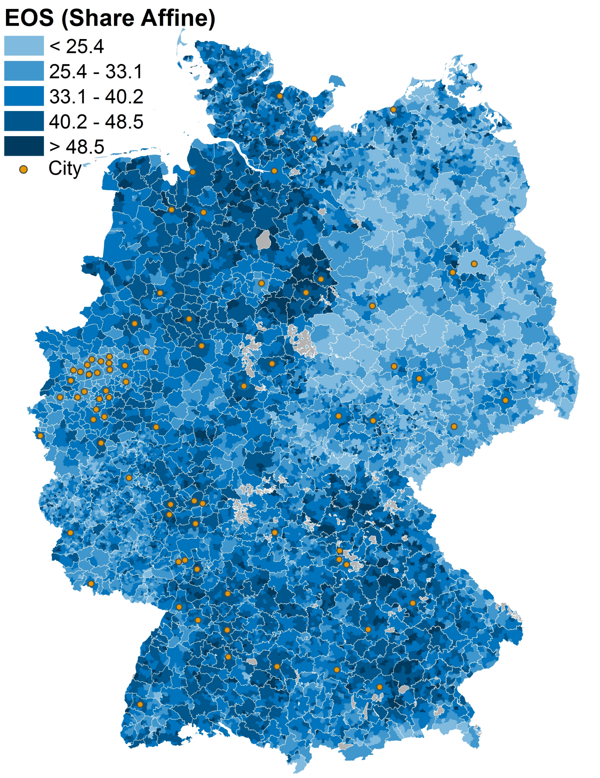
We obtain an aggregation of OSA_b that measures the *regional exposure to online shopping* at the municipality level (EOS_m) in 2015. To that end, Nexiga weights OSA_b by the number of households in building b (HH_b) and calculated the share of households per municipality that are associated with an OSA_b of 7 or higher. Formally:

$$EOS_m = \frac{1}{\sum_{b \in m} HH_b} \sum_{b \in m} (OSA_b \cdot HH_b) \mathbf{1}\{OSA_b \geq 7\},$$

where $\mathbf{1}\{OSA_b \geq 7\}$ is an indicator function assuming the value of 1 if OSA_b is greater than or equal to 7 and 0 otherwise. That is, EOS_m measures the share of households per municipality that have a high affinity for online shopping.

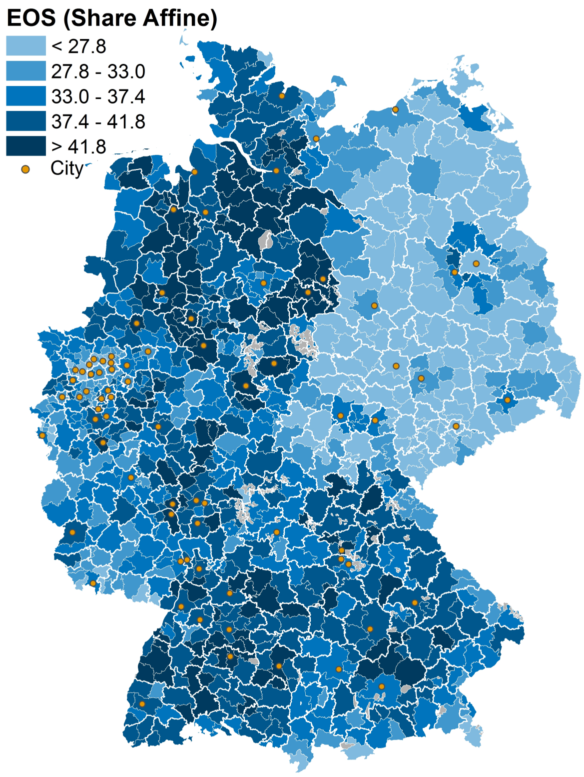
Figure 3.3 illustrates the regional distribution of EOS_m. Considering online shopping preferences at this fine spatial scale reveals the striking insight that high values of EOS_m do not have to coincide with the location of large agglomerations ("cities" are marked with yellow dots). For example, most of the municipalities in the Ruhr region, the largest agglomeration area in Germany, falls in the second lowest quintile of EOS_m whereas many of the highest values can be found around or in-between larger cities. In the area of the New German Länder, however, this relationship is less pronounced. This may be at least partly explained by the general difference in the average levels of EOS_m between East and West Germany. Taking these observations together, the figure suggests on one hand, that the regional exposure to online shopping is positively correlated with purchasing power at higher levels of aggregation. On the other hand, the lower uptake of online shopping in densely populated municipalities might be explained by the fact that these areas typically provide better access to traditional shopping infrastructure.

Figure 3.3 : Regional Distribution of EOS_m



Note: The map illustrates the regional distribution of EOS (share of highly online shopping affine population) at the municipality level. White lines define regional retail markets. Orange dots mark municipalities with at least 100,000 residents. Gray shaded areas are unincorporated territories ("Gemeindefreie Gebiete") or water areas. Own illustrations. Data: Nexiga GmbH (2017), Regionalstatistik. Map: GeoBasis-DE / BKG 2014.

Figure 3.4 : Regional Distribution of EOS_r



Note: The map illustrates the regional distribution of EOS (share of highly online shopping affine population) at the retail-market level. Thin (thick) white lines define regional retail markets (regional labor markets). Orange dots mark municipalities with at least 100,000 residents. Gray shaded areas are unincorporated territories ("Gemeindefreie Gebiete") or water areas. Own illustrations. Data: Nexiga GmbH (2017), Regionalstatistik. Map: GeoBasis-DE / BKG 2014.

In 1999, roughly 83.2% of all municipalities in Germany did not have a brick-and-mortar bookstore, implying that we have to assign these municipalities to those local retail markets in which their residents were most likely to buy books before they were able to shop online. In the absence of comprehensive data on individual shopping behavior, we address this challenge by relying on an administrative definition of the Federal Institute for Building, Urban Affairs and Spatial Development (BBSR). The BBSR defines "Mittelbereiche" to describe areas in which residents are able to meet their demand for high(er) quality goods and services.[3] That is, these areas are explicitly designed to account for retail-specific linkages between regions. In our book-specific context, using this definition for local retail markets (r) passes an important robustness check in that only 31 of the 878 retail markets in Germany did not have at least one traditional bookstore in 1999.

Since municipalities are mostly nested within retail markets, we can calculate our measure for regional exposure to online shopping at the retail-market level (EOS_r) as a population-

[3] "Mittelbereiche" are primarily used for the purpose of spatial observation and not in regional planning.

weighted average of EOS_m. Figure 3.4 illustrates EOS_r across all 878 retail markets in Germany. On average, 35% of the households in a retail market are highly affine to online shopping. Even though the average retail market nests 13 municipalities, the patterns observed at the municipality level are also visible after aggregation. For example, the difference between East and West Germany even becomes more pronounced when considering retail markets.

It is important to note that regional variation in exposure to online shopping is largely driven by differences in local population characteristics rather than by technological factors. For example, as online book retailers offered their services throughout the whole country soon after market entry, their services were not only available to a limited set of selected municipalities. Furthermore, local availability of internet infrastructure did not restrict the opportunity to shop online as integrated services digital networks (ISDN) have been available to all German households since the end of the 1990s.[4] Even though access to higher internet bandwidths may has fostered shopping online, we argue that this is a negligible factor for the uptake of online shopping as compared to personal characteristics such as age, education, and income. As a result, EOS_r can best be thought of as a regional aggregate of the individual decisions made by people living in retail market r.

3.3.2 Administrative Data

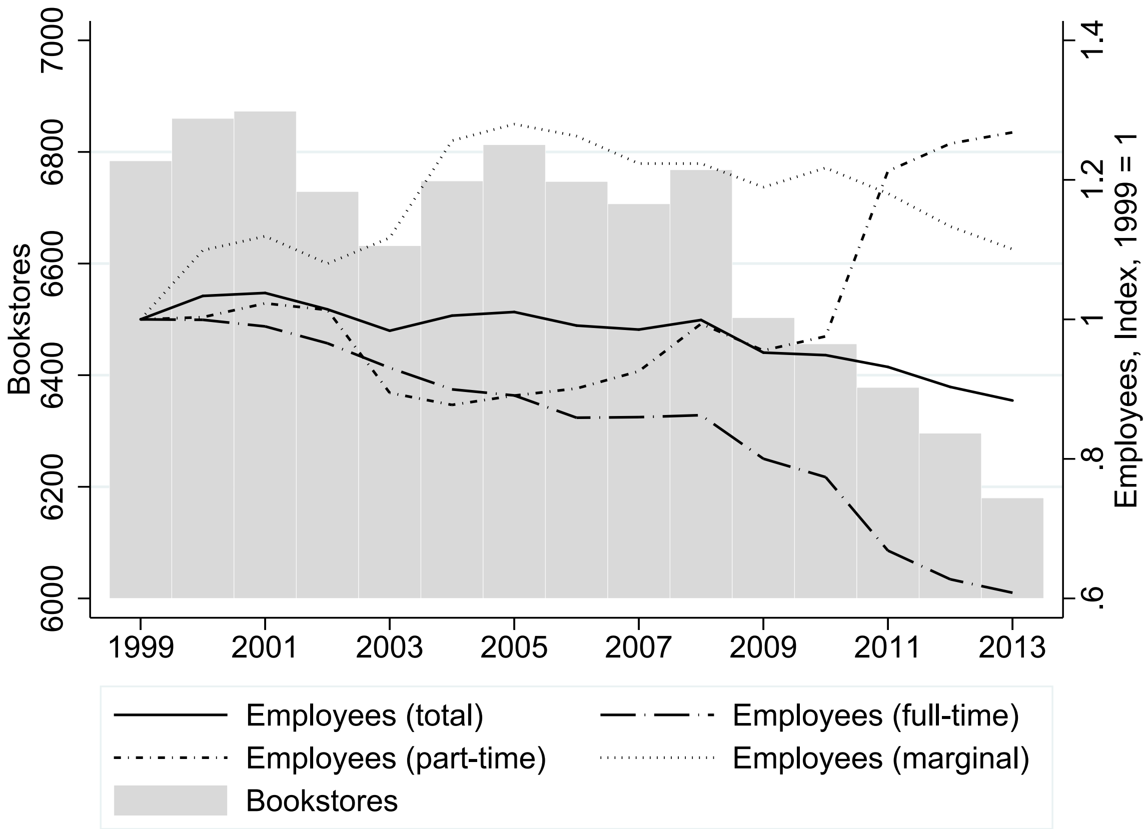
Our main source of administrative data is the Establishment History Panel (BHP) of the Institute for Employment Research (IAB). The BHP provides detailed information on the universe of all German establishments that have at least one employee covered by social security contributions on June 30th in each year. Although our version of the BHP spans the years from 1975 to 2014 (East German establishments are included since 1992), the detailed industry classifications necessary to identify brick-and-mortar bookstores are available only from 1999 onward.[5] Prior to this year, the BHP did not include any information on establishments solely employing so-called marginal workers, i.e., part-time employees not subject to social security and with very low working hours and wages (currently up to 450 euro per month). Since these workers may be of particular interest in the retail sector, our empirical analysis uses administrative data only from 1999 onward. One of the largest book publishing houses in Germany ("Verlagsgruppe Weltbild") went bankrupt in 2014. As this publishing house also operated numerous traditional brick-and-mortar stores, we exclude this year from our analysis to avoid identifying short-term fluctuations. Our final sample covers the years from 1999 to 2013.

With regard to brick-and-mortar bookstores, we clean the BHP by excluding all establishments that had more than 100 employees in 2013 (i.e., that are above the 99.5% percentile). This is necessary as some large online retailers for books are misclassified. To match our cleaned

[4] Large online retailers optimize their websites to make efficient use of available bandwidths.

[5] We define a consistent measure of stationary bookstores using five-digit industry classifications from the German classification systems of 1993 and 2008, i.e., the "Wirtschaftszweigsklassifikationen".

Figure 3.5 : Brick-and-Mortar Bookstores and their Employees



Note: The left axis depicts the number of German brick-and-mortar bookstores over time. The right axis indicates the development of book-sector employment over time; it is indexed to 1999. Data: BHP 7514 (for documentation, see Schmucker et al. (2016)).

establishment-level information with the regional geomarketing data, we aggregate the BHP to the retail-market-by-year level. For each retail market, this aggregation provides us with the number of brick-and-mortar bookstores, the number of employees (also distinguished by full-time, part-time, and marginal), and the average wage of full-time employees. By considering all industries at once (i.e., rather than focusing on traditional bookstores only), the BHP also allows us to define a comprehensive set of control variables that measure employment-related characteristics of the retail markets. To this end, we count the number of employees by qualification, nationality, and gender, as well as measure the average wage across all industries.

Figure 3.5 shows the number of brick-and-mortar bookstores in Germany in levels as well as the development of employment therein indexed to the year 1999. Between 1999 and 2013, the number of bookstores decreased by roughly 10%, which matches the decline in the share of sales observed in Figure 3.1. In terms of retail employment, there is strong divergence across types of workers. While the number of full-time employees has decreased by 40% since 1999, the number of regular and marginal part-time workers increased by roughly 25% and 10%, respectively. Overall, the number of employees working in traditional bookstores decreased

Table 3.1 : Descriptive Statistics

	1999		2013		$\Delta_{2013-1999}$	
	Mean	SD	Mean	SD	Mean	SD
Panel A: Dependent Variables						
Bookstores	0.77	0.40	0.71	0.36	−0.06	0.39
Employees (total)	3.48	2.35	3.19	2.12	−0.29	1.87
Employees (full-time)	1.44	1.11	0.86	0.80	−0.57	0.90
Employees (part-time)	0.60	0.66	0.81	0.76	0.22	0.73
Employees (marginal)	1.25	1.10	1.38	1.08	0.13	1.08
Panel B: Explanatory Variables						
EOS	0.00	0.00	0.35	0.06	0.35	0.06
EOS^{bartik}	0.00	0.00	0.33	0.01	0.33	0.01
Share Population <20	0.22	0.02	0.18	0.02	−0.04	0.01
Share Population 20–29	0.11	0.01	0.11	0.02	−0.00	0.01
Share Population 30–39	0.16	0.01	0.11	0.01	−0.05	0.01
Share Population 40–49	0.15	0.01	0.15	0.01	0.00	0.01
Share Population 50–59	0.12	0.01	0.16	0.01	0.05	0.02
Share Population ≥60	0.24	0.02	0.28	0.03	0.04	0.02
Panel C: Control Variables						
Population Density	0.15	0.15	0.15	0.15	0.00	0.00
Internet Usage	0.00	0.00	0.29	0.09	0.29	0.09
Share Female	0.48	0.04	0.49	0.04	0.02	0.02
Share German	0.94	0.04	0.93	0.04	−0.01	0.01
Share Medium Skilled	0.73	0.04	0.74	0.06	0.00	0.03
Share High Skilled	0.07	0.03	0.12	0.04	0.04	0.02
Share Employed	0.34	0.07	0.40	0.09	0.06	0.04
Average Hourly Wages	58.64	19.67	73.01	27.95	14.36	17.66
N	792		792		792	

Note: Dependent variables are measured per 10,000 residents. EOS denotes online shopping exposure and measures the share of residents within a retail market who have a high preference for shopping online. EOS^{bartik} is a Bartik-type instrument for EOS (for more details on its computation, see Section 3.4.2). Population density is measured in terms of 100 residents per km^2. Average hourly wages are calculated for all full-time workers within a retail market. Data: Nexiga GmbH (2017), Regionalstatistik, BHP (for a documentation of the BHP7514, see Schmucker et al. (2016)).

by around 15%. At least in the aggregate, these developments imply that the dramatic change in full-time employees has been partly compensated for by increased employment of part-time employees.

We complement our aggregation of the BHP and the data from Nexiga with a series of regional variables that are available from public sources. These include data on population (by age and gender), the geographic area of municipalities (as well as all aggregations thereof), and information on regional infrastructure. Table 3.1 presents descriptive statistics for all variables of interest. In 1999, the average retail market had 0.77 traditional bookstores per 10,000 residents. By 2013, this number had decreased by 0.06 or 7.8%, which is close to the change in aggregate numbers illustrated in Figure 3.5. All dependent variables are measured per 10,000 residents, i.e., denote densities, to account for size differences between retail markets. With regards to employment, it is worth noting that the national developments in the composition of employees (shown in Figure 3.5) are also reflected in the average numbers rather than being limited to a few selected places like large cities.

3.4 Empirical Framework

3.4.1 Baseline Regression

To determine whether regional exposure to online shopping affects the development of brick-and-mortar bookstores, we estimate the following long difference model:

$$\Delta Y_r = \beta \Delta EOS_r + \Delta \mathbf{X}_r' \theta + \varepsilon_r, \tag{3.4.1}$$

where ΔY_r denotes a change in outcome Y of retail market r between 1999 and 2013. Our main outcomes of interest are the local stock of brick-and-mortar bookstores and the number of employees working in them (both measured per 10,000 residents). Our explanatory variable of interest, ΔEOS_r, measures the change in regional exposure to online shopping over the same time period. In 1999, we assume that EOS_r equals zero as only a negligible share of books was purchased online at that time. Consequently, any variation observed in ΔEOS_r is equal to the regional variation in 2013. Since this variable describes long-term changes in shopping behavior, we use Nexiga's data for 2015 as a proxy for 2013. $\Delta \mathbf{X}_r$ is a vector of changes in control variables which includes information on internet usage, population density, and various characteristics of the local labor market.

In Equation (3.4.1), we address many endogeneity concerns related to the assumption that ΔEOS_r is primarily driven by population-specific aspects with fixed effects as well as a comprehensive set of control variables. With regard to fixed effects, our long difference approach implicitly accounts for any type of time-constant differences across retail markets. These include, among others, the effects of persistent differences in age, education, and income levels. However, as we continue to be concerned that unobserved shocks to the local population structure (e.g., heterogeneous trends of agglomeration) could compromise our results, we also include fixed effects for regional labor markets (RLMs). We use the IAB's definition of RLMs, which describes 258 geographic areas in Germany that are economically integrated and characterized by a high share of work-related commuting. Empirically, using RLM fixed effects implements a *geographic* conditional independence assumption (CIA). That is, by only comparing retail markets within the same RLM, we account for unobserved shocks that vary smoothly across space within confined geographical regions (i.e, RLMs in this case). Among others, this allows us to control for changes in average local income levels due to the closing of a large plant or the decline of an industry clustered in a specific region.

Even though these fixed effects eliminate unobserved heterogeneity in various dimensions, they may still fail to adequately capture other changes in the local population structure. Our greatest concern is that using the internet in general may have direct effects on the development of brick-and-mortar bookstores. For example, internet usage may be positively correlated with the number of books purchased at traditional stores as the information available online helps customers identify books they want to buy. Furthermore, there could also be a negative relationship if spending time online causes people to read less or replace traditional media like encyclopedias, dictionaries, and tourist guides with similar online services. As internet usage is likely determined by the same personal characteristics as online shopping, our estimates in Equation (3.4.1) might be biased if we neglect these effects. Therefore, we include the change in internet usage across retail markets between 1999 and 2013 as a control variable in our regression.

To assess the sensitivity of our results, we also control for changes in other population characteristics between 1999 and 2013. At the retail-market level, we account for changes in agglomeration by including population density. Furthermore, we control for differences in population shares by employment status, skill level, gender, and nationality as they might be correlated with the adoption of new technologies. With regard to income levels, we include the change in the average hourly wage of full-time workers.

3.4.2 Identification

Even though our combination of fixed effects and time-varying control variables alleviates some endogeneity concerns related to ΔEOS_r, others remain. For example, our measure of regional exposure to online shopping might be subject to measurement error. On the one hand, measurement error could be introduced by Nexiga's extrapolation of the Schuber survey data across space and time. On the other hand, our population-weighted aggregation of municipality-level information could fall prey to similar problems. Furthermore, reverse causality could affect our OLS estimates if the availability or quality of traditional shopping infrastructure (e.g., the number of stores) affect the uptake of online shopping. Similarly, unobserved shocks to local retail markets, for example, the opening or closing of a large mall, could cause biases in the same direction.

In the absence of technological frictions or political incentives that could lead to exogenous variation in ΔEOS_r, we employ a Bartik-type identification approach to address these endogeneity concerns. First introduced by Bartik (1991), this type of identification approach has been used in many fields of economics. Most closely related to this paper are Acemoglu and Restrepo (Acemoglu and Restrepo) and Dauth et al. (2017), who use Bartik-type instruments to identify the effects of industrial robots usage on regional employment. We adapt their empirical framework to our retail-specific context and construct a Bartik-type instrument that exploits regional variation in the uptake of online shopping based on different age structures. Specifically, we take advantage of the fact that the share of book expenditures online differs considerably across age groups (see the survey data illustrated in Figure 3.6). Acemoglu and Linn (2004) exploit similar variation to identify the effect of market size on pharmaceutical innovations.

Figure 3.6 shows that almost half the people in the age group from 30 to 39 prefer buying their books online. In contrast, only 19% of the people who are 60 and older used this retail channel. Assuming that all books at the beginning of our sample period were bought in traditional stores, and keeping everything else constant, this figure implies that traditional bookstores have lost, on average, 31.6% of their business since the introduction of online shopping. However, since the share of online expenditures varies strongly across age groups ($OnlShare_a$), we do not expect that all brick-and-mortar bookstores are affected equally by this aggregate development. Instead, traditional bookstores that are located in retail markets with a population more prone to online shopping should also be subject to stronger effects of online shopping.

Formally, we exploit this relationship by constructing a Bartik-type instrument:

$$\Delta EOS_r^{bartik} = \sum_a \underbrace{\frac{Pop_{a,r,2001}}{Pop_{r,2001}}}_{\text{Historic Share}} \left(\underbrace{\frac{Pop_{a,2013} \cdot OnlShare_{a,2013}}{Pop_{a,1999}} - \frac{Pop_{a,1999} \cdot 0}{Pop_{a,1999}}}_{\text{Aggregate Change}} \right), \tag{3.4.2}$$

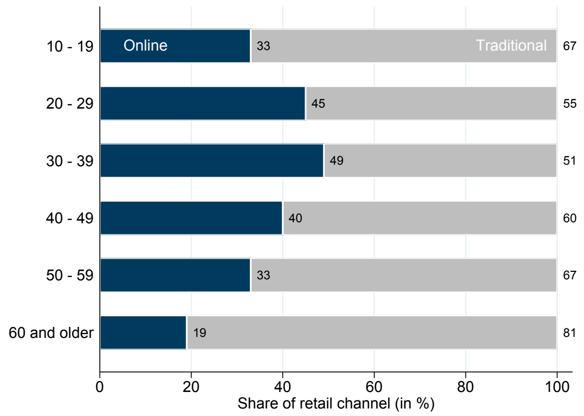
where the first factor in the product denotes the share of people in age group a, living in retail market r, in year 2001, and the second factor indicates the aggregate change in “online shoppers” standardized by the population in 1999. Since online shopping only started to take off in the early 2000s, we assume that the share of online expenditures for books in 1999 was zero. The age shares in 2001 are intended to approximate the regional age structure before online shopping was introduced. We use 2001 as it is the first year for which we have comprehensive information on age shares across all retail markets. The persistence in regional age structures should ensure the validity of this choice.

We use our instrument to estimate the following first stage relationship:

$$\Delta EOS_r = \delta \Delta EOS_r^{bartik} + \Delta \mathbf{X}'_r \theta + \eta_r. \tag{3.4.3}$$

We illustrate this relationship in Figure 3.7, which is an added variables plot of our most restrictive specification. It includes RLM fixed effects as well as all the control variables described above. The solid red line is an OLS estimate of the slope coefficient δ in Equation (3.4.3). The coefficient δ is greater than 3, indicating a strong and positive correlation between ΔEOS_r and our Bartik-type instrument ΔEOS_r^{bartik}.

Figure 3.6 : Book Shopping by Channel & Age



Note: Book shopping distinguished by retail channel and age group. The survey was conducted in 2015 and is representative of the German population. The traditional retail channel comprises purchases from brick-and-mortar stores as well as other stationary retailers. Textbooks are excluded. Data: Börsenverein des Deutschen Buchhandels (2015), $n = 60{,}070$.

Figure 3.7 : First Stage of ΔEOS_r on ΔEOS_r^{bartik}

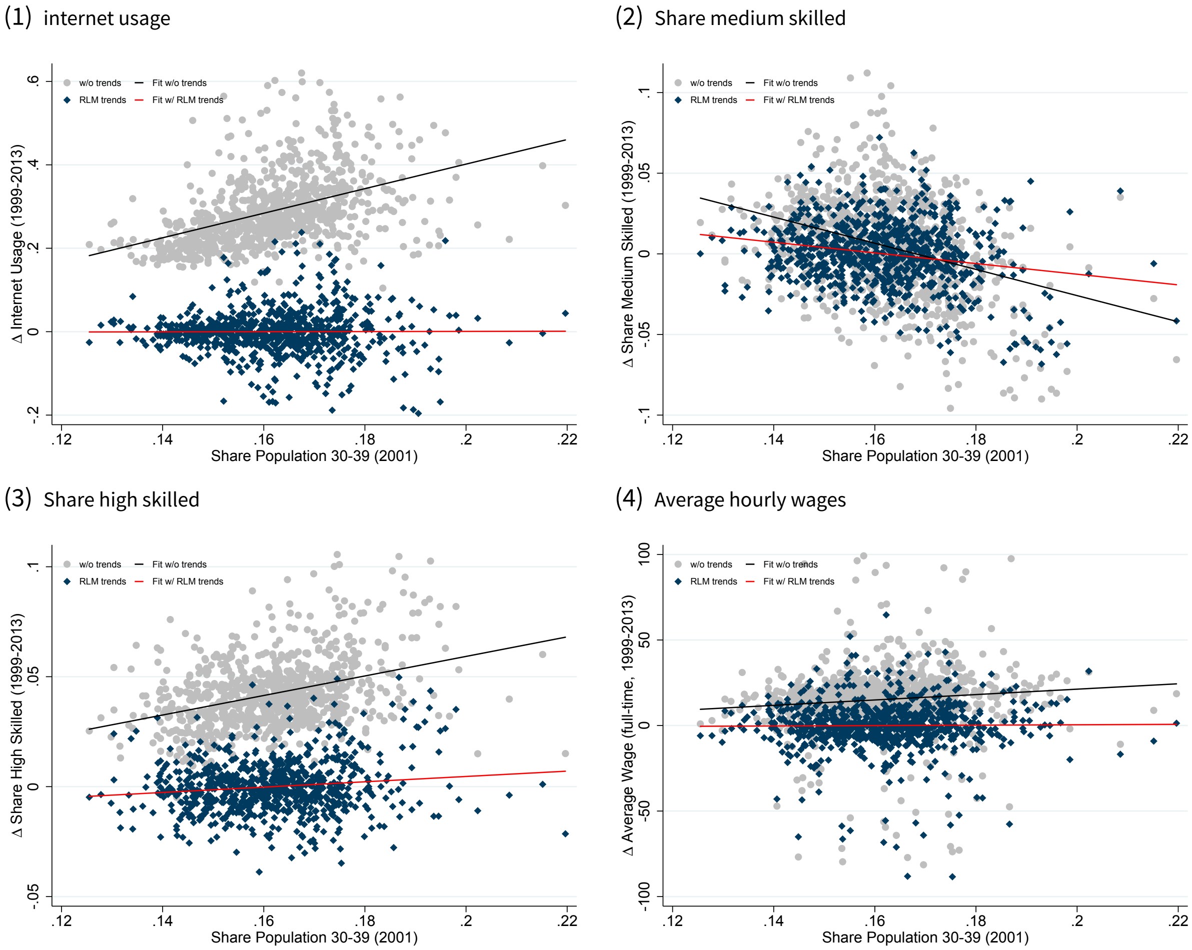
Note: Added variables plot of ΔEOS_r and ΔEOS_r^{bartik}, accounting for RLM-fixed effects, changes in population density, online usage, share female, share foreign, share medium skilled, share high skilled, share employed, and average hourly wages. Data: Nexiga GmbH (2017), Regionalstatistik, BHP (for a documentation of the BHP7514, see Schmucker et al. (2016)).

To explain why this coefficient is greater than 1, recall that both measures for online shopping differ within two dimensions. On the one hand, ΔEOS_r reflects a general change in the local population's *affinity* to shop online that is not specific to certain products or services. On the other hand, ΔEOS_r^{bartik} measures *actual online expenditures* for books as a fraction of overall spending. Therefore, it may not be surprising that a high share of actual expenditures requires an even greater share of online shopping affine people as even highly affine people occasionally shop for books offline. Furthermore, general online shopping affinity may be greater than that related to a specific product (in our case books).

There are several reasons why using ΔEOS_r^{bartik} as an instrument addresses most of the endogeneity concerns mentioned above. For instance, many studies have to rely on Bartik-type instruments as proxies for their regional variables of interest simply because there are no adequate data available (e.g., Acemoglu and Restrepo (Acemoglu and Restrepo)). At the very least, ΔEOS_r^{bartik} provides a second measure for ΔEOS_r that should be subject to a different type of measurement error than our explanatory variable of interest. Consequently, using it as an instrument should reduce measurement error in our survey-based information. With regard to omitted variables, the central advantage of a Bartik-type instrument is that it reduces the number of endogeneity concerns to a single dimension. That is, rather than arguing why ΔEOS_r may not be correlated with unobserved characteristics in Equation (3.4.1), we can focus on the question of why the age-related aspect in this measure is exogenous.

Before we discuss the validity of our instrument in detail, recall that, in general, Bartik-type instruments are an interaction of two components: a (historic) local share and an aggregate shock. Most of the empirical literature using this identification approach argues that the aggregate shock (and, therefore, the instrument) is exogenous in its specific context. In some

Figure 3.8 : Relationship Between (Pre-Internet) Age Shares and Changes in Control Variables



Note: This figure plots the relationship between the share of 30 to 39 year old people in 2001 and the change in a selection of control variables between 1999 and 2013. Gray dots (black lines) indicate scatter plots (OLS regression coefficients) of the unconditional relationship between the change in the control variable of interest and the pre-internet share of people between 30 and 39. Blue diamonds (red lines) represent added variable scatter plots (OLS regression coefficients) of the same relationships after accounting for RLM-fixed effects. Data: Nexiga GmbH (2017), Regionalstatistik, BHP (for a documentation of the BHP7514, see Schmucker et al. (2016)).

cases, this also involves instrumentation of the shock (see, e.g., studies using the "China shock" based on the work by Autor et al. (2013)). In recent work, Goldsmith-Pinkham et al. (2018) challenge this line of reasoning and argue that identification using Bartik-type instruments should be discussed in terms of the shares rather than the shocks. To formalize their argument, they show that using a Bartik-type instrument is numerically equivalent to using the local shares directly and that aggregate shocks only contribute to instrument relevance.

We thus discuss the validity of our Bartik-type instrument from both angles. Regarding the exogeneity of the aggregate shock in our context, it can be argued that online shopping only became possible because the internet was invented. Since this was an international-level technological advance, it should not be correlated with regional developments in Germany. In a similar vein, using an aggregate shock also allows us to abstract from selection problems caused by regional migration as the aggregate population in Germany is only affected by demographic factors and international migration.

With regards to the exogeneity of our shares, it is less straightforward to argue why the regional age structure in 2001 is neither directly related to developments in regional book markets nor affected by any omitted factors. For example, our fixed effects structure could fail to adequately account for regional trends that heterogeneously affect local retail markets. To address this problem, we follow a suggestion by Goldsmith-Pinkham et al. (2018) and conduct a simple empirical test. Specifically, we investigate whether our regional age shares in 2001 are balanced across the set of control variables available to us. This test is inspired by the idea that, ideally, a valid instrument should not be affected too strongly by observable characteristics as this suggests that unobservables may also be important (see, e.g., Oster (2018) for an elaboration on this argument). Even though one may argue that controlling for observable confounders solves any problems that can be detected with this test (in fact, we also build on a conditional independence assumption above), it still provides important insights in our context.

As an example, Figure 3.8 shows a selection of four scatter plots that indicate how balanced the share of 30 to 39 year old people is across four important control variables (i.e., the change in internet usage, average hourly wages of full-time workers, and the share of medium as well as high skilled employees).[6] The gray dots indicate the unconditional relationship between the change in the control variable of interest and the (pre-internet) share of the 30 to 39 year old residents (black lines denote corresponding coefficients of OLS regressions).

As can be seen in the upper left graph, there is a strong (unconditional) positive correlation between the pre-internet share of 30 to 39 year old people and the change in internet usage. However, if we compare retail markets only within RLMs, this relationship vanishes entirely. Qualitatively similar observations can also be made in the other graphs in that adding RLM-fixed effects at least reduces the relationship observed in the unconditional case considerably. Taking these observations together with the fact that we can draw similar conclusions for all other age shares, we conclude that our Bartik-type instrument is valid only when we include RLM-fixed effects to account for underlying regional trends. Conditional on doing so, however, our approach can address the endogeneity concerns discussed above.

[6] In the Appendix, we show the corresponding graphs for the remaining control variables used in the regressions with respect to the share of 30 to 39 year old people. Conducting the same visual inspections for all other age shares leads to similar conclusions (results are available from the authors upon request).

3.5 Results

3.5.1 Baseline OLS and IV Results

Table 3.2 shows the results of our baseline OLS and IV estimations for the number of brick-and-mortar bookstores (measured per 10,000 residents). Across all specifications, we include RLM fixed effects and cluster standard errors at the same level. We exclude retail markets below the bottom and above the top percentile as we do not want our effects to be driven by outliers and also drop retail markets that do not have at least one bookstore in 1999. Since we are estimating long difference models, the number of observations corresponds to the number of retail markets considered (i.e., $N = 792$).

Our OLS estimates show that regional online shopping exposure has robust negative, though statistically insignificant, effects on the number of brick-and-mortar bookstores. As expected, including the change in internet usage (i.e., in Column (2)) decreases the estimated effect. This supports the idea that online shopping and internet usage are positively correlated and should both affect traditional retailers. Neglecting internet usage therefore leads to an upward bias of the coefficient for ΔEOS in Column (1). In our most comprehensive OLS specification (Column (3)), we also include a set of socio-demographic as well as labor market controls. Among these, only the change in average hourly wages of full-time workers has a statistically significant effect; suggesting that changes in income levels affect the development of local bookstores.

With regard to the IV results in Columns (4) to (6), all regressions are associated with strong first stage results. Across specifications, a 1 percentage point increase in ΔEOS^{bartik} leads to an increase in online shopping exposure by more than 3 percentage points. The corresponding Kleibergen-Paap F-statistics are either close to or above 100, rejecting the null hypothesis that we face a weak instrument problem. Regarding the second stage, we find that our IV estimates for ΔEOS are considerably smaller than their OLS counterparts and become statistically significant at the 10% level (Column (6)).

There are (at least) two reasons that may help to explain the differences between our OLS and IV estimates. On the one hand, an upward bias in our OLS regressions could result from unobserved shocks that are positively correlated with ΔBookstores and ΔEOS, respectively. For example, if our controls fail to adequately account for changes in local income levels (which are likely positively correlated with EOS, see the maps in Section 3.3.1), this would explain the pattern observed. On the other hand, Nexiga constructs our measure of EOS based on interpolations across space and time. If these interpolations induce classical measurement error, this would lead to an attenuation bias that could also explain our smaller IV estimates.

Considering Column (6) as our preferred specification, the point estimate of –1.771 implies that an increase in regional online shopping exposure by one standard deviation (i.e., 6 percentage points) reduces the local stock of brick-and-mortar bookstores per 10,000 residents by 0.11.

Table 3.2 : Baseline OLS and IV Results (Bookstores)

ΔBookstores	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
ΔEOS	−0.614	−0.773	−0.908	−1.453	−1.702	−1.771*
	(0.563)	(0.581)	(0.574)	(0.922)	(1.084)	(1.018)
ΔInternet Usage		0.257	0.308		0.347	0.389*
		(0.205)	(0.212)		(0.219)	(0.229)
ΔPopulation Density		−9.213	−21.59		66.93	42.03
		(127.8)	(106.3)		(153.7)	(130.8)
ΔShare Female			−0.528			−0.553
			(0.781)			(0.784)
ΔShare German			1.939			1.879
			(1.435)			(1.433)
ΔShare Medium Skilled			−0.286			−0.267
			(0.930)			(0.933)
ΔShare High Skilled			−1.504			−1.347
			(1.502)			(1.535)
ΔShare Employed			0.238			0.310
			(0.527)			(0.503)
ΔAverage Hourly Wages			0.008***			0.008***
			(0.001)			(0.001)
First Stage						
ΔEOS^{bartik}				3.619***	3.470***	3.517***
				(0.323)	(0.352)	(0.354)
Kleibergen-Paap F-Stat.				125.40	97.19	98.67
RLM-FE	Yes	Yes	Yes	Yes	Yes	Yes
N	792	792	792	792	792	792

Note: Dependent variable is the change in the number of bookstores per 10,000 residents between 1999 and 2013. All specifications include RLM-fixed effects. Standard errors are clustered at the RLM level. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Data: Nexiga GmbH (2017), Regionalstatistik, BHP (for a documentation of the BHP7514, see Schmucker et al. (2016)).

Given that the average retail market in 1999 had 0.77 bookstores per 10,000 residents, this effect corresponds to a reduction by approximately 14%.

In Table 3.3, we present the IV results for changes in employment (i.e., the total number of employees as well as employees distinguished by type). As for bookstores, the employment variables are measured per 10,000 residents. Our IV estimates for these outcomes fit into the picture observed for bookstores. For example, Column (1) shows that an increase in the regional exposure to online shopping substantially reduces overall employment in brick-and-

Table 3.3 : Baseline IV Results (Employees)

ΔEmployees	Total	Full-Time	Part-Time	Marginal
	(1)	(2)	(3)	(4)
ΔEOS	−7.297*	−2.896	−1.953	−2.318
	(4.261)	(2.232)	(1.644)	(2.912)
ΔInternet Usage	0.532	0.436	−0.331	0.604
	(1.594)	(0.851)	(0.573)	(0.776)
ΔPopulation Density	452.9	203.7	29.44	205.3
	(663.1)	(359.8)	(260.6)	(340.5)
ΔShare Female	2.148	1.578	0.251	0.366
	(3.490)	(1.839)	(1.899)	(2.075)
ΔShare German	−1.784	−0.174	−5.830*	3.340
	(7.282)	(4.627)	(3.235)	(4.266)
ΔShare Medium Skilled	6.133	1.273	4.188	−0.710
	(5.186)	(2.261)	(2.753)	(2.827)
ΔShare High Skilled	−1.596	−9.298***	6.268*	1.849
	(7.253)	(3.198)	(3.734)	(4.395)
ΔShare Employed	1.727	−0.006	0.871	1.007
	(2.126)	(1.040)	(0.970)	(1.154)
ΔAverage Hourly Wages	0.020***	0.004***	0.005***	0.001***
	(0.004)	(0.002)	(0.002)	(0.002)
First Stage				
ΔEOS^{bartik}	3.517***	3.517***	3.517***	3.517***
	(0.354)	(0.354)	(0.354)	(0.354)
Kleibergen-Paap F-Stat.	98.67	98.67	98.67	98.67
RLM-FE	Yes	Yes	Yes	Yes
N	792	792	792	792

Note: Dependent variables are measured per 10,000 residents and indicate changes in the number of bookstores and employees between 1999 and 2013, respectively. All specifications include RLM-fixed effects. Standard errors are clustered at the RLM level. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Data: Nexiga GmbH (2017), Regionalstatistik, BHP (for a documentation of the BHP7514, see Schmucker et al. (2016)).

mortar bookstores. This effect is statistically significant at the 10% level. The point estimate of −7.297 implies that an increase in regional online shopping exposure by one standard deviation leads to a reduction in total brick-and-mortar employment by 0.44 employees per 10,000 residents. In terms of 1999 employment levels, this corresponds to a reduction by approximately 13%. Regarding our results in Columns (2) to (4), the degree of precision associated with our IV estimates prohibits us from drawing strong conclusions about whether ΔEOS has heterogeneous effects across employment types. However, considering the point estimates, our results suggest that full-time employees may be affected most severely by the introduction of online shopping and contribute the largest part to the overall effect.

3.5.2 Robustness and Heterogeneity

In Section 3.4.2, we have argued that our estimation strategy is only valid conditional on a set of fixed effects. While we have also motivated the importance of our time-varying controls, our estimates have been relatively robust toward their inclusion. So far, one important limitation of our time-varying controls is that they only account for general characteristics of the local population and labor market, but lack a book market specific component. Incorporating this concern, we test the robustness of our results by including the number of bookstores (per 10,000 residents) in 1999 as a control. Arguably, this should provide a good approximation of book market specific aspects of the local retail market before the internet (and, hence, online shopping) was introduced. Table 3.4 presents the results for all outcomes of interest when controlling for this lag. Except for part-time employees, we find that the density of bookstores in 1999 has a negative and statistically highly significant influence on the development of local book markets. This result suggests that retail markets with a previously high number of brick-and-mortar bookstores are associated with more severe decreases between 1999 and 2013.

With regard to our estimates for ΔEOS, we find that, after including this lag, our estimates become slightly smaller and are more precisely estimated. For example, our point estimate for the change in bookstores decreases from –1.771 (in our preferred IV specification of Table 3.2) to –2.139. While these two point estimates are not statistically significant different from each other, the latter is associated with substantially lower standard errors. Similarly, our estimates for changes in the total number of employees as well as the number of full-time workers (per 10,000 residents) do not change much in magnitude, but are now statistically significant at the 5% and 10% level, respectively.

An additional benefit of including the lagged density of bookstores is that it enables us to investigate the homogeneity of our effects for ΔEOS. More specifically, it allows asking whether regional exposure to online shopping has heterogeneous effects along the distribution of the number of local bookstores in 1999. Table B.1 in the Appendix provides summary statistics for a selection of variables within six intervals to illustrate how retail markets differ along this distribution. Apart from the average bookstore density in 1999, we also include the pre-internet number of bookstore employees (by type) per store to characterize local book markets. In the second panel, we list most of our control variables (i.e., those that are not zero in 1999) and the total numbers of residents and employees. With regard to book market specific aspects in 1999, this table reveals that retail markets with higher bookstore densities in 1999 have, on average, “smaller” bookstores (i.e., lower employees per bookstore ratios). This relationship does not only hold for the total number of employees, but across all types. Maybe surprisingly, the retail markets with the highest bookstore densities are not the largest

Table 3.4 : Robustness Test of IV Results

	ΔBookstores	ΔEmployees			
		Total	Full-Time	Part-Time	Marginal
	(1)	(2)	(3)	(4)	(5)
ΔEOS	−2.139***	−8.266**	−3.388*	−2.009	−2.689
	(0.764)	(4.208)	(2.224)	(1.649)	(2.861)
ΔInternet Usage	0.353*	0.439	0.789	−0.337	0.568
	(0.195)	(1.483)	(0.851)	(0.573)	(0.736)
ΔPopulation Density	46.90	465.7	210.2	30.18	210.2
	(119.6)	(668.9)	(349.1)	(261.8)	(348.0)
ΔShare Female	−0.295	2.827	1.922	0.290	0.626
	(0.630)	(3.496)	(1.945)	(1.895)	(2.047)
ΔShare German	−0.069	−6.903	−2.771	−6.125*	1.380
	(1.184)	(7.529)	(4.747)	(3.206)	(4.489)
ΔShare Medium Skilled	−1.215	3.639	0.008	4.044	−1.665
	(0.759)	(4.899)	(2.178)	(2.747)	(2.728)
ΔShare High Skilled	−0.925	−0.487	−8.736***	6.332*	2.274
	(1.193)	(6.787)	(2.814)	(3.767)	(4.196)
ΔShare Employed	0.039	1.016	−0.354	0.830	0.735
	(0.485)	(2.149)	(0.995)	(0.977)	(1.173)
ΔAverage Hourly Wages	0.008***	0.019***	0.004*	0.005***	0.010***
	(0.001)	(0.004)	(0.002)	(0.002)	(0.002)
L.Bookstores	−0.573***	−1.506***	−0.764***	−0.087	−0.577***
	(0.050)	(0.221)	(0.125)	(0.086)	(0.151)
First Stage					
ΔEOS^{bartik}	3.504***	3.504***	3.504***	3.504***	3.504***
	(0.347)	(0.347)	(0.347)	(0.347)	(0.347)
Kleibergen-Paap F-Stat.	102.11	102.11	102.11	102.11	102.11
RLM-FE	Yes	Yes	Yes	Yes	Yes
N	792	792	792	792	792

Note: Dependent variables are measured per 10,000 residents and indicate changes in the number of bookstores and employees between 1999 and 2013, respectively. All specifications include RLM-fixed effects. Standard errors are clustered at the RLM level. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Data: Nexiga GmbH (2017), Regionalstatistik, BHP (for a documentation of the BHP7514, see Schmucker et al. (2016)).

ones, e.g., with respect to the total number of residents or employees. Instead, there is an inverted u-shape relationship between absolute retail market size and bookstore density. In terms of our controls variables, there are no stark differences between retail markets in the intervals considered. The only notable exception is that higher initial bookstore densities are also associated with higher average wages of full-time workers.

To estimate heterogeneous effects with respect to bookstore density in 1999, we standardize the number of bookstores per 10,000 residents in 1999 at five different percentiles ($p \in \{10, 25, 50, 75, 90\}$) and estimate the following model:

$$\Delta Y_r = \beta_1 \Delta EOS_r + \beta_2 (\Delta EOS_r \cdot L.Bookstores_r^p) + \beta_3 L.Bookstores_r^p + \Delta \mathbf{X}_r^{'} \theta + \varepsilon_r, \tag{3.5.1}$$

where $L.Bookstores_r^p$ denotes the number of bookstores per 10,000 residents in 1999, standardized at percentile p. This standardized measure enters the interaction term ($\Delta EOS_r \cdot L.Bookstores_r^p$). Consequently, the coefficients β_1 and β_3, i.e., the "main effects", reflect conditional relationships. That is, β_1 denotes the effect of ΔEOS_r conditional on $L.Bookstores_r^p$ being zero, whereas β_3 is the effect of the lagged (and standardized) measure for pre-internet bookstore density in absence of a change in ΔEOS_r. Consequently, if one is interested in the effect of a change in ΔEOS_r for a specific percentile $q \neq p$, this can be calculated as $\beta_1 + \beta_2 \cdot (|q - p|)$.

In Table 3.5, we present our results for all outcomes (Panels) and percentiles (Columns) of interest. For the sake of brevity, we omit the point estimates for all control variables as well as for the lagged value of bookstore density. In Panel A, we depict how changes in regional online shopping exposure heterogeneously affect the development of the local stock of brick-and-mortar bookstores. Considering Column (1), an increase in ΔEOS by one standard deviation leads to a decrease of approximately 0.11 bookstores per 10,000 residents (i.e., $-1.895 \cdot 0.06$) at the 10th percentile of the lagged density distribution. In contrast, an identical increase at the 90th percentile has an effect of 0.14. Although the difference between these two point estimates has some economic meaning, we lack the statistical power to conclude that it is also statistically significant.

Turning to the total number of employees in Panel B, we find an economically and statistically more distinct pattern. While retail markets at the bottom of the density distribution do not reduce their total number of bookstore employees significantly, there is a strong and steady increase in the absolute size of the coefficients at higher percentiles. At the 75^{th} and 90^{th} percentile, an increase in ΔEOS by one standard deviation leads to a decrease by approximately 0.62 and 0.93 employees per 10,000 residents, respectively. These estimates are statistically significant at the 1% level.

To understand where these heterogeneous effects come from, we further investigate the effects across types of bookstore employees (Panels C to E). With regard to full-time workers, we find a pattern that is qualitatively similar to the one observed for changes in the number of bookstores, but lacks statistical significance. If we take the point estimates for this outcome at face value, retail markets across the whole density distribution respond to introduction of online shopping in a similar manner, i.e., by reducing full-time employment. Consequently, the small differences in the effects for full-time workers cannot explain the heterogeneity observed for the total number of bookstore employees.

With respect to part-time and marginal workers, our heterogeneity analysis reveals that retail markets at higher percentiles of the bookstore density distribution experience larger decreases. Again, most of our point estimates are not statistically significant (with the only exception being the estimate for marginal employees in Column (5)), but the general patterns associated with them are in line with the heterogeneity observed for the total number of employees. This implies that our estimates for the total number of employees mask heterogeneous effects among employment categories. For example, considering low density retail markets, the very

Table 3.5 : Heterogeneity of IV Results

Percentile	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
	(1)	(2)	(3)	(4)	(5)
Panel A:	ΔBookstores				
ΔEOS	-1.895^{**}	-1.979^{**}	-2.077^{***}	-2.203^{***}	-2.367^{**}
	(0.934)	(0.828)	(0.767)	(0.804)	(1.017)
$\Delta EOS \cdot L$.Books.	−0.202	−0.202	−0.202	−0.202	−0.202
	(0.517)	(0.517)	(0.517)	(0.517)	(0.517)
Panel B:	ΔEmployees (Total)				
ΔEOS	−0.388	−3.091	−6.250	-10.30^{***}	-15.580^{***}
	(5.265)	(4.690)	(4.203)	(3.975)	(4.427)
$\Delta EOS \cdot L$.Books.	-6.491^{***}	-6.491^{***}	-6.491^{***}	-6.491^{***}	-6.491^{***}
	(2.308)	(2.308)	(2.308)	(2.308)	(2.308)
Panel C:	ΔEmployees (Full-Time)				
ΔEOS	−3.083	−3.188	−3.310	−3.466	−3.670
	(2.754)	(2.410)	(2.206)	(2.334)	(3.038)
$\Delta EOS \cdot L$.Books.	−0.251	−0.251	−0.251	−0.251	−0.251
	(1.607)	(1.607)	(1.607)	(1.607)	(1.607)
Panel D:	ΔEmployees (Part-Time)				
ΔEOS	−0.033	−0.908	−1.580	−2.442	−3.565
	(2.163)	(1.865)	(1.670)	(1.738)	(2.284)
$\Delta EOS \cdot L$.Books.	−1.381	−1.381	−1.381	−1.381	−1.381
	(1.271)	(1.271)	(1.271)	(1.271)	(1.271)
Panel E:	ΔEmployees (Marginal)				
ΔEOS	2.321	0.602	−1.406	−3.981	-7.339^{**}
	(3.473)	(3.133)	(2.878)	(2.833)	(3.254)
$\Delta EOS \cdot L$.Books.	-4.127^{***}	-4.127^{***}	-4.127^{***}	-4.127^{***}	-4.127^{***}
	(1.572)	(1.572)	(1.572)	(1.572)	(1.572)
First Stage					
ΔEOS^{bartik}	3.205^{***}	3.293^{***}	3.396^{***}	3.529^{***}	3.702^{***}
	(0.471)	(0.400)	(0.351)	(0.365)	(0.492)
$\Delta EOS^{bartik} \cdot L$.Books.	3.867^{***}	3.779^{***}	3.675^{***}	3.543^{***}	3.370^{***}
	(1.182)	(1.098)	(1.009)	(0.913)	(0.831)
Kleibergen-Paap F-Stat.	13.332	13.332	13.332	13.332	13.332
RLM-FE	Yes	Yes	Yes	Yes	Yes
Controls (incl. L.Books.)	Yes	Yes	Yes	Yes	Yes
N	792	792	792	792	792

Note: Dependent variables denote standardized changes in the number of bookstores, and employees therein, per 10,000 residents between 1999 and 201. All specifications include RLM-fixed effects. Standard errors are clustered at the RLM level. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Data: Nexiga GmbH (2017), Regionalstatistik, BHP (for a documentation of the BHP7514, see Schmucker et al. (2016)).

small (and statistically insignificant) reduction in the total number of employees seems to be the result of a replacement of full-time workers with marginal employees. In contrast, at the upper end of the distribution, the absolute large effects stem from a joint decline in all employment categories.

Summing up, our heterogeneity analysis indicates that retail markets with a higher density of bookstores in 1999 are more strongly affected by the introduction of online shopping. While retail markets at the lower end of the density distribution seem to substitute different types of employees (i.e., full-time vs marginal), we observe that bookstores in retail markets at the high end decrease employment across all types of workers. One possible explanation for this heterogeneity is that bookstores in less concentrated retail markets face higher competition from other bookstores and are therefore more vulnerable to the additional competition induced by online shopping.

3.6 Conclusion

The (local) decline of brick-and-mortar retailers is often associated with online shopping. Considering the case of traditional bookstores in Germany, we combine novel geomarketing data with administrative information on traditional bookstores to investigate this relationship empirically. Based on a long difference framework, we implement a Bartik-type IV strategy that exploits local variation in pre-internet age structures to identify the causal impact of regional online shopping exposure on the development of brick-and-mortar bookstores between 1999 and 2013.

In summary, we find that regional exposure to online shopping has a robust negative effect on the development of brick-and-mortar bookstores. For example, an increase in the uptake of online shopping by one standard deviation leads to a reduction in the local stock of bookstores by approximately 0.11 establishments per 10,000 residents. Relative to the levels in 1999 (i.e., the pre-internet era), this corresponds to a decline by 14%. In terms of total employment, we find effects of similar relative size.

Across different types of employment, we find that our effects are heterogeneous with respect to the density of bookstores in 1999. Although parts of our heterogeneity analysis lack statistical power, our point estimates indicate that all retail markets reduce full-time workers in a similar manner. While retail markets at the lower end of the density distribution seem to substitute these full-time workers by marginal employees, we find consistently negative effects for all types of employment at higher percentiles. Taking into account that retail markets at the upper end of this density distribution were less concentrated in 1999, one possible explanation for our results may be that the additional competition induced by online shopping affected those retail markets the most that were also more competitive in the past.

Finally, we have analyzed the impact of online shopping on the development of brick-and-mortar bookstores in Germany because the institutional setting is particularly well suited to identify how the availability of a convenient distribution channel affects traditional retailers. While the convenience aspect is arguably important for most customers and, therefore, also retailers, it may not be the dominant one in every industry. For example, in some industries, limited regional supply or the demand for niche products may be more important. It would therefore be promising to know whether our effects also persist in other industries and, in comparison, how much convenience matters relative to other channels. Furthermore, we view our paper as a step towards explaining the development of brick-and-mortar retailers as this constitutes an important aspect for city planners and local policy makers. As general equilibrium effects are therefore missing from our study (i.e., we do not investigate what happens to former employees of bookstores who loose their jobs), this topic provides further promising avenues for future research.

4 The Effect of Broadband internet on Entrepreneurship*

4.1 Introduction

How to ensure equivalence of living conditions and avert depopulation of rural areas? This is a key question faced by many local policy makers, who answer with calling for investment in adequate infrastructure in rural areas. At the top of the list: broadband internet expansion. According to the Digital Agenda for Europe, all households should have access to internet connections providing data downstream rates of at least 100 Mbps by 2025. Among others, this is intended to reduce the digital rural-urban divide and to provide rural households with high-speed internet access. Further, local policy makers typically hope for young, innovative companies to settle where the quality of internet connections is high.

We take this expectation to a test and examine the effect of local broadband internet availability on local business formation in mostly rural areas. We use data on establishment startups across German municipalities between 1992 and 2009 and combine it with municipality-level information on household broadband coverage at a data downstream rate of at least 384 Kbps. Identifying the causal effect of broadband internet on business formation is challenging since profit maximizing network operators do not roll-out infrastructure at random. They are inclined to develop infrastructure (first), where demand and willingness to pay are high. In order to determine the causal effect of broadband internet availability on business formation, we make use of an instrumental variable strategy which exploits technological peculiarities in broadband internet provision. In combination with a quasi-random allocation of important network infrastructure elements these technological peculiarities cause exogenous variation in broadband internet availability across rural German municipalities.

We find that broadband internet increases the overall establishment startup rate across the manufacturing, services and trade sector. In particular, our instrumental variables results suggest that an increase in broadband coverage by ten percentage points raises the number of new establishments per 100 employees by 0.28. Compared to the average startup rate between 1994 and 1996, prior to the introduction of broadband internet, this corresponds to an increase by 18.42%. We further find that the overall positive effect is driven by a rise in startup activity in the trade sector, while startup activity actually declines in the manufacturing sector. A complier analysis suggests that our IV estimates reflect the effect of broadband coverage for a sub-population of our sample municipalities that has (has no) broadband access because of technological peculiarities facilitating (complicating) broadband provision. This

* This chapter is based on joint work with Oliver Falck, and Bastian Stockinger

sub-population mainly consists of municipalities that are classified as rural by the Federal Institute for Building, Urban Affairs and Spatial Development (BBSR). Consequently, the results we find are not necessarily generalizable throughout Germany. This is important to keep in mind, when evaluating potential benefits of broadband deployment in a particular municipality or region. Finally, our placebo estimates generally support the validity of our results, in particular with regards to the positive and negative effects in trade and manufacturing, respectively.

This paper contributes to the literature on the economic effects of telecommunications service and broadband internet provision and adoption. A recent comprehensive review of this literature is provided by Bertschek et al. (2015). With regards to aggregate economic effects of telecommunications networks in general and broadband networks in particular, **?** and Czernich et al. (2011), respectively, have identified significant positive effects. Besides, Briglauer and Gugler (2017) examine the contribution of high-capacity broadband adoption in fostering economic growth. They find that the effect of fiber-based broadband internet adoption on GDP – net of basic broadband internet effects – is small but significantly different from zero. A variety of studies has dealt with productivity effects (e.g. Bertschek et al., 2013; Canzian et al., 2015; Colombo et al., 2013; De Stefano et al., 2014; Haller and Lyons, 2015) and labor market effects (e.g. Fabritz, 2013; Kandilov and Renkow, 2010; Kolko, 2012; Whitacre et al., 2014) of broadband internet. Overall, these studies mostly suggest a positive (but small) or zero effect. Akerman et al. (2015) and Atasoy (2013) further show that the gains from broadband provision are unequally distributed, depending on the skill distribution of employees and the tasks that need to be performed.

More closely related to our study are papers investigating broadband internet effects on firm entry and the stock of firms. There exist several studies that are similar to ours in terms of studying the effect of broadband provision in rural areas during the early 2000s. For example, Kim and Orazem (2017) study the effect of broadband provision in postcode areas in rural Iowa between 2000-2002. They pursue a differences-in-differences approach and find that broadband deployment significantly positively affects firm entry. Their results further suggest that positive effects are driven by rural postcode areas that are closely located to urban areas while remote postcode areas are unaffected. Kandilov and Renkow (2010) examine how firm stock responds to broadband internet provision. Similar to Kim and Orazem (2017), they find that the positive effects of a broadband loan program in rural U.S. postcode areas between 2002-2003 are driven by communities that are closely located to urban areas. Whitacre et al. (2014) investigate how broadband adoption, availability and bandwidth change firm stock in rural U.S. postcode areas between 2001 and 2010. While they can show positive effects for broadband adoption they do neither find broadband availability nor bandwidth to affect the stock of firms. Most closely related to our paper in terms of geographic scope and the time horizon investigated is a study by Audretsch et al. (2014). The authors use data on German counties for a time period between 2000-2005 to study the role of infrastructure, namely highway, railway, knowledge, and broadband infrastructure on startup rates across different sectors. With regards to broadband infrastructure, their results suggest a statistically significant and economically sizable effect on overall startup rates. They further show that the effects found concentrate in (consumer related) services and retail trade while they are rather small or even zero in manufacturing.

We advance the study by Audretsch et al. (2014) and the literature on broadband internet and business formation in general in two important ways. First, we use data on the smallest administrative regional unit in Germany – municipalities – to study the effect of broadband internet on establishment startup rates. Second, our fine-grained regional data allows us to pursue an instrumental variable approach to identify the causal effect of broadband internet provision on entrepreneurial activity. In spite of substantial research on the role of broadband internet in shaping entrepreneurship such causal evidence is still scant, in particular for Germany. Given the high cost of broadband internet expansion, more research is required that pins down the magnitude of effects in different scenarios. Think of policy makers. On ground of such estimates provided in the literature, they may draw conclusions about whether the expected economic gains of a particular broadband deployment project justify its cost.

The rest of this paper proceeds as follows. Section 4.2 introduces and discusses our empirical strategy. Section 4.3 presents the data at hand and provides descriptive statistics. Section 4.4 presents the results of our analysis and provides placebo estimates. Finally, Section 4.5 concludes.

4.2 Empirical Framework

To determine the relationship between broadband internet and business formation in rural areas we estimate the following model:

$$\Delta EA_m = \beta_1 \Delta BB_m + \Delta \mathbf{X}'_m \beta_2 + \alpha_r + \varepsilon_m \tag{4.2.1}$$

ΔEA_m indicates the difference in average startup rates in municipality m between the pre-internet period (1994-1996) and the internet period (2004-2006). Broadband coverage is zero before 1999. Therefore, ΔBB_m measures average broadband coverage in our sample municipalities in the internet period. Since we have data on broadband coverage from 2005 onwards only, we use broadband coverage in 2005 as a proxy for average broadband coverage between 2004 and 2006. It measures the share of households within a municipality that are covered by broadband networks, providing at least 384 kbps.

Estimating our model in first differences we implicitly control for potentially unobservable structural differences between municipalities that are constant over time. Nevertheless, there might be unobservable, time-varying differences that are correlated with broadband roll-out as well as changes in startup rates over time. Therefore, we control for changes in the local labor force. In particular, $\Delta \mathbf{X}_m$ accounts for changes in the share of female full-time employees, the share of foreign full-time employees, the share of medium skilled full-time employees, and the share of high-skilled full-time employees between the pre-internet and internet period.

Further, we include fixed effects, α_r, for regional labor markets (RLM). There exist 258 RLM across Germany, nesting more than 11,000 municipalities. Including these RLM fixed-effects, we account for common unobservable shocks that are shared by municipalities within the same RLM. For example, the location of a major company that triggers both, broadband deployment as well as the startup of (component) suppliers across *all* municipalities in the RLM. Further, since we are estimating our model in first differences, including RLM fixed-effects is equivalent to allowing for differences in linear trends in startup rates across RLMs. In summary, including RLM fixed-effects allows us to abstract from potential endogeneity concerns that arise from the fact that more aggregate differences (on the level of RLMs or higher) are correlated with both the development of startup rates over time as well as broadband coverage.

We address remaining concerns, arising from potentially unobservable, time varying confounding variables by instrumenting broadband coverage. The corresponding first stage is as follows:

$$\Delta BB_m = \theta_1 Fr4200_m + \Delta \mathbf{X'}_m \theta_2 + \alpha_r + \mu_m \tag{4.2.2}$$

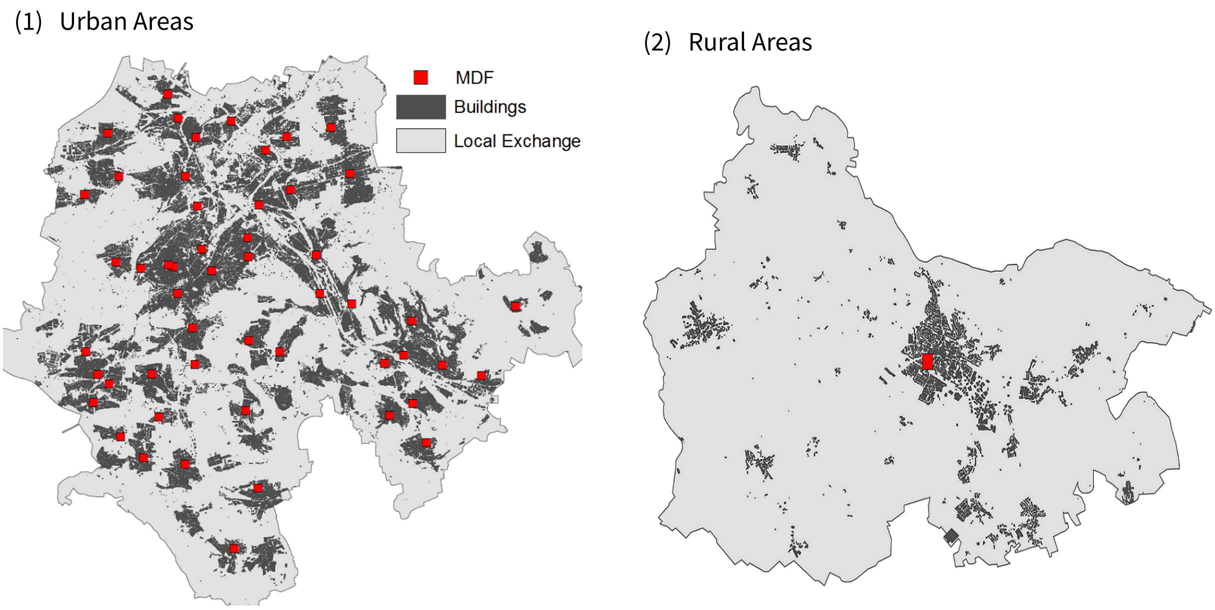
$Fr4200_m$ is our instrumental variable, while ΔBB_m, $\Delta \mathbf{X'}_m$, and α_r are as in Equation 4.2.1. In what follows, we explain our instrumental variable in more detail.

As in many countries across Europe, internet access in Germany until the late 1990s was restricted to dial-up connections. Eventually, faster broadband internet connections became available in 1999, when the Digital Subscriber Line (DSL) technology was introduced[1]. However, not everyone could instantly profit from the new technology. Many households still had to get along with much slower dial-up connections because of technological peculiarities in broadband internet provision via DSL. Whether a particular household was adversely affected - and thus could not instantly adopt the new technology - by these technological peculiarities hinged on it's geographic location. In particular, it was determined by the distance of the households to its main distribution frame (MDF), a node in the public switched telephone network (PSTN) connecting subscriber lines to the backbone network (see Figure A.2 in the Appendix). This is because households are connected to their MDF via copper cables. Since copper cables suffer line loss increasing in the length of the cable, DSL subscriptions could only be offered to households that were located close enough to their MDF. Specifically, DSL subscriptions at the lowest data downstream rate of 384 Kbps were marketed only if line loss was below 55dB. This was typically reached at a distance between households and their MDF of 4200 meters. Providing high-speed internet access to households that were located farther away required replacement of existing copper wires. Replacing copper wires involved costly construction works[2] causing considerable delay in DSL coverage of municipalities with high

[1] Note that in Germany DSL was the predominant broadband technology for many years. Even in 2006, 96% of all broadband connections were still realized via DSL, 3.5% via cable, 0.5% via power line or satellite (Bundesnetzagentur, 2006).

[2] The roll-out of fiber wire is particularly expensive as it needs to be rolled-out subsurface. Therefore it comes at a cost of 80,000 euros. An additional 10,000 euros need to be invested to install a node, connecting potentially remaining parts of copper wires to the fiber wire

Figure 4.1 : Number of MDFs in Rural and Urban LEs



Note: Figure (1) and (2) illustrate MDF location in a stereotypical urban and rural local exchange area (LE), respectively. Data: Deutsche Telekom, OpenStreetMap contributors (2018b).

shares of households that are located beyond a distance of 4200 meters from their MDF. The idea of our instrument is to calculate the share of a municipality's households that are located within 4200 meters from the MDF to which they are connected to. The rationale behind is as follows: the greater the share of favorably located households, the higher broadband coverage within the municipality.

Calculating our instrument is subject to three issues, which we can solve using geo-data and geo-processing software. First, we do not observe the exact location of households across Germany. However, we observe the location of buildings as well as their ground area. Instead of households, we thus rely on the buildings to calculate our instrument: the share of a municipality's total built-up area that is located within 4200 meters from its MDFs. Second, households are assigned to MDFs within local exchange areas[3] (LE) while our data is available on municipality level. Since LEs and municipalities are not nested within each other, buildings may be connected to MDFs in different LEs even though they are located in the same municipality. We take this into account by identifying the building ground area of a LE that is located within 4200 meters from its MDF in a first step. Using geo-information system (GIS) software, we recode this information to the municipality level in a second step. Third, when calculating our instrument, we also face the problem that LEs not always host a single MDF. Figure 4.1 illustrates that larger, more densely populated LEs typically host multiple MDFs in contrast to rural LEs, that typically host a single MDF. Since we do not observe the boundaries

[3] A local exchange area is a geographically restricted section of the telephony access network. Historically, a LE in Germany comprised all subscribers with the same local area code.

Table 4.1 : Average Sample vs. Average Western Municipality

	Western Germany		Sample	
	Mean	SD	Mean	SD
Employees (FT, in 1000)	1.99	13.36	0.61	1.26
No. of New Estab.	1.43	4.38	1.52	4.66
Share Female	0.29	0.14	0.29	0.15
Share Foreign	0.07	0.07	0.06	0.07
Share Medium-Skilled	0.71	0.15	0.71	0.15
Share High-Skilled	0.03	0.04	0.02	0.04
DSL (≥384 kbits)	0.76	0.21	0.74	0.22
Share Rural	0.63	-	0.76	-
Share Semi-Urban	0.13	-	0.12	-
Share Urban	0.23	-	0.12	-

Note: Table depicts differences between the average Western German municipality and the average municipality in our sample during 1994-1996. Data: TÜV Rheinland (Breitbandatlas), BHP 7510 (v. 2.2.1) (for documentation, see Gruhl et al. (2012)).

of the area served by a particular MDF, we can not clearly assign a particular building to its MDF in LEs that host multiple MDFs.

To avoid error in calculating our instrumental variable, we therefore restrict our sample to LEs with a single MDF[4]. As a consequence, we get a restricted set of municipalities which is the baseline sample of the analysis that follows. Figure 4.3 illustrates that this sample does not include municipalities around major German cities with 100,000 and more inhabitants. Consequently, our sample likely over-represents rural municipalities. This is also suggested by Table 4.1. It shows that about 76% of our sample municipalities are located in rural areas, while this is true for 63% of all Western German municipalities. Furthermore, the number of (full-time) employees is substantially smaller in our sample municipalities than the average Western municipality. However, the share of female full-time employees, foreign full-time employees as well as medium and highly skilled full-time employees is similar in the average Western and the average municipality in our sample.

[4] The specific procedure applied to derive the sample of LEs included in our analysis takes into account that we can clearly assign whether buildings are within 4200 meters from the MDF they are connected to irrespective of the total number of MDFs in their LE when i) they are located within 4200 meters of all MDFs or ii) they are located more than 4200 meters away from all MDFs in their LE. However, it turns out that this results in a sample of LEs with a single MDF

The sample restriction arising from our instrumental variables strategy is valuable for sample definition with regards to the political discussion about the lack of broadband infrastructure in rural areas. This is because the resulting sample only includes municipalities, where (at least some) buildings are located far away from their MDF. Consequently, the cost of deploying broadband infrastructure is particularly high in these regions. We argue that these regions comprise the municipalities local policy makers think of, when demanding for infrastructure deployment to ensure equivalence of living conditions and avert depopulation.

4.3 Data

We use fine grained municipality level data from two different sources, comprising yearly information on new business formation and broadband internet coverage for nearly all Western German municipalities between 1992-2009 and 2005-2009, respectively.

4.3.1 Business Formation

We measure the startup rate as the number of new establishments per 100 employees within a particular municipality and year. We follow Hethey and Schmieder (2010) and define a new establishment as an establishment, whose workforce newly came together to produce goods or services[5]. Apart from new innovative businesses associated with *entrepreneurship* this also includes traditional businesses such as bakeries or butcheries. The ability to measure business formation comprehensively is a strength of the paper since it allows us to abstract from applying any ad-hoc selection criteria regarding the type of businesses investigated. However it does not limit us in studying potentially heterogeneous effects of broadband internet availability on the creation of business of different types of establishments, including more and less “entrepreneurial” types.

We gather data on new establishments from the Establishment History Panel (EHP) of the Institute for Employment Research (IAB). For 1999-2009, it provides information on all German establishments in the administrative employment records of the Federal Employment Agency with at least one registered employee that is liable to social security contributions on June 30th. Some of the new establishments are operated by the proprietor without a single employee. We thus obtain a positive selection of new establishments in terms of their labor market impact and establishment age. However, there are good reasons to use this data. First, the data reliably covers all regular employees and employers across Germany. Since the data is used to

[5] In detail, new establishments are defined according to the concept of maximum clustered inflow (MCI). MCI refers to the single largest group of workers that work together at a new employer and have all worked together at the previous employer. We say an establishment is new if it’s MCI made up for at most 80% of the MCI’s previous employer’s workforce. If the threshold is exceeded, we only count those establishments as new, where MCI makes up for less than 30% of the new employer’s workforce. This definition includes spin-offs of existing establishments. Yet, robustness test indicate, that excluding spin-offs does not alter our results (c.p. Hethey and Schmieder (2010) for further detail).

Table 4.2 : Descriptive Statistics

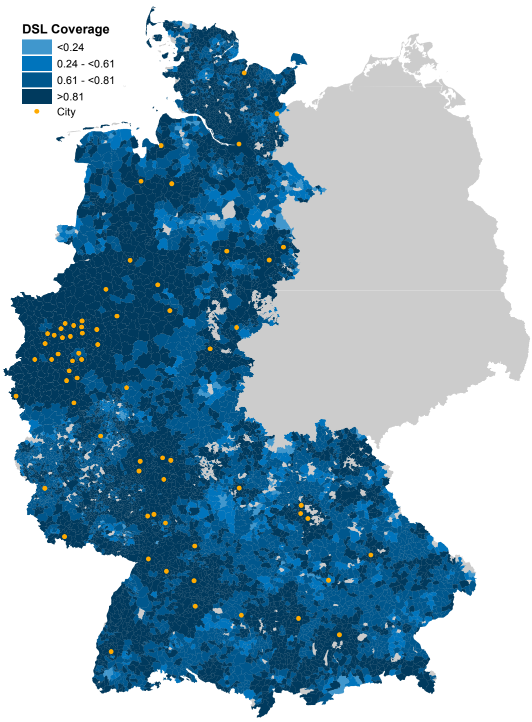
	I (1994-1996)		II (2004-2006)		$\Delta(II-I)$	
	Mean	SD	Mean	SD	Mean	SD
Panel A: Startup Rates (No. of new estabs. p. 100 empl.)						
Overall	1.52	4.66	1.23	3.25	–0.30	5.39
Manufacturing	0.20	1.28	0.15	0.66	–0.05	1.45
Services	0.99	3.99	0.80	2.65	–0.19	4.59
Knowledge-Int. Serv.	0.65	3.28	0.49	1.67	–0.15	3.66
Trade	0.33	1.96	0.28	1.52	–0.05	2.45
Panel B: Broadband Coverage & Instrumental Variable						
DSL (≥384 kbits)	0.00	-	0.74	0.22	0.74	0.22
Fr.4200	0.74	0.35	0.74	0.35	-	-
Panel C: Control Variables						
Share Female	0.29	0.15	0.29	0.15	0.00	0.14
Share Foreign	0.06	0.07	0.05	0.06	-0.01	0.07
Share Medium-Skilled	0.71	0.15	0.67	0.17	-0.05	0.18
Share High-Skilled	0.02	0.04	0.03	0.05	0.01	0.05

Note: Table depicts municipality averages between 1994-1996 and 2004-2006. DSL coverage as of 2005. Fr4200 measures the share of the total building ground of a municipality that is located within a radius of 4200 meters around their MDF. Data: TÜV Rheinland (Breitbandatlas), BHP 7510 (v. 2.2.1) (for documentation, see Gruhl et al. (2012)), OpenStreetMap contributors (2018b)

calculate social security contributions, misreporting is subject to potentially severe pecuniary sanctions. Second, the data contains a wealth of information on employee demographics such as gender, nationality, gross daily wages, type of employment (e.g. full-time, part-time) and qualification, among others. Third, establishment-specific information includes industry codes, information on establishment size and establishment location at the municipality level. We use industry codes to identify entering establishments' sectoral affiliation and to categorize their activities, with respect to knowledge intensity. Since we are interested in business formation on municipality level, it is important that the location of new businesses is reported unambiguously. In this context, observing establishments rather than firms is an advantage. While firms may run multiple businesses in various locations, establishments are clearly defined as one or several business units operating in the same industry and locality. We are thus confident to identify all new establishment that start-up within a municipality in a particular year.

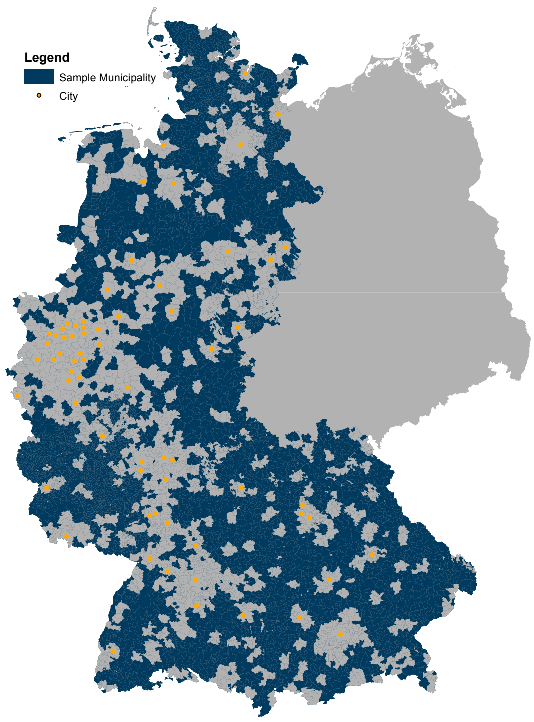
Table 4.2 summarizes our data for the municipalities in our sample. Between 1994-1996, on average 1.52 establishments per 100 employees entered the business across manufacturing,

Figure 4.2 : Average DSL Coverage 2005-2009



Note: Figure illustrates DSL coverage at a data downstream rate of at least 384 kbits across Germany in 2005. Own illustrations. Data: TÜV Rheinland (Breitbandatlas),OpenStreetMap contributors (2018b). Map: GeoBasis-DE / BKG 2014.

Figure 4.3 : Sample Municipalities



Note: Figure illustrates the sample of (rural) municipalities that are included in our analysis. Red marks illustrate cities with 100,000 inhabitants or more. Own illustrations. Data: OpenStreetMap contributors (2018b). Map: GeoBasis-DE / BKG 2014.

services, or trade. The service sector accounts for more than 65.13% of this overall startup rate. The trade sector accounts for 21.71%, and manufacturing for 13.16%. Startup rates generally decrease over time. Table 4.2 illustrates that they are on average smaller across all sectors during 2004-2006 than during 1994-1996.

4.3.2 Broadband Coverage

Data on broadband coverage stems from the German broadband atlas (Breitbandatlas Deutschland). It includes yearly information on broadband internet availability across German municipalities between 2005 and 2009. In detail, it indicates the share of households within a municipality that has access to broadband internet at a minimum data downstream transfer rate of 384 kbps. For our empirical analysis, we rely on broadband coverage in 2005 as a proxy for average broadband coverage between 2004 and 2006. Table 4.2 shows that in 2005, 74% of all households in the average municipality of our sample were able to use DSL at such bandwidth. Further, Figure 4.2 illustrates regional difference in broadband coverage across German municipalities. It shows that broadband coverage is generally higher in urban areas and municipalities close to major German cities. Instead, more remote municipalities tend to depict lower rates of broadband coverage.

4.4 Results

4.4.1 Baseline Results

Table 4.3 presents our baseline results regarding the effect of DSL on the overall establishment startup rate. While Column (1) to (3) provide results for our OLS estimations, Columns (4) to (6) present the corresponding IV results. We include RLM-fixed-effects in all specifications to allow for differential linear trends in startup rates across labor markets. In Column (2) and (5), we control for changes in the composition of the population of the municipalities considered. Furthermore, we include fixed-effects in Column (3) and (6) that allow for differences in the trend of startup rates that are shared by municipalities in rural, partly urban, and urban regions, respectively.

Throughout all OLS as well as IV specifications we find a statistically significant positive effect of DSL coverage on the overall startup rate in manufacturing, services and trade. Coefficients do not vary substantially across the different OLS specifications in Columns (1) to (3) and across the different IV specification in Columns (4) to (6). Yet, comparing OLS coefficients with the respective results from IV regressions reveals that IV coefficients are generally larger. For example, the IV coefficient from our most comprehensive specification in Column (6) is more than three times larger than the respective OLS coefficient in Column (3). While a downward bias is generally possible, endogenous roll-out favoring economically strong areas would rather suggest OLS coefficients to be upward biased. According to Imbens and Angrist (1994), one potential explanation for such differences between OLS and IV results could be as follows: rather than an average treatment effect (ATE), our IV estimates might reflect a local average treatment effect (LATE) for the sub-population of municipalities that are covered by DSL networks to the extend that is suggested by the share of buildings within 4200 meters from their MDF. This means that municipalities comply with their instrument, i.e. municipalities with a high share of buildings that are located within a radius of 4200 meters from their MDFs have a high DSL coverage while municipalities with a low share have a low DSL coverage. To examine this, we follow the approach suggested by Akerman et al. (2015) and calculate the share of compliers among rural, partly urban, and urban municipalities. To that end, we estimate the following first stage equation for the whole sample and the sample of rural, partly urban, and urban municipalities separately:

$$\Delta BB_m = \theta_1^s Fr4200_m + \alpha_s + \mu_m \tag{4.4.1}$$

ΔBB_m is again DSL coverage in 2005, $Fr4200_m$ is our instrumental variable, and α_s are three dummy variables that are equal to one if a municipality is characterized as a rural, partly urban, and urban municipality, respectively. It is zero otherwise. In the next step, we calculate the share of compliers in each of the three groups, proceeding as follows: we calculate the ratios

Table 4.3 : Baseline Results

	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
Δ DSL	0.779*	0.830**	0.811*	2.957**	2.704**	2.763*
	(0.407)	(0.420)	(0.419)	(1.345)	(1.376)	(1.498)
Δ Frac. Female		0.752	0.750		0.741	0.737
		(1.893)	(1.893)		(1.867)	(1.865)
Δ Frac Foreig.		-9.776***	-9.787***		-9.871***	-9.883***
		(3.550)	(3.549)		(3.577)	(3.574)
Δ Frac. Med. Skilled		-3.393**	-3.393**		-3.401**	-3.399**
		(1.592)	(1.592)		(1.588)	(1.588)
Δ Frac. High Skilled		-4.183	-4.204		-4.189	-4.172
		(2.706)	(2.706)		(2.688)	(2.690)
First stage						
Fr.4200				0.158***	0.158***	0.150***
				(0.017)	(0.017)	(0.018)
Kleib.-Paap F-Stat.				86.30	85.98	72.09
RLM-FE	Yes	Yes	Yes	Yes	Yes	Yes
Type FE	No	No	Yes	No	No	Yes
N	5870	5870	5870	5870	5870	5870

Note: Dependent variable is the change in the average startup rate between 1994-1996 and 2004-2006. Δ DSL measures DSL coverage in 2005. RLM-FE refer to fixed effects on the level of regional labor markets. Type FE identify rural, partly urban, and urban municipalities. Standard errors are clustered at the RLM level. Significance levels: p<0.10, **p<0.05, ***p<0.01.

between the group specific $\hat{\theta}_1^s$ and the overall $\hat{\theta}_1$ and multiply it by the share of municipalities belonging to the respective group. The share of compliers in group s is thus defined as:

$$Compliers^s = \frac{\hat{\theta}_1^s}{\hat{\theta}_1} \times \frac{N^s}{N} \tag{4.4.2}$$

where N_s and N indicates the number of municipalities in group s and the overall number of municipalities in our sample, respectively. Column (3) of Table 4.4 illustrates that municipalities in rural regions are highly over-represented among the compliers. Column (4) further shows that in the rural municipalities employment levels are substantially smaller compared to partly urban and urban municipalities. This explains why our IV effects are larger than our OLS effects. A single additional establishment startup in a municipality with a low level of employment increases the startup rate more than a single establishment startup in a municipality with a high level of employment. Consequently, the large difference in the size

Table 4.4 : Analyzing Complier Municipalities

	Composition of Sample (1)	$\hat{\theta}$ (2)	Composition of Compliers (3)	Avg. No. of Employees (4)	Avg. Startup Rate (94-96) (5)	Number of obs. (6)
Rural	0.76	0.16	0.81	348.73	1.66	4,472
Partly Urban	0.12	0.07	0.06	2226.92	1.04	686
Urban	0.12	0.07	0.06	1936.18	1.09	712
Total	1	0.15	1	759.14	1.52	5870

Note: We partition the baseline sample into three groups, consisting of rural, partly urban and urban municipalities. Column (1) depicts the size of the groups relative to the overall sample. Column (2) presents results from estimating equation 4.4.1 for the respective groups. Column (3) shows the distribution of compliers across different groups of municipalities, which is calculated according to formula 4.4.2. Columns (4) and (5) characterize complier municipalities with regards to their average number of employees and their average startup rate between 1994 and 1996. Column (6) provides the number of observations in the respective group.

of our estimated coefficients between our OLS and IV specifications is likely explained by the fact that the IV reflects a group-specific effect for rural municipalities rather than an average treatment effect. Keeping this in mind, the most comprehensive specification of Column (6) in Table 4.3 suggests that an increase in DSL coverage by ten percentage points increases the number of new establishments per 100 employees by approximately 0.28. DSL coverage in the average municipality is approximately 0.74. Consequently, expanding the DSL network in the average municipality to full coverage would result in 0.72 new establishments per 100 employees. This corresponds to 47.37% of the average startup rate between 1994 and 1996 (1.52). Focusing on the average rural municipality (as suggested by our IV estimates, reflecting a LATE) gives similar results since DSL coverage is somewhat lower, while the average startup rate between 1994 and 1996 is slightly larger. In summary, our coefficient estimate suggests that expanding the DSL network in the average rural municipality to full coverage would results in an increase of the startup rate in rural municipalities by 48.27% compared to the respective average startup rate between 1994 and 1996.

In Table 4.5 we examine whether DSL coverage has a different impact on establishment startup rates across different sectors of the economy. Column (1) repeats our most comprehensive IV specification from Column (6) in Table 4.3. Column (2) to (5) present the corresponding results for the manufacturing sector, the service sector, the trade sector, and the knowledge-intensive service sector. Given the rather small levels of average startup rates between 1994-1996, all our estimates are notable in terms of their economic size. Yet, the degree of precision that is associated with some of our estimates prevents us from drawing strong conclusions with regards to the effect of DSL coverage on startup rates in some of these sectors. However, manufacturing and trade are an exception. Column (2) of Table 4.5 suggests that an increase in DSL coverage by ten percentage points statistically significantly reduces the startup rate in the manufacturing sector by about 0.06. Given a baseline level for the average startup rate in

Table 4.5 : Results by Sector

	Overall (1)	Manufac. (2)	Services (3)	Trade (4)	KI-Services (5)
Δ DSL	2.763*	-0.593*	1.322	2.034***	1.150
	(1.498)	(0.350)	(1.296)	(0.772)	(1.000)
Δ Frac. Female	0.737	0.373*	0.839	-0.475	0.834
	(1.865)	(0.195)	(1.618)	(0.402)	(1.604)
Δ Frac Foreig.	-9.883***	-0.708***	-7.037**	-2.138***	-6.133**
	(3.574)	(0.260)	(3.530)	(0.562)	(2.796)
Δ Frac. Med. Skilled	-3.399**	-0.155	-2.816**	-0.427	-1.520
	(1.588)	(0.212)	(1.415)	(0.429)	(1.299)
Δ Frac. High Skilled	-4.172	-0.346	-3.200	-0.626	0.0565
	(2.690)	(0.411)	(2.428)	(0.758)	(1.968)
First stage					
Fr. 4200	0.150***	0.150***	0.150***	0.150***	0.150***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Kleib.-Paap F-Stat.	72.09	72.09	72.09	72.09	72.09
LMR FE	Yes	Yes	Yes	Yes	Yes
Type FE	Yes	Yes	Yes	Yes	Yes
N	5870	5870	5870	5870	5870

Note: Dependent variable is the change in average startup rates between 1994-1996 and 2004-2006. Δ DSL measures DSL coverage in 2005. The different Columns present results for the respective sectors indicated in the Column heads. RLM-FE refer to fixed effects on the level of regional labor markets. Type FE identify rural, partly urban and urban municipalities. Standard Errors are clustered at the RLM level. Significance levels: $p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

the manufacturing sector between 1994 and 1996 of 0.2, this corresponds to a reduction of 30%. Column (4) suggests that an increase in DSL coverage by ten percentage points increases startup rates in the trade sector by 0.2 percentage points. This corresponds to 60.6% of the sector's average startup rate between 1994 and 1996.

In summary, we find a statistically significant and economically sizable effect of DSL coverage on the joint startup rates across the manufacturing, service and trade sector in a sample of rural German municipalities, where broadband expansion tends to be rather expensive. Further, our results suggest that the effect is driven by the trade sector, while startup rates in the manufacturing sector may even decrease in response to rising DSL coverage. Pursuing an instrumental variables approach, we identify a local average treatment effect for the sub-population of municipalities that are covered by DSL to the extend that is suggested by the

Table 4.6 : Placebo Estimations

	Overall (1)	Manufac. (2)	Services (3)	Trade (4)	KI-Services (5)
Δ DSL	2.506	0.411	1.247	0.848	1.260
	(1.657)	(0.658)	(0.810)	(1.085)	(1.058)
Δ Frac. Female	1.041	0.0992	1.742	-0.800	2.158**
	(1.730)	(0.426)	(1.256)	(0.835)	(1.010)
Δ Frac Foreig.	1.442	-0.229	0.861	0.809	-1.288
	(2.661)	(0.260)	(2.316)	(1.709)	(1.458)
Δ Frac. Med. Skilled	-2.768*	-0.563	-1.182	-1.023	-1.923**
	(1.440)	(0.475)	(0.834)	(1.269)	(0.933)
Δ Frac. High Skilled	0.142	-0.882	-0.903	1.927	-0.970
	(2.172)	(0.616)	(1.269)	(1.435)	(1.790)
First stage					
Fr. 4200	0.149***	0.149***	0.149***	0.149***	0.149***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Kleib.-Paap F-Stat.	74.19	74.19	74.19	74.19	74.19
LMR FE	Yes	Yes	Yes	Yes	Yes
Type FE	Yes	Yes	Yes	Yes	Yes
N	5897	5897	5897	5897	5897

Note: Dependent variable is the change in average startup rates between 1992-1994 and 1997-1999. Δ DSL measures DSL coverage in 2005. The different Columns present results for the respective sectors indicated in the Column heads. RLM-FE refer to fixed effects on the level of regional labor markets. Type FE identify rural, partly urban and urban municipalities. Standard Errors are clustered at the RLM level. Significance levels: $p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

share of buildings within 4200 meters around their MDF. Therefore, our results mainly reflect the effect for rural municipalities, where employment levels are generally low.

4.4.2 Placebo Estimates

To substantiate the validity of our findings we provide a placebo test that consists in estimating the same model as before but for a time period prior to the introduction of DSL. Consequently, for our instrumental variables approach to be valid, a necessary condition for our placebo study is to find zero effects. In a first step, we calculate average startup rates for an imaginary pre-DSL period (1992-1994) and an imaginary DSL period (1997-1999). Importantly, both time periods are before DSL was actually introduced[6]. In the second step, we calculate the change in startup rates as well as changes in the composition of the population between our two placebo time periods. In the third step, we regress the changes in startup rates between our imaginary pre-internet and internet period on DSL coverage in 2005 as well as the respective control variables.

Table 4.6 presents the results of our placebo test with regards to the effect of broadband internet on startup rates across different sectors. All coefficient estimates are not statistically significantly different from zero. Generally, this speaks in favor of the validity of our instrumental variables results. For a placebo test to be credible, point estimates should not only be insignificantly estimated but also small in terms of their economic size. This is not generally true. In fact, the coefficient estimates of our placebo estimates for the effect of DSL on overall startup rates as well as startup rates in the service and knowledge-intensive service sector are quite sizable. However, placebo estimates for the manufacturing and the trade sector speak in favor of the negative and positive effects found in our instrumental variables estimations. For the manufacturing sector, our placebo estimate suggests that, if at all, higher rates of DSL coverage are correlated with unobservable factors that are positively related to the startup rate in manufacturing. If anything, our IV estimate would thus be upward biased, implying that we would underestimate the negative effect found. With regards to the trade sector, we find that our placebo coefficient in Column (4) of Table 4.6 is not only estimated less precisely but also substantially smaller (about 40% the size of the IV coefficient). This is reassuring with regards to our finding of positive effects of DSL coverage on establishment startup in the trade sector across rural German municipalities.

4.5 Conclusions

For many local policy makers across Europe, ensuring equivalence of living conditions and economic strength hinges on the provision of adequate broadband internet infrastructure, not least because they expect young, innovative companies to settle where the quality of internet connections is high. In this paper, we examine whether this expectation is true analyzing the effect of broadband internet on business formation in rural municipalities across Germany.

We combine data on the entire number of establishment startups across German municipalities between 1992 and 2009 with municipality-level data on broadband coverage at a data downstream rate of at least 384 Kbps. To estimate causal effects, we exploit technological peculiarities in the provision of broadband internet via DSL technology that - depending on their geographic location within the municipality - excluded some households across Germany from readily adopting the new technology, when it became available in 1999.

[6] DSL was introduced in Germany in 1999 only in a few large cities across Germany. None of these cities is comprised in our sample. We can thus rely on 1999 as pre-internet period.

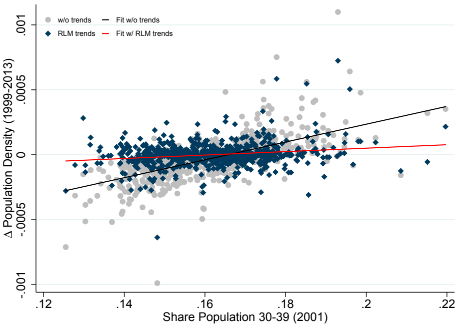
Our results indicate that an increase in DSL coverage by ten percentage points increases the overall startup rate across manufacturing, services, and trade by 0.28, i.e. and additional 0.28 new establishments per 100 employees. The effect turns out to be driven by the trade sector, while the startup rate in the manufacturing sector even decreases. Furthermore, we find that our instrumental variable results mainly reflect the effect for the sub-population that is (is not) covered by DSL because of the fact that many households could (could not) readily adopt DSL as a result of being favorably (unfavorably) located. A more detailed analysis shows that this mostly concerns rural municipalities where employment levels are generally small, such that a single establishment startup affects the startup rate substantially. Overall, this suggests that the effect of broadband internet on establishment startup rates may be smaller in the average German municipality.

Even though the relationship between broadband internet and entrepreneurship attracts substantial interest from both policy makers as well as academics, robust, causal evidence quantifying the effect is surprisingly rare. Yet, reliable estimates are of major importance for policy makers who want to evaluate, for example, the potential benefits relative to the cost of a broadband deployment project. In summary, this paper contributes to catering demand for causal evidence on the relationship between broadband provision and entrepreneurship, providing robust empirical evidence for rural municipalities across Germany.

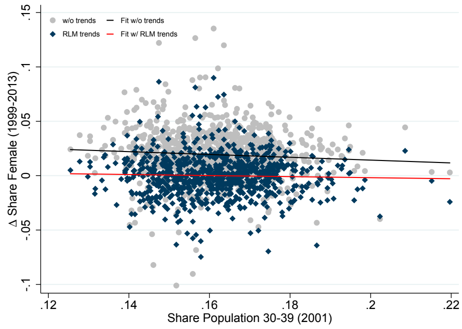
A Figures

Figure A.1 : Relationship Between (Pre-Internet) Age Shares and Changes in Control Variables

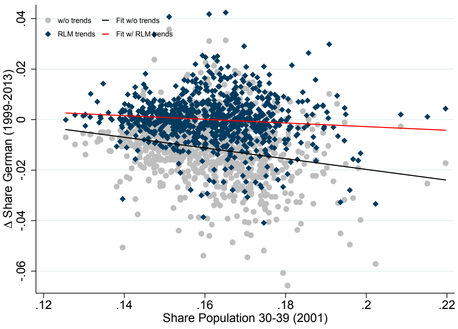
(1) Population density



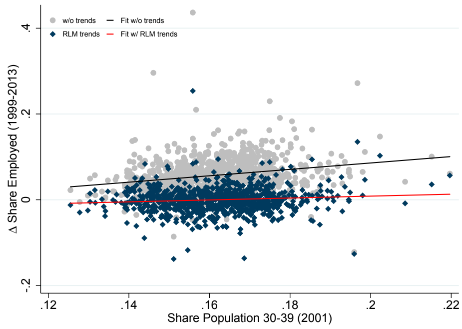
(2) Share female



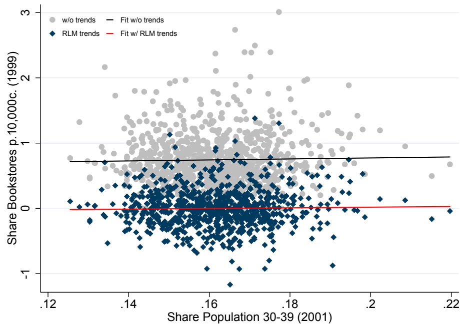
(3) Share German



(4) Share employed

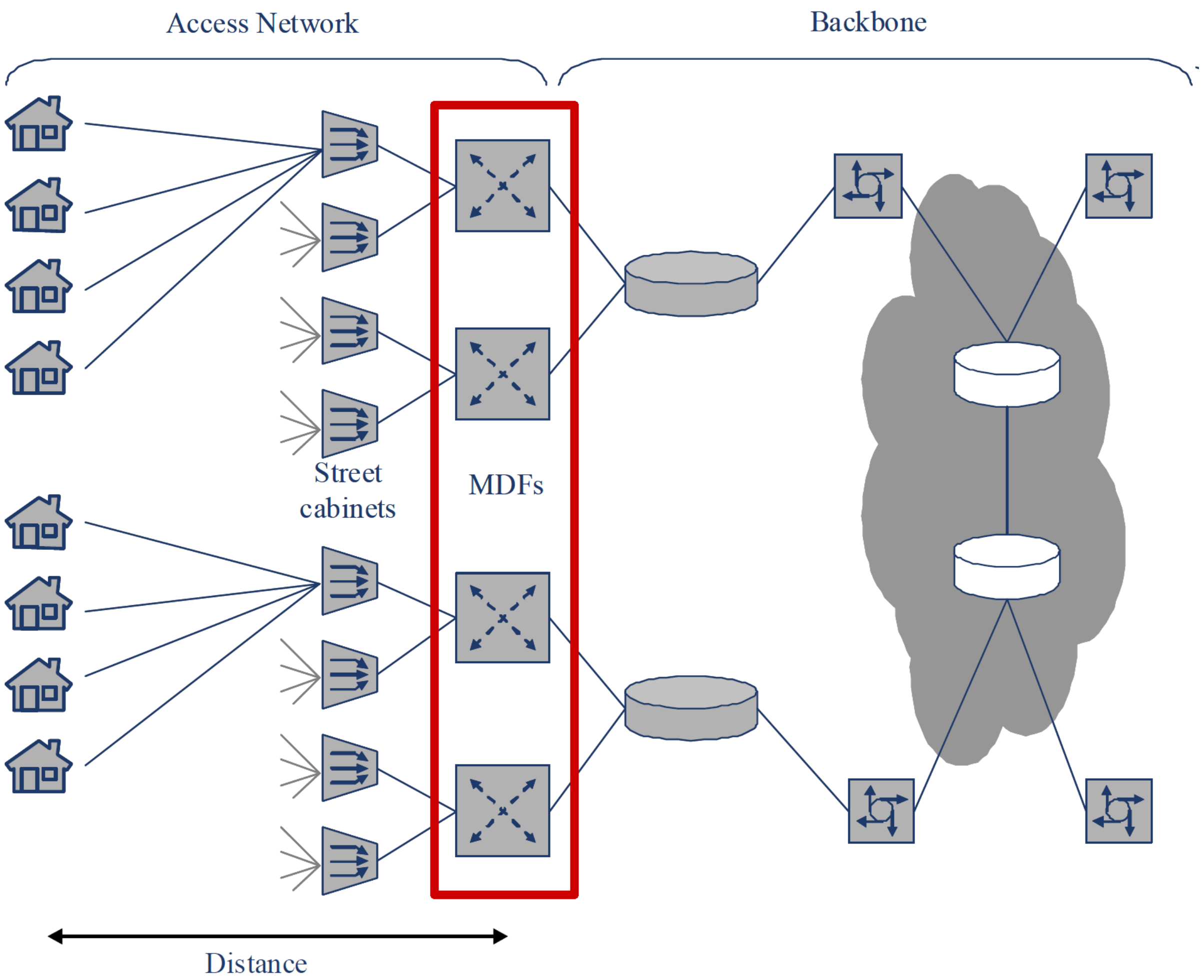


(5) Share employed



Note: This figure plots the relationship between the share of 30 to 39 year old people in 2001 and the change in a selection of control variables between 1999 and 2013. Gray dots (black lines) indicate scatter plots (OLS regression coefficients) of the unconditional relationship between the change in the control variable of interest and the pre-Internet share of people between 30 and 39. Blue diamonds (red lines) represent added variable scatter plots (OLS regression coefficients) of the same relationships after accounting for RLM-fixed effects. Data: Nexiga GmbH (2017), Regionalstatistik, BHP (for a documentation of the BHP7514, see Schmucker et al. (2016)).

Figure A.2 : Structure of the Public Switched Telephone Network



Note: Figure illustrates the structure of the public switched telephone network (c.p. Falck et al. (2014)

B Tables

Table B.1 : Summary Statistics Along the Distribution Bookstore Density in 1999

Percentiles	$(0,10]$	$(10,25]$	$(25,50]$	$(50,75]$	$(75,90]$	$(90,100]$
	(1)	(2)	(3)	(4)	(5)	(6)
Local Book Markets in 1999						
Bookstores	0.27	0.43	0.60	0.81	1.10	1.63
Employees (total) p.b.	5.76	5.54	4.70	4.73	4.40	3.63
Employees (full-time) p.b.	2.46	2.34	1.92	2.03	1.82	1.36
Employees (part-time) p.b.	0.93	1.00	0.82	0.81	0.80	0.56
Employees (marginal) p.b.	1.92	1.85	1.64	1.62	1.56	1.50
Local Population in 1999						
Population (total, tsd.)	49.17	72.92	87.39	121.27	97.81	77.60
Population Density	0.11	0.14	0.16	0.17	0.15	0.11
Employees (total, tsd.)	15.60	25.2	30.71	48.81	41.30	31.20
Share Employed	0.32	0.33	0.33	0.35	0.36	0.36
Share Female	0.47	0.47	0.47	0.48	0.48	0.48
Share German	0.94	0.95	0.95	0.95	0.94	0.94
Share Medium Skilled	0.73	0.74	0.74	0.74	0.72	0.73
Share High Skilled	0.06	0.07	0.07	0.08	0.08	0.07
Average Hourly Wages	48.93	53.50	56.90	60.40	64.72	66.97
N	80	118	198	198	119	79

Note: Variables describing the structure of the local book market are either measured per 10,000 residents (bookstores) or per bookstore (total number of employees as well as full-time, part-time, and marginal employees). The total number of residents and employees is expressed in thousands. Population density is measured in terms of 100 residents per km^2. Average hourly wages are calculated for all full-time workers within a retail market. Data: Nexiga GmbH (2017), Regionalstatistik, BHP (for a documentation of the BHP7514, see Schmucker et al. (2016)).

Table B.2 : Baseline OLS Results (Not Purged from ME)

	lnBandwidth			lnLatency		
	(1)	(2)	(3)	(4)	(5)	(6)
Mobile Quality	-0.00191	-0.00188	-0.00165	0.00853***	0.00778***	0.00674***
	(0.00146)	(0.00142)	(0.00139)	(0.00141)	(0.00143)	(0.00142)
Log of FL p.hh. ’11	0.0482***	0.0465***	0.0422***	0.0467***	0.0455***	0.0417***
	(0.00585)	(0.00592)	(0.00593)	(0.00579)	(0.00586)	(0.00588)
Log HH. in ’11	-0.00332*	0.000728	0.00294	-0.00229	0.000653	0.00275
	(0.00178)	(0.00195)	(0.00202)	(0.00176)	(0.00191)	(0.00199)
HH Dens. ’11	-0.0332***	-0.0309***	-0.0267***	-0.0322***	-0.0305***	-0.0263***
	(0.00413)	(0.00420)	(0.00449)	(0.00408)	(0.00414)	(0.00444)
Cov. Cable (Yes/No)	-0.0398***	-0.0382***	-0.0387***	-0.0392***	-0.0378***	-0.0384***
	(0.00317)	(0.00316)	(0.00313)	(0.00313)	(0.00313)	(0.00311)
lnDown. p.hh ’11	0.00141	0.00452	0.00634	0.00341	0.00578	0.00728*
	(0.00382)	(0.00384)	(0.00386)	(0.00379)	(0.00380)	(0.00383)
lnArea Parks p.hh		0.00208	0.00352**		0.000902	0.00258
		(0.00155)	(0.00171)		(0.00155)	(0.00171)
lnArea Lakes p.hh		-0.000760	-0.000366		-0.000951	-0.000535
		(0.000668)	(0.000671)		(0.000663)	(0.000667)
lnDist. River		0.000715	-0.000582		0.000626	-0.000573
		(0.00157)	(0.00156)		(0.00154)	(0.00153)
lnDist. Station		0.0110***	0.00910***		0.0105***	0.00878***
		(0.00154)	(0.00161)		(0.00152)	(0.00159)
Frac. >65 ’11			0.0526			0.0317
			(0.0449)			(0.0443)
Frac. Students ’11			0.314*			0.291*
			(0.178)			(0.176)
Frac. Sec. ’11			-0.161***			-0.151***
			(0.0373)			(0.0368)
Frac. Tert. ’11			-0.00937			-0.00509
			(0.0440)			(0.0438)
Frac. Female ’11			-0.453***			-0.405***
			(0.126)			(0.127)
Frac. Unempl. 2011			-0.0441			-0.0442
			(0.0565)			(0.0564)
Frac. Foreig. ’11			-0.102***			-0.0987***
			(0.0370)			(0.0369)
Frag. FE	Yes	Yes	Yes	Yes	Yes	Yes
N	2127	2127	2127	2127	2127	2127

Note: Dependent variable is the percentage change in incumbent fixed-line market penetration between 2011 and 2014. Mobile Quality as indicated in the column head. Frag. FE refers to a set of dummy variables indicating the minimum number of a postcode’s separated built-up areas to cover ≥ 50% of its total built-up area. Huber-White standard errors applied. Significance levels: *p<0.10, **p<0.05, ***p<0.01.

Table B.3 : The Role of Population Composition in Fixed-Mobile Voice Substitution (Latency)

	(1) 10th percentile	(2) 50th percentile	(3) 90th percentile
Panel A: Fraction Population > 65 in 2011			
Mobile Quality	0.0217**	0.0162**	0.00897
	(0.00901)	(0.00733)	(0.00833)
Mobile Quality x Frac. > 65	-0.00509	-0.00509	-0.00509
	(0.00381)	(0.00381)	(0.00381)
Panel B: Fraction Population Max. Tertiary Educated in 2011			
Mobile Quality	0.0182**	0.0162**	0.0122
	(0.00921)	(0.00759)	(0.00767)
Mobile Quality x Frac. tert. educ.	-0.00319	-0.00319	-0.00319
	(0.00491)	(0.00491)	(0.00491)
Panel C: Fraction Population Unemployed in 2011			
Mobile Quality	0.00280	0.0124*	0.0303***
	(0.00885)	(0.00730)	(0.0116)
Mobile Quality x Frac. unemployed	0.0148**	0.0148**	0.0148**
	(0.00751)	(0.00751)	(0.00751)
Panel D: Fraction Foreigners in 2011			
Mobile Quality	-0.0134	0.00705	0.0510***
	(0.0134)	(0.00787)	(0.0191)
Mobile Quality x Frac. foreigners	0.0306**	0.0306**	0.0306**
	(0.0136)	(0.0136)	(0.0136)
Frag. FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
N	2127	2127	2127

Note: Dependent variable is the percentage change in incumbent fixed-line market penetration between 2011 and 2014. Mobile Quality measured as latency. Frag. FE refers to a set of dummy variables indicating the minimum number of a postcode's separated built-up areas to cover ≥ 50% of its total built-up area. Controls include all controls of Column (6) in Table 1.2. Throughout all specification *Mobile Quality* is purged from measurement error by instrumenting it with higher quality mobile coverage. In Column (1), (2), and (3), the panels' respective demographic variable is standardized to have a value of zero at the 10th, 50th, 90th percentile. Huber-White standard errors applied. Significance levels: *p<0.10, **p<0.05, ***p<0.01.

Table B.4 : The Effect of Hybrid Coverage on Competition (OLS)

	(1)	(2)	(3)	(4)
Min. Hybrid M (2015)	0.0120***	0.0128***	0.0152***	0.0158***
	(0.00446)	(0.00467)	(0.00451)	(0.00457)
Log FL p.hh. (2014)	-0.00647	-0.00846	-0.0438***	-0.0468***
	(0.00789)	(0.00892)	(0.0104)	(0.0106)
Log Pop. (2014)		-0.00141	-0.00456**	-0.00437**
		(0.00187)	(0.00178)	(0.00183)
HH density (2014)		0.00212	-0.00443	-0.00195
		(0.00658)	(0.00575)	(0.00611)
Cov. LTE op. A (2016)			-0.000486	-0.0000941
			(0.00478)	(0.00468)
Cov. LTE op. B (2016)			-0.00530	-0.00476
			(0.00398)	(0.00393)
Cov. FTTX (2016)			-0.0380***	-0.0382***
			(0.00653)	(0.00661)
Cov. HFC (2016)			-0.0402***	-0.0427***
			(0.00484)	(0.00507)
Cov. 6-16 Mbps inc. (2016)			0.0112	0.0131
			(0.0208)	(0.0211)
Cov. 16-50 Mbps inc. (2016)			0.0499***	0.0493***
			(0.0105)	(0.0105)
Cov. ≥50 Mbps inc. (2016)			0.0940***	0.0932***
			(0.0106)	(0.0105)
Share Pop. 18-34 (2014)				-0.0407
				(0.139)
Share Pop. 35-64 (2014)				0.0725
				(0.140)
Share Pop. ≥65 (2014)				0.101
				(0.0899)
Share Female (2014)				0.0798
				(0.156)
County FE	Yes	Yes	Yes	Yes
LEs	5142	5142	5142	5142

Note: Dependent variable is the percentage change in incumbent fixed-line market penetration between 2014-2016. Min. Hybrid M is the share of households that are covered by the hybrid product. All regressions are weighted with the number of households in 2014. Standard errors clustered at county level. Significance levels: $^{*}p < 0.1,^{**} 0.05,^{***} p < 0.01$.

Table B.5 : The Effect of Hybrid Coverage on Competition (IV)

	(1)	(2)	(3)	(4)
Min. Hybrid M (2015)	0.0154	0.0169	0.0227**	0.0243**
	(0.0103)	(0.0114)	(0.0108)	(0.0108)
Log FL p.hh. (2014)	-0.00523	-0.00820	-0.0437***	-0.0468***
	(0.00808)	(0.00859)	(0.0100)	(0.0103)
Log Pop. (2014)		-0.00176	-0.00505***	-0.00490**
		(0.00207)	(0.00189)	(0.00192)
HH density (2014)		0.00206	-0.00445	-0.00184
		(0.00637)	(0.00560)	(0.00594)
Cov. LTE op. A (2016)			-0.000854	-0.000514
			(0.00469)	(0.00460)
Cov. LTE op. B (2016)			-0.00591	-0.00544
			(0.00382)	(0.00376)
Cov. FTTX (2016)			-0.0382***	-0.0384***
			(0.00634)	(0.00642)
Cov. HFC (2016)			-0.0410***	-0.0437***
			(0.00491)	(0.00515)
Cov. 6-16 Mbps inc. (2016)			0.0127	0.0148
			(0.0197)	(0.0199)
Cov. 16-50 Mbps inc. (2016)			0.0505***	0.0499***
			(0.0103)	(0.0102)
Cov. ≥50 Mbps inc. (2016)			0.0944***	0.0936***
			(0.0103)	(0.0102)
Share Pop. 18-34 (2014)				-0.0453
				(0.136)
Share Pop. 35-64 (2014)				0.0725
				(0.136)
Share Pop. ≥65 (2014)				0.103
				(0.0865)
Share Female (2014)				0.0756
				(0.150)
County FE	Yes	Yes	Yes	Yes
LEs	5142	5142	5142	5142

Note: Dependent variable is the percentage change in incumbent fixed-line market penetration between 2014-2016. Min. Hybrid M is the share of households that are covered by the hybrid product. All regressions are weighted with the number of households in 2014. Standard errors clustered at county level. Significance levels: $^{*}p < 0.1,^{**} 0.05,^{***} p < 0.01$.

Bibliography

Acemoglu, D. and J. Linn (2004). Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry. *Quarterly Journal of Economics 119*(3), 1049–1090.

Acemoglu, D. and P. Restrepo. Robots and Jobs: Evidence from US Labor Markets. *NBER Working Paper 23285*.

Akerman, A., I. Gaarder, and M. Mogstad (2015). The Skill Complementarity of Broadband Internet. *Quarterly Journal of Economics 130*(4), 1781–1824.

Atasoy, H. (2013). The Effect of Broadband Internet Expansion on Labor Market Outcomes. *Industrial and Labor Relations Review 66*(2), 315–345.

Audretsch, D. B., D. Heger, and T. Veith (2014). Infrastructure and Entrepreneurship. *Small Business Economics 44*(2), 219–230.

Autor, D. H., D. Dorn, and G. H. Hanson (2013). The China Syndrome: Local Labor Market Impacts of Import Competition in the United States. *American Economic Review 103*(6), 2121–2168.

Barth, A.-K. and U. Heimeshoff (2014a). Does the Growth of Mobile Markets Cause the Demise of Fixed Networks? Evidence from the European Union. *Telecommunications Policy 38*(11), 945–960.

Barth, A.-K. and U. Heimeshoff (2014b). What is the Magnitude of Fixed-Mobile Call Substitution? Empirical Evidence from 16 European Countries. *Telecommunications Policy 38*(8-9), 771–782.

Bartik, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* Kalamazoo: W.E. Upjohn Institute for Employment Research.

Basker, E. (2005). Job Creation or Destruction? Labor Market Effects of Wal-Mart Expansion. *The Review of Economics and Statistics 87*(1), 174–183.

Behrens, J. (2017). *Book on Demand: Auswirkungen auf den Deutschen Buchmarkt*. Norderstedt: BoD Books on Demand.

Bertschek, I., W. Briglauer, K. Hüschelrath, B. Kauf, and T. Niebel (2015). The Economic Impacts of Broadband Internet: A Survey. *Review of Network Economics 14*(4), 201–227.

Bertschek, I., D. Cerquera, and G. J. Klein (2013, sep). More Bits - More Bucks? Measuring the Impact of Broadband Internet on Firm Performance. *Information Economics and Policy 25*(3), 190–203.

Borraz, F., J. Dubra, D. Ferrés, and L. Zipitría (2014). Supermarket Entry and the Survival of Small Stores. *Review of Industrial Organization 44*(1), 73–93.

Börsenverein des Deutschen Buchhandels (2015). Buchkäufer und -leser 2015: Profile, Motive und Einstellungen. Accessed: 06/01/2018, https://docplayer.org/22304330-Buchkaeufer-und-leser-2015-profile-motive-einstellungen-frankfurt-am-main-14-oktober-2015.html.

Börsenverein des Deutschen Buchhandels (2018). Umsatzanteil von E-Books im Publikumsmarkt in Deutschland in den Jahren 2010 bis 2018. In *Statista - Das Statistik-Portal*. Accessed: 06/01/2018, https://de.statista.com/statistik/daten/studie/303339/umfrage/umsatzanteil-von-e-books-im-buchmarkt/.

Bourreau, M. and P. Dogan (2006). "Build-or-Buy" Strategies in the Local Loop. *American Economic Review 96*(2), 72–76.

Briglauer, W. (2015). How EU Sector-Specific Regulations and Competition Affect Migration from Old to New Communications Infrastructure: Recent Evidence from EU27 Member States. *Journal of Regulatory Economics 48*(2), 194–217.

Briglauer, W., C. Cambini, T. Fetzer, and K. Hüschelrath (2017). The European Electronic Communications Code: A Critical Appraisal with a Focus on Incentivizing Investment in Next Generation Broadband Networks. *Telecommunications Policy 41*(10), 948–961.

Briglauer, W. and K. Gugler (2017). Go for Gigabit? First Evidence on Economic Benefits of (Ultra-)Fast Broadband Technologies in Europe. *Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3006513*.

Briglauer, W., A. Schwarz, and C. Zulehner (2011). Is Fixed-Mobile Substitution Strong Enough to De-Regulate Fixed Voice Telephony? Evidence from the Austrian Markets. *Journal of Regulatory Economics 39*(1), 50–67.

Briglauer, W. and I. Vogelsang (2017). A Regulatory Roadmap to Incentivize Investment in New High-Speed Broadband Networks. *DigiWorld Economic Journal 106*, 143–160.

Brynjolfsson, E., Y. Hu, and M. S. Rahman (2009). Battle of the Retail Channels: How Product Selection and Geography Drive Cross-Channel Competition. *Management Science 55*(11), 1755–1765.

Brynjolfsson, E. and M. D. Smith (2000). Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science 46*(4), 563–585.

Bundesnetzagentur (2006). *Jahresbericht 2006*. Bonn.

Bundesnetzagentur (2015). *Marktdefinition und Marktanalyse des Marktes Nr. 3b der Märkteempfehlung der EU-Kommission vom 09. Oktober 2014*. Bonn.

Bundesnetzagentur (2017). *Jahresbericht 2017*. Bonn.

Bundesrechnungshof Deutschland (2009). *Die Postreform in Deutschland: Eine Rückschau*. Stuttgart: W. Kohlhammer GmbH.

Bundesverband E-Commerce und Versandhandel and Boniversum (2017). Würden Sie die folgenden Produkte lieber im Online- und Versandhandel oder lieber im klassischen Einzelhandel kaufen? In *Statista - Das Statistik-Portal*. Accessed: 10/09/2017, https://de.statista.com/statistik/daten/studie/201914/umfrage/einkaufsverhalten-im-onlinehandel-vs-einzelhandel-nach-produktgruppen/.

Bundesverband E-Commerce und Versandhandel Deutschland (2017). Warengruppen im Online-Handel nach dem Umsatz in Deutschland in den Jahren 2014 bis 2016 (in Millionen Euro). In *Statista - Das Statistik-Portal.* Accessed: 10/09/2017, https://de.statista.com/statistik/daten/studie/253188/umfrage/umsatzstarke-warengruppen-im-online-handel-in-deutschland/.

Canzian, G., S. Poy, and S. Schüller (2015). Broadband Diffusion and Firm Performance in Rural Areas: Quasi-Experimental Evidence. *IZA Working Paper 9429*.

Cavallo, A. (2017). Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers. *American Economic Review 107*(1), 283–303.

Colombo, M. G., A. Croce, and L. Grilli (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.

Cronau, S. (2006). *Buch und Buchhandel in Zahlen 2006*. Frankfurt am Main: Börsenverein des Deutschen Buchhandels e.V.

Cronau, S. (2011). *Buch und Buchhandel in Zahlen 2011*. Frankfurt am Main: Börsenverein des Deutschen Buchhandels e.V.

Cronau, S. (2013). *Buch und Buchhandel in Zahlen 2013*. Frankfurt am Main: Börsenverein des Deutschen Buchhandels e.V.

Cronau, S. (2014). *Buch und Buchhandel in Zahlen 2014*. Frankfurt am Main: Börsenverein des Deutschen Buchhandels e.V.

Czernich, N., O. Falck, T. Kretschmer, and L. Woessmann (2011). Broadband Infrastructure and Economic Growth. *Economic Journal 121*(552), 505–532.

Dauth, W., S. Findeisen, J. Südekrum, and N. Wößner (2017). German Robots - The Impact of Industrial Robots on Workers. *IAB-Discussion Paper 30/2017*.

De Stefano, T., R. Kneller, and J. Timmis (2014). The (Fuzzy) Digital Divide : Universal Broadband Access and Firm Performance. *University of Nottingham Discussion Paper 14/06*.

Economides, N., K. Seim, and V. B. Viard (2008). Quantifying the Benefits of Entry into Local Phone Service. *RAND Journal of Economics 39*(3), 699–730.

Einav, L., D. Knoepfle, J. Levin, and N. Sundaresan (2014). Sales Taxes and Internet Commerce. *American Economic Review 104*(1), 1–26.

Ellershaw, J., J. Riding, A. Lee, A. V. Tran, L. J. Guan, R. Tucker, T. Smith, and E. Stumpf (2009). Deployment Costs of Rural Broadband Technologies. *Telecommunications Journal of Australia 59*(2), 29.1–29.16.

Ellison, G. and S. F. Ellison (2009). Tax Sensitivity and Home State Preferences in Internet Purchasing. *American Economic Journal: Economic Policy 1*(12), 53–71.

European Commission (2018). Broadband Coverage in Europe 2017. Technical report.

Fabritz, N. (2013). The Impact of Broadband on Economic Activity in Rural Areas: Evidence from German Municipalities. *Ifo Working Paper 166*.

Falck, O., R. Gold, and S. Heblich (2014). E-lections: Voting Behavior and the Internet. *American Economic Review 104*(7), 2238–2265.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2018). Notes on Bartik Instruments. *NBER Working Paper No. 24408*.

Goolsbee, A. (2000). In a World without Borders. The Impact of Taxes on Internet Commerce. *The Quarterly Journal of Economics 115*(2), 561–576.

Goolsbee, A. (2001). Competition in the Computer Industry: Online versus Retail. *The Journal of Industrial Economics 49*(4), 487–499.

Goolsbee, A., M. F. Lovenheim, and J. Slemrod (2010). Playing With Fire : Cigarettes , Taxes , and Competition from the Internet. *American Economic Journal: Economic Policy 2*(1), 131–154.

Graetz, G. and G. Michaels (2018). Robots at Work. *Review of Economics and Statistics 100*(5), 753–768.

Grajek, M. and L.-H. Röller (2012). Regulation and Investment in Network Industries: Evidence from European Telecoms. *Journal of Law and Economics 55*, 189–216.

Gruhl, A., A. Schmucker, and S. Seth (2012). Establishment History Panel 1975-2010 (v. 2.2.1.). *FDZ-Datenreport*.

Grzybowski, L. (2014). Fixed-to-Mobile Substitution in the European Union. *Telecommunications Policy 38*(7), 601–612.

Grzybowski, L., M. Hasbi, and J. Liang (2018). Transition from Copper to Fiber Broadband: The Role of Connection Speed and Switching Costs. *Information Economics and Policy 42*, 1–10.

Grzybowski, L. and F. Verboven (2016). Substitution between Fixed-Line and Mobile Access: The Role of Complementarities. *Journal of Regulatory Economics 49*(2), 113–151.

Haller, S. A. and S. Lyons (2015). Broadband Adoption and Firm Productivity: Evidence from Irish Manufacturing Firms. *Telecommunications Policy 39*(1), 1–13.

Hethey, T. and J. F. Schmieder (2010). Using Worker Flows in the Analysis of Establishment Turnover - Evidence from German Administrative Data. *FDZ-Methodenreport 06/2010*, 1–42.

Igami, M. (2011). Does Big Drive Out Small? Entry, Exit, and Differentiation in the Supermarket Industry. *Review of Industrial Organization 38*(1), 1–21.

Imbens, G. W. and J. D. Angrist (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica 62*(2), 467–475.

Jia, P. (2008). What Happens When Wal-Mart Comes to Town? An Empirical Analysis of the Discount Retailing Industry. *Econometrica 76*(6), 1263–1316.

Kandilov, I. T. and M. Renkow (2010). Infrastructure Investment and Rural Economic Development: An Evaluation of USDA's Broadband Loan Program. *Growth and Change 41*(2), 165–191.

Kim, Y. and P. Orazem (2017). Broadband Internet and New Firm Location Decisions in Rural Areas. *American Journal of Agricultural Economics 99*(1), 285–302.

Klumpp, T. and X. Su (2010). Open Access and Dynamic Efficiency. *American Economic Journal: Microeconomics 2*(2), 64–96.

Kolko, J. (2012). Broadband and Local Growth. *Journal of Urban Economics 71*(1), 100–113.

Lange, M. R. J. and A. Saric (2016). Substitution Between Fixed, Mobile, and Voice over IP Telephony – Evidence from the European Union. *Telecommunications Policy 40*(10-11), 1007–1019.

Langendorf, B. (2003). *Buch und Buchhandel in Zahlen 2003.* Frankfurt am Main: Börsenverein des Deutschen Buchhandels e.V.

Lengsfeld, J. H. (2011). An Econometric Analysis of the Sociodemographic Topology of the Digital Divide in Europe. *The Information Society 27*(3), 141–157.

Liu, Y. H., J. Prince, and S. Wallsten (2018). Distinguishing Bandwidth and Latency in Households' Willingness-to-Pay for Broadband Internet Speed. *Information Economics and Policy 45*, 1–15.

Michailidis, A., M. Partalidou, S. A. Nastis, A. Papadaki-Klavdianou, and C. Charatsari (2011). Who Goes Online? Evidence of Internet Use Patterns from Rural Greece. *Telecommunications Policy 35*(4), 333–343.

Mills, B. F. and B. E. Whitacre (2003). Understanding the Non- Metropolitan-Metropolitan Digital Divide. *Growth and Change 34*(2), 219–243.

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

Neumark, D., J. Zhang, and S. Ciccarella (2008). The Effects of Wal-Mart on Local Labor Markets. *Journal of Urban Economics 63*(2), 405–430.

OpenStreetMap contributors (2018a). Austria. Retrieved on 2018/01/20 from `https://download.geofabrik.de/europe/austria.html`.

OpenStreetMap contributors (2018b). Germany. Retrieved on 2018/02/12 from `https://download.geofabrik.de/europe/germany.html`.

Oster, E. (2018). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business and Economic Statistics, forthcoming*.

Salemink, K., D. Strijker, and G. Bosworth (2017). Rural Development in the Digital Age: A Systematic Literature Review on Unequal ICT Availability, Adoption, and Use in Rural Areas. *Journal of Rural Studies 54*, 360–371.

Schmucker, A., S. Seth, J. Ludsteck, J. Eberle, and A. Ganzer (2016). Establishment History Panel 1975-2014. *FDZ-Datenreport*.

Schrape, J.-F. (2011). Der Wandel des Buchhandels durch Digitalisierung und Internet. *SOI Discussion Paper 2011-01*.

Townsend, L., A. Sathiaseelan, G. Fairhurst, and C. Wallace (2013). Enhanced Broadband Access as a Solution to the Social and Economic Problems of the Rural Digital Divide. *Local Economy 28*(6), 580–595.

van Dijk, J. and K. Hacker (2003). The Digital Divide as a Complex and Dynamic Phenomenon. *The Information Society 19*, 315–326.

Vogelsang, I. (2003). Price Regulation of Access to Telecommunications Networks. *Journal of Economic Literature 41*(3), 830–862.

Vogelsang, I. (2010). The Relationship between Mobile and Fixed-Line Communications: A Survey. *Information Economics and Policy 22*(1), 4–17.

Vogelsang, I. (2017). Regulatory Inertia versus ICT Dynamics: The Case of Product Innovations. *Telecommunications Policy 41*(10), 978–990.

Ward, M. R. and S. Zheng (2012). Mobile and Fixed Substitution for Telephone Service in China. *Telecommunications Policy 36*(4), 301–310.

Whitacre, B., R. Gallardo, and S. Strover (2014). Does Rural Broadband Impact Jobs and Income? Evidence from Spatial and First-Differenced Regressions. *The Annals of Regional Science 53*(3), 649–670.

Xiao, M. and P. F. Orazem (2011). Does the Fourth Entrant Make any Difference? Entry and Competition in the Early U.S. Broadband Market. *International Journal of Industrial Organization 29*(5), 547–561.

Zentner, A., M. Smith, and C. Kaya (2013). How Video Rental Patterns Change as Consumers Move Online. *Management Science 59*(11), 2622–2634.

Zhang, C. and X. Lei (2015). Short-Term Jobs and Long-Term Job Loss of Large Supermarket Entry. *China Economic Journal 8*(1), 74–85.