Incumbency Advantages: Price Dispersion, Price Discrimination and Consumer Search at Online Platforms *

Klaus Gugler† Sven Heim‡ Maarten Janssen§ Mario Liebensteiner¶

November 14, 2021

Abstract

We study how consumer search affects pricing when consumers observe the baseline price of their current provider and decide whether or not to search for an alternative tariff at an online platform. In such an environment, incumbent providers can price discriminate between consumers with lower and higher search costs. Using panel data on German electricity retail markets, we show that local incumbents increase their baseline and decrease their online tariffs, while entrants decrease their tariffs when consumer search intensifies. For incumbents, intensifying price discrimination is a viable option in markets with more consumer search. In a theoretical model, we show that these pricing patterns are consistent with the strategic interaction of an incumbent and entrants.

Keywords: Search, Price Dispersion, Price Discrimination, Electricity

JEL Classification: D43, D83, L11, L13, Q40

*We thank *ene't* for giving us access to their data on consumer search at price comparison websites. We thank Klenio Barbosa, Anette Boom, Tomaso Duso, Daniel Garcia, Alessandro Gavazza, Ulrich Laitenberger, Dieter Pennersdorfer, Steven Puller, Karl Schlag, Philipp Schmidt-Dengler, Nils-Hendrik von der Fehr, Mike Waterson, Matthijs Wildenbeest, Biliana Yontcheva, and Christine Zulehner for valuable comments. We also thank *Acxiom* for giving us data on household characteristics. The project was financially supported by the State Government of Baden-Württemberg, Germany, through the research program Strengthening Efficiency and Competitiveness in the European Knowledge Economies (SEEK). Janssen acknowledges financial support from the Austrian Science Foundation FWF, under project number P 27995-G27.

†Vienna University of Economics and Business (WU), klaus.gugler@wu.ac.at

‡Mines ParisTech, Paris, and ZEW Center for European Economic Research Mannheim, sven.heim@mines-paristech.fr

§University of Vienna, National Research University Higher School of Economics Moscow and CEPR, maarten.janssen@univie.ac.at

¶Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), mario.liebensteiner@fau.de

1 Introduction

Many markets are characterized by a substantial asymmetry between an incumbent provider and competing firms in that consumers know the contract with their current provider, but have to pay a search cost to be informed of alternative contracts. Once they know the alternative available contracts, consumers also have to be willing to switch to change providers. This is the case, for instance, in markets such as electricity or gas, where liberalizations have taken place but the former incumbent still serves a large fraction of the consumers. The incumbent may use this informational advantage and price discriminate between consumers with high search costs and those with lower search costs who may consider leaving.

This paper studies the asymmetric strategic interaction between an incumbent and alternative providers, taking into account consumers' willingness to search for and switch to alternative providers, and the possibility for the incumbent to price discriminate. In our empirical analysis, we investigate this interaction using data from the German retail electricity markets, which are characterized by local incumbents from the pre-liberalization era and many local retailers that have entered the market since.[1] In particular, we investigate how the extent of price discrimination and price dispersion depends on the fraction of consumers searching for alternative providers. Our theoretical model shows how price discrimination may help the incumbent increase its profit, given the heterogeneity of its customers.

The model has the following elements. Consumers observe the baseline price the incumbent sets to all customers at no cost. Having observed the baseline price, consumers decide whether or not to search. Search is costly and allows the consumers to observe all other prices in the market by consulting a platform/price comparison website. At the platform, consumers choose between buying from the lowest-price entrant or staying with the incumbent at the incumbent's discount platform price. Consumer-specific brand loyalty implies that even if the incumbent's discount price is not the lowest price on the platform, some consumers will stay with the incumbent.[2] The incumbent engages in second-degree price discrimination in that all consumers face a menu of prices and self select into whether or not they search. However, there is also an element of third-degree price discrimination in that once a consumer visits the platform, the incumbent infers it is a low-search-cost consumer so that it can target that

[1] This setup applies to electricity markets around the world. See, e.g., Cabral (2017) for evidence related to different European countries and Hortacsu et al. (2017) for evidence on the USA.

[2] We call these costs also transaction costs, i.e. all the objective and psychological costs consumers face if they switch.

group with a different price. We show that by varying the search cost distribution, this simple model can accommodate a rich pattern of pricing behavior, including the ones we find in our empirical analysis. If the fraction of consumers with low search costs increases, price dispersion and price discrimination increase simultaneously with the fraction of consumers who search online, while the incumbent raises its baseline price to consumers who do not search.

The empirical part of our analysis employs a unique data set on retail electricity prices and consumer search intensity at the German zip-code level for the period 2011–2014. The German retail electricity market has been liberalized at the end of the previous millennium, where former local monopolies have been replaced by local retail competition. Since then, local incumbent suppliers compete with new entrants, and all consumers that initially had to consume electricity at the incumbent's baseline tariff have the freedom to search for cheaper offers and switch to any alternative tariff that is available in their zip code. Even though in recent years most consumers use online platforms to search for cheaper rates,[3] in 2015 76% of all households were still served by the incumbent, with 33% remaining at the expensive baseline tariff, while 43% have switched to a cheaper incumbent tariff, and only 24% have switched to an entrant (BNetzA, 2015). Hence, some two decades after liberalization the incumbent still prices well above costs, strategically price discriminating between different types of consumer groups.

A key aspect of our dataset is the fact that we have information on consumer search intensity in panel format at the zip code and year level. In particular, we have data on the actual number of households' search queries at online price comparison platforms, and given that most of the search for lower prices is via these platforms, we interpret this data as a direct measure of search intensity at the local level. With some notable, recent exceptions (such as De los Santos et al., 2012; Blake et al., 2016; Coey et al., 2020), other empirical studies on consumer search markets often rely on indirect measures of consumer search activity.[4]

In terms of prices, we have access to the incumbent's baseline and cheapest online price and the lowest online price of an entrant at the zip code level. Using these data, we empirically show that incumbents increase their baseline rates if consumer search more. We also

[3]According to a 2011 survey, 80% of the switchers searched online for alternative providers (www.atkearney.at/documents/3709812/3710656/BIP_Der_Strom_und_Gasvertrieb_im_Wandel.pdf, March 3, 2016). This number is likely to have increased in more recent years.

[4]Brynjolfsson and Smith (2014), for example, use access to the internet as a proxy for lower search costs. Similarly, for retail gasoline markets, Pennersdorfer et al. (2020) use commuters versus non-commuters to distinguish between informed and uninformed consumers.

find that entrants reduce their tariffs. Moreover, the incumbent increases the extent of price discrimination and lowers its cheapest online tariff significantly when consumers search more at platforms. We estimate that a one-standard deviation increase in within-zip-code search intensity explains nearly 40% of observed price discrimination. Hence, one key take-away message of our analysis is that an incumbent who is confronted with competitors entering the market and who has a heterogeneous customer base of loyal and searching consumers can increase its profit using price discrimination. As search intensity and pricing are likely to affect each other simultaneously (as search intensity is a function of prices (e.g. Byrne and De Roos, 2017; Heim, 2021; Tappata, 2009) and retailers' pricing strategies depend on consumers' search efforts), we address the potential reverse causality of pricing on search intensity by employing an instrumental variables approach.

The German electricity market is just one example of a market where important market liberalizations have taken place over the last 20–30 years. Many of these markets (including electricity markets in several states of the USA, Canada, other EU Member States) share important features with the German electricity market. In all these markets, new firms have entered, incumbents may engage in price discrimination, and there is an important asymmetry as consumers know the base price of the incumbent, whereas they have to incur a search cost to learn prices set by entrants. Other liberalized sectors, such as natural gas, telecommunications, health insurance, railways, postal services, and airlines share similar features. A key dividing line between these examples is whether or not consumers have an ongoing relation with their suppliers. Thus, markets, such as electricity, telecommunications, and health insurance markets, have the feature that consumers are naturally informed about their current supplier and will automatically continue their contract as long as they do not search for and switch to alternatives. The role of incumbency effects is also of importance to sectors beyond the liberalization context, such as retail banking, where (online) searching consumers may get much better deals than loyal consumers.[5]

As the focus of this paper is on the impact of consumer search on price dispersion and price discrimination in markets with incumbents, it contributes to different strands of literature.

[5]For example, the article "American banks pay depositors less than online accounts", in The Economist (15 February 2018, https://www.economist.com/news/finance-and-economics/21737031-they-seem-be-relying-power-inertia-retain-their-customersa-risky) states that established U.S. banks generally offer substantially lower interest rates on savings accounts compared to online rates offered to clients at internet portals. Houde et al. (2019) show that banks have an incumbency advantage for mortgage services because the large majority of consumers combines day-to-day banking and mortgage services opening the possibility to price discriminate between consumers with different outside options and/or search costs.

There is a large and varied theoretical literature on how consumer search affects price dispersion in homogeneous goods markets (see, e.g., Stahl, 1989, and Janssen and Moraga-González, 2004). Most of the theoretical models are based on random search, and do not take into account the incumbency effect or the possibility of price discrimination.

Several empirical studies focus on price dispersion and search intensity (see, e.g., Sorensen, 2000; De los Santos et al., 2012). Tang et al. (2010) find that an increase in shopbot use reduces average prices and price dispersion in online book retailing. Lach and Moraga-González (2017) show that competition may be more beneficial for consumers who are better informed. Pennersdorfer et al. (2020) find an inverted U-shaped relation between price dispersion and the share of informed consumers (as proxied by the share of commuters) in the Austrian gasoline retail market. This literature does not, however, deal with incumbency effects.

A growing literature explicitly deals with search in electricity markets, but most of these papers mainly focus on how consumers search without considering the implications for price setting. Giulietti et al. (2014) analyze the retail electricity market in the United Kingdom and find that roughly half the households had relatively high search costs. Hortacsu et al. (2017) analyze switching in the Texas retail electricity market and find that even though households rarely switch to alternative retailers, they do switch more after experiencing a "bill shock". Moreover, they also find that households attach a brand advantage to the incumbent. Both papers do not observe the actual search behavior of consumers, however. Dressler and Weiergraeber (2019) use a structural demand model of the Belgian electricity market focussing on switching costs and limited awareness. In contrast, Byrne et al. (2019) use a field experiment to study how heterogeneous search frictions are used by electricity firms in Australia to differentiate between consumers by combining posted prices and sequential bargaining with individual households.

Another related literature argues that entry may lead to higher incumbency prices and/or profits (see, e.g., Perloff et al., 1995; Ishibashi and Matsushima, 2009). In all these models, because of either horizontal or vertical product differentiation, after entry the incumbent will focus on a more targeted group of consumers that are less price sensitive. Using a similar logic, Doganoglu (2010) shows that small switching costs may lead to lower prices relative to a situation without switching costs. Even though the mechanism of our theoretical model also relies on the incumbent targeting a specific group of consumers, our focus is different as we take entry as given and analyze the incumbent's price discrimination strategy and how it depends on search and switching behavior.

There is a small literature dealing with price discrimination and incumbency. For the UK retail electricity market, Davies et al. (2014) present evidence suggesting that firms deliberately differentiated their tariff structures, resulting in market segmentation according to consumers' usage. For the US airline industry, Goolsbee and Syverson (2008) indicate that incumbents respond to the threat of entry by substantially reducing average fares on the directly threatened routes, but that they do not cut prices on routes to nearby airports in the same market. This bears some relationship to our result that the incumbent price discriminates between searching consumers who may choose an alternative option and non-searching consumers who do not.

At a theoretical level, the idea that a firm would like to price discriminate against consumers with higher search cost is not new (see, e.g., Salop, 1977). Salop (1977) studies a monopoly setting and his argument critically depends on the assumption that the monopolist is committed to charging prices according to a price distribution, while consumers can somehow react to changes in the price distribution (assuming they observe the distribution, but not the prices) by adopting a different search strategy. Cabral (2016) analyzes conditions for which switching costs may lead to higher or lower equilibrium prices in markets where sellers discriminate between locked-in and not locked-in consumers. Cabral and Gilbukh (2020) also model firms engaging in price discrimination between active and passive searchers. Unless they pay a search cost, consumers buy from the high price of a firm. The focus of Cabral and Gilbukh (2020) is, however, very different from ours in that they study symmetric firms facing cost shocks, where we focus on how asymmetric pricing is affected by the presence of more searching consumers.

The rest of the paper is structured as follows. The next section describes the German retail electricity market in more detail. Section 3 provides a theoretical model to guide the empirical approach and findings. Section 4 describes the empirical identification strategy and discusses the data. Section 5 presents the econometric results. Section 6 concludes.

2 Institutional Details

In 1999 Germany's electricity liberalization brought about the end of local monopolies by allowing entry to local markets. It was believed that increased competition and freedom of consumer choice would eventually result in lower retail margins with large economic benefits for consumers. Prior to market liberalization, the local incumbent served all customers in its distribution grid area at a regulated tariff. Since liberalization, the incumbents have been legally obliged to supply electricity at a baseline tariff to all households which do not proactively choose another supplier. Moreover, a household that moves to another zip code is automatically supplied by the local incumbent at its baseline tariff.[6] The incumbents' baseline tariffs are no longer regulated and households are free to change to alternative tariffs that are offered by one of the many new entrants or by their local incumbent.

Retail entry in the German electricity market follows a regulated, non-discriminatory procedure and involves low entry costs and risks. This is also witnessed by the large number of active retailers: there are on average 133 electricity retailers per zip code area with a range of 55 to 192. Another important characteristic is that retailers competing in a zip code have almost identical costs: some cost components, such as grid charges and concession fees, differ over time and across zip codes, but are equal for all retailers in a zip code. Other cost components, such as subsidies for renewable energies, only change over time but do not have local variation. Costs for purchasing wholesale electricity are also almost identical across retailers since wholesale electricity prices are determined centrally at the European Energy Exchange (EEX).[7] Some other costs, such as administrative or advertisement costs, may differ across retailers but account for only a minor part of (variations in) the retail costs. Thus, while costs are similar for all retailers in the local market, they vary substantially across local markets in Germany.

Many incumbents operate only at a very local level and 46% of the incumbents only have a single zip code in their incumbency area. These small incumbents are mostly municipal utilities. Larger incumbents often have several zip codes in their incumbency area. The incumbency areas of incumbents with more than one zip code cover five zip codes at the median and 32 at the mean. Hence, as the costs differ between zip codes, incumbents serving more than one zip code area face different costs within their incumbency area, and on average they set 3.5 different prices in their incumbency areas. Incumbents that have more than one price zone set prices that differ on average by 10.4 Euro/3.5 MWh.[8,9] Thus, retailers set local prices that vary (in most cases) at the zip-code level.

At the start of the liberalization process, all consumers were automatically supplied by their

[6]By law, the incumbent in a zip code area is defined as the local retailer with the largest customer base. Thus, even though in theory a different retailer may become incumbent, in practice the original incumbent almost never changed. The few exceptions are due to mergers between municipal utilities.

[7]Even if firms buy electricity through direct contracts with electricity producers, the spot price still represents the opportunity costs of purchasing electricity.

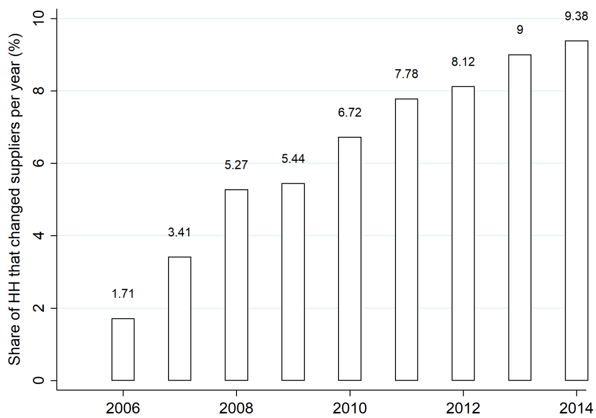
[8]The largest observed difference of the base tariffs within the same incumbency area of 134 Euro/MWh was in 2012 by E.ON Avacon Vertrieb GmbH, which served 189 zip codes with 14 different price zones.

[9]Online Appendix Figure F1 illustrates the base tariffs set by Envia Mitteldeutsche Energie GmbH within its incumbency area.

local incumbent and paid the baseline tariff. Consumers can switch away from the *incumbent baseline* tariff at any time with two weeks' notice. Consumers who switch generally take a one-year contract with their new supplier, which is automatically renewed if the consumer does not cancel the contract in time.[10]

In recent years, most households who consider changing their supplier visit an online price comparison platform. The largest platforms are Verivox, Toptarif, and Check24. Besides Toptarif, our database also covers all search activities conducted on several other well-known online price comparison platforms including Stromtipp.de (power hint), Energieverbraucherportal.de (energy consumption portal) and mut-zum-wechseln.de (courage-to-switch). Verivox started to provide search services in electricity in 2000, Toptarif in 2007, and Check24 in 2008. Despite this fairly recent trend of searching through platforms, in 2011 80% of the switchers had already searched online for alternative providers (A.T. Kearney, 2012).

Figure 1: Average switching rates of households in German retail electricity markets



Notes: Data on supplier changes are obtained from Germany's regulatory authority (BNetzA, 2015), data on the number of German households are from the German Federal Statistical Office.

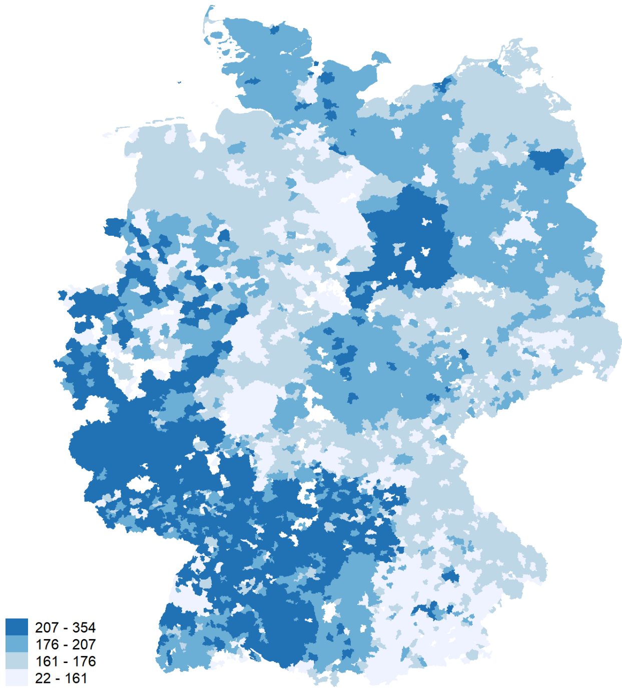
The switching rate has been growing in recent years (see Figure 1), as online price comparison platforms have significantly reduced the costs of searching for cheaper providers (something that is also acknowledged in other markets; see e.g. Bar-Isaac et al., 2012). A comparison portal requires a consumer to enter all relevant details (zip code, expected yearly electricity consumption, whether the contract is for private or commercial use). Then, there are several options to choose from, such as whether or not to only consider "green" electricity, whether or not prices are guaranteed throughout the year[11] and whether or not the listed tariffs should

[10] According to a market report by the German regulatory authority BNetzA (2013, p. 150), the average contract period is 10 months, suggesting that most consumers choose yearly contracts.

[11] Consumers have an extraordinary termination right if their retailer adjusts the price. Retailers may also adjust

include one-off bonuses. The platform then lists the "personalized" prices of all providers that are active in the indicated zip code ranked from lowest to highest. For each tariff, the platform also provides information on how much consumers can save over the year compared to the incumbent's baseline price. Thus, the search process costs some time and effort, but for all consumers who are familiar with online shopping, the search costs are relatively small compared to the potential savings of switching from the incumbents' baseline tariff to the overall cheapest tariff, which are on average almost 200 Euro per year for a standard two-person household with 3,500 kWh consumption. The potential gains from search range from 77 to 354 Euro per year, depending on the households' location, as shown in Figure 2 for the year 2012.

Figure 2: Potential gains from search (2012)



There is a tiered pricing system in Germany (two-part tariffs with a fixed and a variable component). However, as electricity consumption is inelastic it is typically not the case that consumers start consuming more if prices decline. Consumers have a typical amount of electricity consumption, which depends on how they heat, whether they use air conditioning, how much time they watch television, etc. Thus, consumers simply sort into different categories (types) and in the empirical analysis we consider households with the most common level of electricity consumption (i.e. 3,500 kWh, BNetzA, 2015; this is also the default consumption

tariffs despite having them advertised as "guaranteed", e.g. in the event of changing grid fees or a change in the renewable energy surcharge.

level suggested by price comparison platforms, see Figure 5) and calculate the respective annual tariffs.

Not only have search costs declined over time, switching costs have also been significantly reduced as switching is now an automated process and conducted entirely by the new provider who automatically arranges all switching activities for new customers, such as unsubscribing from the old supplier and registration, at no additional cost.[12]

Finally, as there are no retailer specific differences regarding the quality of supply, retail electricity can be considered a fairly homogeneous product, which helps us to rule out product differentiation as a possible explanation for price dispersion.[13] If an entrant fails to deliver, the incumbent provider has the legal obligation to continue electricity supply at the baseline tariff without interruption. However, not all consumers may be aware of this safety net. Hence, theoretically it should not matter for the end-consumer which retailer delivers the electricity, although to some extent it still may matter in practice, equipping the incumbent with some brand value.

As other prices than the incumbent base tariff can only be observed by consumers who proactively search, an incumbent is able to set alternative tariffs online and price discriminate. The incumbent's online tariff is typically set below the baseline tariff and above the cheapest overall tariff set by an entrant. Figure 3 shows that there are considerable price differences between the incumbent's high baseline tariff P_H^I, the incumbent's lowest tariff P_L^I), and the overall cheapest entrant tariff P^E. As consumers who switch away from the incumbent most likely choose the cheapest tariff available, we focus on the cheapest entrant price.[14] As a result, we observe three forms of price dispersion: (i) Overall *price dispersion* ($P_H^I - P^E$), which is the difference between the incumbent's baseline tariff and the overall cheapest tariff; (ii) *price discrimination* by the incumbent ($P_H^I - P_L^I$), measured by the difference between the incumbent's baseline tariff and the incumbent's cheaper online tariff; and (iii) *online price dispersion* measured by the difference between the incumbent's cheaper tariff and the cheapest entrant tariff ($P_L^I - P^E$).[15]

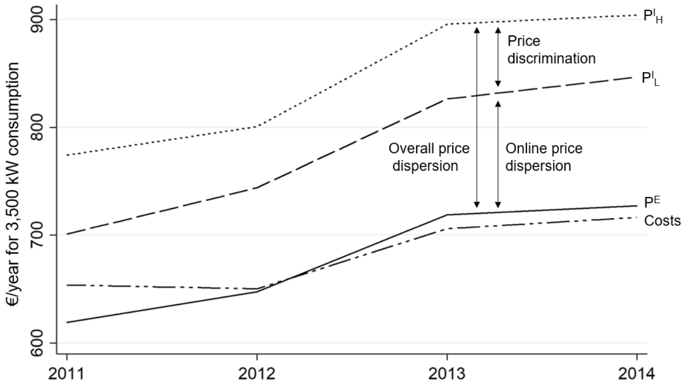
[12]In many other countries, the switching process for electricity providers is comparable to the one in Germany's retail electricity markets. E.g., studying the UK market, Giulietti et al. (2014, p. 561) argue that "search is perceived by consumers as being significantly more difficult than switching." A similar point has been made by Hortacsu et al. (2017) for Texas.

[13]The main product differentiating feature is the differentiation between regular tariffs and "green" tariffs, which exclusively sell electricity produced by renewable sources. As consumers choosing green certified tariffs only present a very small share of all searching consumers, in the empirical analysis we omit these consumers.

[14]This is supported by, for example, Baye et al. (2006) in the market for handheld PCs.

[15]We employ the price range as our dispersion measure, which is a commonly used measure in the literature

Figure 3: Average tariffs and costs (€/year for 3,500 kWh)



Note: P_H^I, P_L^I, P^E denote the incumbent's baseline tariff (price incumbent high), the incumbent's cheapest online tariff (price incumbent low), and the overall cheapest entrant tariff (price entrant), respectively. Costs (C) are presented net of value added taxes.

Figure 3 also depicts the (approximated) costs of retailers (see Section 4.2 for more details). We see that costs and prices have increased over time (mostly due to increased taxes and levies to finance the integration of renewables). Evidently, even nearly two decades after the retail liberalization in the industry, the incumbent baseline tariff remains well above costs. Moreover, the figure emphasizes that incumbents price discriminate with the cheaper incumbent's price still being well above costs. By contrast, the cheapest tariffs set by entrants are very close to cost.

3 A Simple Search Theoretic Model

The main purpose of the theoretical model is to explain that the observed patterns are consistent with the incentives of firms and consumers.[16] In particular, a profit-maximizing incumbent may well increase its baseline prices and increase price discrimination in regions where a larger fraction of consumers have low search and switching cost. To do so, we build the simplest possible model where an incumbent competes with entrants for a homogeneous product and the incumbent is able to price discriminate between consumers that are heterogeneous in their search and transaction costs.

The model closely follows the institutional details described above and we view the decision

(Baye et al., 2006). In our case, the price range best reflects the potential gains from search.

[16]In Section D of the Online Appendix, we also use our theoretical model for a counterfactual analysis of the welfare impacts of a prohibition of price discrimination. We find that *on average* consumers are worse off under price discrimination, whereas searching consumers are better off and loyal consumers are considerably worse off. We also find that with price discrimination, there will be fewer consumers switching to entrants than if price discrimination were banned.

problem of the consumer as a static problem that is repeated every year. All consumers observe the regular (baseline) price P_H^I of the incumbent at no additional cost.[17] There is an online price comparison website consumers can consult at a search cost s, which is distributed according to a distribution function $F(s;z)$, where we use z to represent exogenous parameters that determine the shape of the search cost distribution. (In the empirical part of the paper, z are the instruments that are exogenous to the prices, but that do affect differences in search behavior across different zip codes.) The search cost reflects the time it takes consumers to get familiar with the tariff-comparison platform and to enter the required personal information on the price comparison website. At the website, consumers will see the online price P^E of the overall cheapest firm (usually an entrant) and the cheapest (online) price P_L^I of the incumbent. Once on the website, a consumer compares prices without additional search cost.

Apart from their search cost, consumers also pay a transaction cost if they want to switch away from the incumbent. These costs also differ between individuals and refer to all the objective and psychological costs consumers face if they switch. As explained in Section 2, the objective switching costs are small, but consumers may perceive the incumbent as more trustworthy. To keep the analysis simple, we assume that these transaction costs are proportional to the search cost, *i.e.*, the transaction cost of a consumer with search cost s is denoted by θs.[18] Thus, once a consumer with search cost s is online and observes both prices P^E and P_L^I, he will continue to buy from the incumbent if $P_L^I - \theta s < P^E$.

The sequence of actions is as follows. In the first-stage, the incumbent and entrant choose prices P_H^I, P_L^I and P^E simultaneously.[19] At the beginning of the second-stage, consumers only observe P_H^I and decide whether or not to search based on their expectation regarding online prices. If they do not search, they buy at P_H^I from the incumbent. If they do search, they observe the online prices and buy where it is best for them, taking the transaction cost into account. We use perfect Bayesian equilibrium with passive beliefs as our solution concept. Thus, we

[17]Note that our model is static and treats incumbents and entrants as asymmetric. In real markets, the following dynamic aspects may be important: once some consumers have switched to entrants, they later gain some incumbency effect as these consumers will have to search if they want to switch away from their provider. Thus, over time entrants and incumbents may become more symmetric to each other. In the theoretical model, we have abstracted from these considerations as individual entrants in the German electricity market typically have a very small market share. In Table 1, we report that on average there are more than 130 firms active in every zip code, though the incumbent provider continues to have around 76% market share. This implies that on average entrants have less than 0.2% market share.

[18]For example, consumers with higher search costs may be older and more wealthy and they also have higher transaction cost as they do not want to risk their stable delivery of electricity by switching.

[19]In Section E of the Online Appendix, we consider an alternative "Stackelberg" version of the model where the incumbent first chooses its baseline price P_H^I, and P_L^I and P^E are chosen at the moment P_H^I is given and observed by the entrant. This model yields the same qualitative predictions.

look for an equilibrium where consumers hav correct beliefs about the online prices and where if a consumer observes an unexpected price P_H^I (different from the equilibrium level), he will continue to believe that P_L^I and P^E are at their equilibrium levels.

A natural candidate for an equilibrium is where low search cost consumers with $s < \widehat{s}_2$ search online and all other consumers stay with the baseline price of the incumbent. Moreover, of the consumers that search online, the ones with a transaction cost $\theta s < \theta\widehat{s}_1$, with $\widehat{s}_1 < \widehat{s}_2$, buy from the entrant, while other online consumers, namely those with $\widehat{s}_1 < s < \widehat{s}_2$, buy from the incumbent at its cheapest (online) price. In such an equilibrium, the cut-off values for search costs are $\widehat{s}_1 = \left(P_L^I - P^E\right)/\theta$ and $\widehat{s}_2 = \left(P_H^I - p_L^{I^e}\right)$. Note that in the definition of $\widehat{s}_2$ we have the incumbent's online price $P_L^{I^e}$ consumers *expect* to find *if* they search and not the realized price as when deciding whether or not to search, consumers do not know the online price. Note also that $\widehat{s}_1$ is defined in terms of realized prices as all consumers with an $s < \widehat{s}_2$ visit the platform and decide from whom to buy after observing both prices.

Assuming, without loss of generality, that the firms have no supply cost, the equilibrium prices we derive can be interpreted as firms' margins. Thus, the respective profits of the entrant and incumbent are as follows:

$$\pi_E = F(\widehat{s}_1; z)P^E = F\left(\frac{P_L^I - P^E}{\theta}; z\right)P^E$$

and

$$\begin{aligned}\pi_I &= \left[F(\widehat{s}_2; z) - F(\widehat{s}_1; z)\right]P_L^I + (1 - F(\widehat{s}_2; z))P_H^I \\ &= \left[F(P_H^I - P_L^{I^e}; z) - F\left(\frac{P_L^I - P^E}{\theta}; z\right)\right]P_L^I + (1 - F(P_H^I - P_L^{I^e}); z)P_H^I.\end{aligned}$$

This yields the following F.O.C.s (evaluated at the equilibrium where $P_L^{I^e} = P_L^I$) for the entrant and the incumbent, respectively:

$$F\left(\frac{P_L^I - P^E}{\theta}; z\right) - f\left(\frac{P_L^I - P^E}{\theta}; z\right)\frac{P^E}{\theta} = 0, \tag{1}$$

$$F(P_H^I - P_L^I; z) - F\left(\frac{P_L^I - P^E}{\theta}; z\right) - f\left(\frac{P_L^I - P^E}{\theta}; z\right)\frac{P_L^I}{\theta} = 0, \tag{2}$$

and

$$-f(P_H^I - P_L^I; z)(P_H^I - P_L^I) + (1 - F(P_H^I - P_L^I; z)) = 0, \tag{3}$$

where $f(.)$ is the density function that is associated with $F(.)$. Note that the fraction of actively searching consumers is given by $F(P_H^I - P_L^I; z)$.

For a given z, these three F.O.C.s determine the equilibrium values of $P_H^{I^*}$, $P_L^{I^*}$ and P^{E^*} and the corresponding levels of price discrimination and price dispersion. To explain our observations, we have to see how these equilibrium price levels change with variations in z and/or θ. The effects of the transaction cost parameter θ are clear-cut: all prices and online price dispersion $P_L^{I^*} - P^{E^*}$are increasing in θ, whereas overall price dispersion $P_H^{I^*} - P^{E^*}$is decreasing and price discrimination $P_H^{I^*} - P_L^{I^*}$ is unaffected. The effects of changes in the search cost distribution, reflected by changes in the parameter z, are richer and different patterns are possible.

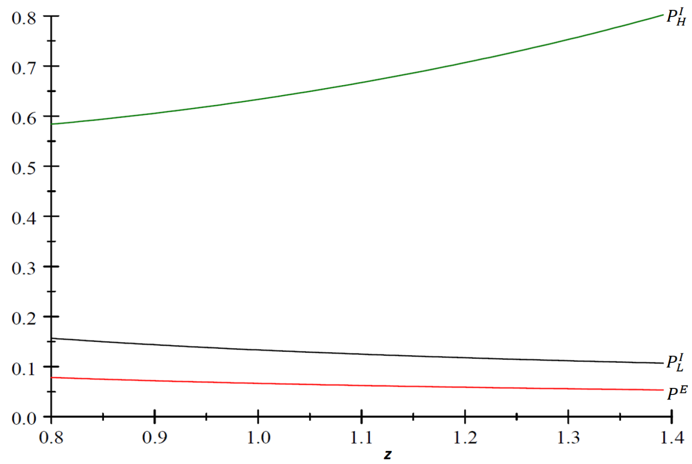
Proposition 1 *The effects of an increase in the transaction cost parameter θ is that all prices and online price dispersion increase, while overall price dispersion decreases and price discrimination is unaffected. The effects of changes in z are as follows. Price discrimination increases if, and only if, the inverse hazard rate evaluated at the equilibrium values* $\frac{1-F(P_H^{I^*}-P_L^{I^*};z)}{f(P_H^{I^*}-P_L^{I^*};z)}$ *is increasing in z. Moreover, P^E and online price dispersion are positively related to P_L^I if the density functions are non-increasing, i.e.* $\partial f\left(\frac{P_L^I-P^E}{\theta};z\right)/\partial(P_L^I - P^E) \leq 0$.

The economic intuition behind the result on price discrimination is as follows: for a given value of P_L^I the incumbent faces a trade-off in its decision whether or not to increase P_H^I. Raising P_H^I increases the profits for all consumers $1-F(P_H^{I^*}-P_L^{I^*};z)$ who stay with the baseline tariff, but a fraction proportional to the density $f(P_H^{I^*} - P_L^{I^*};z)$ will decide to search. At the margin, those that decide to search will eventually buy at the incumbent's online price $P_L^{I^*}$ as the marginal consumer has a higher search and switching cost. The incumbent will lose $P_H^{I^*} - P_L^{I^*}$ per (marginal) consumer who searches. If, evaluated at the equilibrium values, the inverse hazard rate is increasing in z,[20] relatively more consumers will stay on the baseline tariff if z increases, making price discrimination more profitable. Also, to understand online price dispersion, if P_L^I increases, then there is a larger potential demand for the entrant and, under "normal" demand conditions, it should increase its price, but not to the full extent (thereby also increasing sales).

To get more precise results how price discrimination and price dispersion are related to the fraction of searchers, one has to assume a specific form of the search cost distribution. In Section C of the Online Appendix, we analyze the case of a piece-wise linear search cost distribution:

[20]Most distributions covered in standard statistics textbooks have an inverse hazard rate $(1 - F(x))/f(x)$ that is decreasing in x. We ask, however, the inverse hazard rate to be increasing in an exogenous parameter z on the relevant part of the domain of possible search cost values.

Figure 4: Model prediction



Note: The figure predicts price changes as a function of z with $\widetilde{s}_2 = 3/5$ and $\widetilde{s}_1 = 1/5$ and $\theta = 2/5$. P_H^I, P_L^I, and P^E denote the incumbents' baseline tariffs, the incumbents' cheapest (online) tariffs, and the overall cheapest entrants' tariffs, respectively.

$$F(s) = \begin{cases} zs & \text{for } s < \widetilde{s}_1 \\ \alpha + \beta s & \text{for } \widetilde{s}_1 \leq s < \widetilde{s}_2 \\ s & \text{for } 1 \geq s \geq \widetilde{s}_2 \end{cases}$$

where, to have a proper piece-wise linear distribution function, $\alpha = \frac{(z-1)\widetilde{s}_1\widetilde{s}_2}{\widetilde{s}_2-\widetilde{s}_1}$, $\beta = \frac{\widetilde{s}_2-z\widetilde{s}_1}{\widetilde{s}_2-\widetilde{s}_1}$, $\widetilde{s}_2 > \widetilde{s}_1$, and $z > 0$. If $z = 1$, we have the uniform distribution.

Figure 4 depicts how the different prices change as a function of z when $\widetilde{s}_2 = 3/5$ and $\widetilde{s}_1 = 1/5$ and $\theta = 2/5$. Detailed derivations are given in the online Appendix. As $\partial F(P_H^{I^*} - P_L^{I^*})/\partial z$ is a constant positive number, this figure can also be interpreted as how prices are linked to the fraction of searchers. One can see that the incumbent's baseline price is increasing in the fraction of searchers, whereas the other two prices are decreasing, resulting in more price discrimination and overall price dispersion, while online price dispersion is decreasing. This is also what we find in our empirical analysis (see Section 5).

4 Research Design

To examine the effect of consumer search intensity on pricing strategies, we first explain our identification strategy and then describe our data and results.

4.1 Identification and Estimation

We consider the estimation equation

$$\ln(y_{it}) = \beta \ln(\mu_{it}) + \gamma \mathbf{x}_{it} + \delta_i + \eta_t + \epsilon_{it}, \tag{4}$$

where the dependent variable y either denotes an electricity tariff (P_H^I, P_L^I, P^E) or a price difference measure ($P_H^I - P^E$, $P_H^I - P_L^I$, $P_L^I - P^E$) and is a function of consumer search intensity (μ), a set of control variables ($\mathbf{x}$) including regional electricity costs, the number of regional electricity retailers, and some regional socio-economic characteristics, such as income, unemployment rate and average household size among others. Subscript i indicates zip codes with δ_i being the respective location fixed effects. Time t is measured by years, η_t is a set of year effects. ϵ is the error term, which may capture unobservable heterogeneity at the year-by-zip-code level, which is potentially related to search intensity – an issue that we will address by using instrumental variables.

Since we only observe consumer search at the online platforms represented in our sample, but not all consumer search activity, we estimate constant elasticities in a log-log relationship. That is, we include the dependent variables (i.e. tariffs and dispersion measures) as well as search intensity in logs indicating by how much a percentage change in search impacts the dependent variable in percentage terms. Assuming that search patterns at other comparison websites are not different from search at the platforms that we observe in our data, the elasticity estimate allows us to make inferences about the whole market.[21] Hence, our parameter of interest β reflects the percentage change in pricing behavior for a one percent change in search activity.[22]

OLS estimation of Equation (4) recovers the causal impact of search intensity on the outcome variable under the assumption that search intensity is strictly exogenous, conditional on controls ($\mathbf{x}$) as well as on time and location effects. Fixed effects effectively control for unobserved heterogeneity across regions and for aggregate shocks to search intensity. However, we may still face simultaneity between prices and search intensity as consumers search more when the expected gains from searching increase, i.e. when prices are high and/or more dispersed. This is what theory suggests (e.g. Tappata, 2009) and recent empirical work supports this prediction.

[21]This assumption is reasonable: We also have data on consumer search at the platform *Verivox* for the year 2014 and find a correlation coefficient of 85% between search intensity at Verivox and the platforms we use here. Verivox's data is only provided as percentages of search in a respective zip code relative to the overall search in Germany, which is why we cannot merge these data with our actual search data.

[22]However, the estimations are robust if estimated in levels, as we show later.

It has been shown that this is the case in gasoline markets (Byrne and De Roos, 2017) and very recently also for the market analyzed in this paper: Heim (2021) uses the same data and shows that search intensity is indeed substantially affected by changes in prices.

Ignoring this relationship will lead to a biased OLS coefficient on consumer search ($\ln(\mu)$) in Equation (4). To address this issue, we adopt an instrumental-variable (IV) approach. We make use of two potential instruments. One is the regional availability of broadband internet (*BBA*) measured as the share of households in a zip code for which an internet speed of 16 Mbit/s is offered for their local address. Another instrument is the share of young households in a zip code measured as the share of households with a head below the age of 40 (*U40*).

To be a valid instrumental variable, changes in broadband internet availablity and age structure must be (i) correlated with local trends in search intensity, and (ii) unrelated to unobserved local shocks that confound the impact of search intensity on the outcome variable, i.e. the instruments affect pricing strategies only through the search channel. Assumption (i) that the availability of broadband internet and the share of young households are relevant for consumer search intensity is plausible. Faster internet makes online shopping more convenient (i.e. decreases search costs) and younger people are probably more familiar with the internet in general and with online shopping in particular which again decreases search costs. Moreover, this is a testable assumption. The exclusion restriction (ii) is not testable and must be assessed on the grounds of plausibility. There are good reasons to assume that broadband internet availability is unrelated to retail electricity prices. For example, it is unlikely that an electricity provider may set its tariffs in a given zip code according to the availability of broadband internet. Nevertheless, a potential threat to instrument exogeneity may arise because the instruments are potentially correlated with other factors that also affect prices if one does not control for these factors. For instance, internet availabilty may be correlated with population density and urbanization, which in turn may be correlated with the error term. In other words, broadband internet and prices are correlated through other variables if they were omitted. Similarly, it is well possible that – on average – young households also have a poorer financial situation, which in turn may affect pricing decisions (e.g. retailers may face a higher risk of payment shortfalls by younger customers and they may consider this while setting their prices).

In order to overcome this issue we control for a battery of variables in our regression model. We include a set of demographic variables (such as urbanization, population density and household size) in the $\mathbf{x}$ vector. We further control for several variables that capture local financial conditions. We include different income classes – the share of households with an annual income of less than €25,000 (Income <25k €/a) and the share of households with an annual income between €25,000 and €50,000 (Income 25k-50k €/a), with incomes above €50,000 representing the reference group. We also include the average available purchasing power in a zip code as well as the shares of households in a zip code belonging to a certain social status (lower or middle class). We further control for the unemployment rate, the share of socially insured people, the number of households, the average household size, the degree of urbanization and the population density.

By this, we make sure that our instruments are uncorrelated with the error term conditional on these variables. Hence, they should satisfy the exclusion restriction by having an impact on treatment assignment but not on the outcome of interest, conditional on the control variables.

Using these two instruments, we estimate the following linear projection of $\ln(\mu)$ in the first-stage estimation:

$$\ln(\mu_{it}) = \alpha \mathbf{z}_{it} + \gamma^{FS} \mathbf{x}_{it} + \delta_i^{FS} + \eta_t^{FS} + u_{it}, \quad (5)$$

where the vector $\mathbf{z}$ consists of our two instruments *U40* and *BBA*. The superscript *FS* indicates that the parameters are from the first-stage estimation. Plugging the first-stage prediction of search (i.e. $\widehat{\ln(\mu)}$) into Equation (4) should yield a consistent estimate for β.

A remaining potential threat to identification may be that broadband internet availability affects electricity demand or that younger households consume more or less electricity than older households. However, we believe that these demand effects are not likely to translate into direct pricing effects in retail electricity markets, because electricity demand is highly inelastic. Nevertheless, to tackle remaining doubts regarding our identification strategy, we apply several robustness tests. First, we use an alternative set of instruments in the spirit of Hausman (1996). That is, we use average values of the instruments in the 50 surrounding zip codes (while disregarding the actual zip code *i*) as alternative instruments. Thus, even if our instruments were to directly affect prices in some way, the values from neighboring zip codes should not, whereas they should still be correlated with the values in the respective zip code itself (Section G of the Online Appendix provides more details on this approach).[23] Second, we run IV regressions using either internet-based or age-based instruments, instead of both at the same time. The results stay robust to these alternative IV strategies, as we show in the

[23] The share of young households (U40) in neighboring zip codes should be correlated with the zip code of observation due to demographic effects (e.g. trends in migration, existence of a university, etc.). Also, broadband availability (BBA) in zip code *i* is correlated with BBA in neighboring zip codes, e.g. due to regional roll-out.

robustness section G of the Online Appendix.

4.2 Data

We use panel data at the German zip code level for the period 2011–2014.[24] As consumers typically have annual contracts, we aggregate all data to the annual level. Table 1 provides summary statistics of the variables in our regressions. Appendix Table B1 additionally reports the between and within standard deviations of our key variables, indicating that we have sufficient temporal and spatial variation. Figures F2–F8 in the Online Appendix provide heat maps of our main variables, search intensity and tariffs, visualizing their between and within variation.

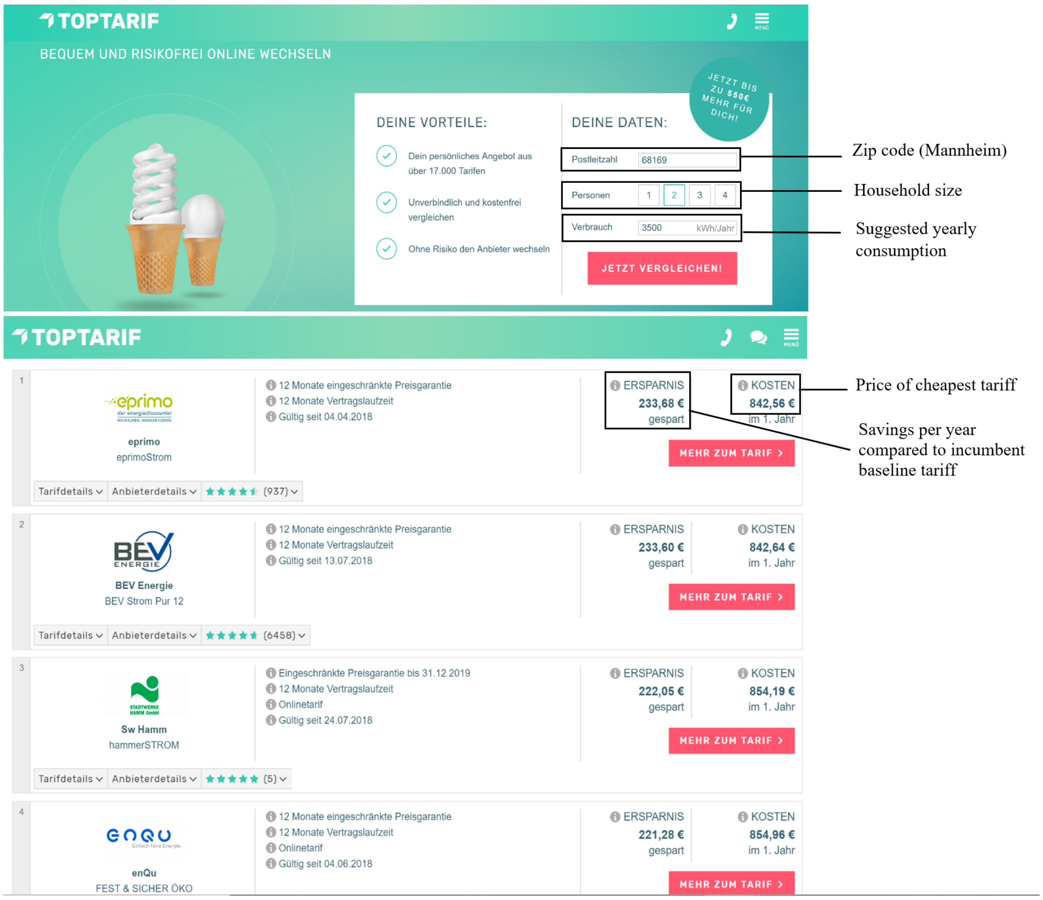
Tariffs. — *ene't*, a German software and data provider for the electricity industry, provided monthly data on retail electricity tariffs and cost components (except for P_L^I, which is already structured annually). In the estimations, we use gross prices (including 19% VAT) since they present the relevant price for end-consumers that are also displayed on the online platforms. We focus on a typical household with an annual consumption level of 3,500 kWh. This is the default consumption level suggested by all major price comparison platforms.[25] The summary statistics in Table 1 show that, on average, a household with 3,500 kWh annual consumption of electricity pays around 1,007 EUR per year if it stays with the incumbent's baseline tariff. The incumbent's cheaper tariff is around 8% lower at 931 EUR, while the overall cheapest entrant tariff is around 808 EUR (which is 20% cheaper than the incumbent default tariff).

Consumer search intensity. — *ene't* also provided the data on individual consumer search queries at several online price comparison sites, which enables us to construct a direct measure of consumer search intensity for each zip code and year. The database covers detailed information on all search queries conducted at several well-known online price comparison platforms including Toptarif.de (top tariff), Stromtipp.de (power hint), Energie-verbraucherportal.de (energy consumer portal), and mut-zum-wechseln.de (courage to switch), of which Toptarif.de is

[24]We have 8,226 zip codes in our data. However, there is an overlap of incumbency areas in some of the zip codes. That is, there may be an incumbent operating one part of a zip code and another incumbent operating another part. We drop all zip codes which have more than one incumbent, reducing the number of zip codes in our data to 7,249. In the robustness section G in the Online Appendix we show that our results change only marginally if we also include these zip codes into our sample and use averages of the different incumbent prices.

[25]3,500 kWh is also the household consumption level that is typically applied by other agencies (e.g. BNetzA, 2015) for comparing retail tariffs. ene't also provided tariff data for other annual consumption levels (2,000 kWh and 4,000 kWh), however only for P_H^I and P^E (but not for P_L^I). Regression estimates using P_H^I and P^E as well as $P_H^I - P^E$ for these alternative consumption levels yield robust results.

Figure 5: Screenshot of a typical online comparison platform



Note: Comparison platforms (here www.toptarif.de) list all available tariffs for a consumer given its expected annual consumption level for its local zip code, starting with the cheapest available tariff (including annual savings compared to the default incumbent baseline tariff). Site accessed on September 18, 2018.

by far the largest platform.[26] For each query, we observe a timestamp, the entered zip code for which the offered electricity tariffs are requested, the (expected) yearly consumption entered into the interface, if the search is performed by a household or an industrial customer, and consumer preferences (e.g. only "green" certificated tariffs). In addition, we are also able to track the search history: each platform user obtains a unique search-session ID (created by *ene't*) indicating the order of the queries from the same user).[27] Figure 5 provides a screenshot of the interface of a typical tariff comparison platform.

In sum, we have information on 35,855,071 search queries from 17,302,530 search sessions of which 96.7% (i.e. 16,778,214 sessions) are conducted by households and the remaining 3.3% (i.e.

[26]Toptarif is one of the three major electricity and gas price comparison websites, along with Verivox and Check24. It was acquired by Verivox in July 2014 but continues to operate as Toptarif.

[27]We are not able to observe actual switching, because clicking on a certain supplier tariff at the online comparison website redirects the searcher to a website where the switch may be finalized. This limitation is common to online data (see Koulayev, 2014). Yet, switching requires searching, so the impact of consumer search on price strategies seems to be consistently estimable. Brynjolfsson and Smith (2001) confirm this and find that factors that drive clicks are reasonable and unbiased indicators of sales, in their study of online book purchases.

524,316 sessions) by industrial customers. As many searchers conduct several search queries within a search session (e.g. comparing prices for different consumption levels) we focus on the number of search *sessions* per year and zip code (rather than on the absolute number of search *queries*).[28] Since our focus is on household consumers, we disregard search by industrial consumers. Furthermore, we exclude 551,256 search sessions, which exclusively consider eco-label (i.e. "green") certified tariffs.[29] Those searches are most likely not predominantly price driven and, on average, €152 more expensive than the cheapest tariff.

We construct our measure of search intensity as the number of search sessions within a zip code per year divided by the number of households:[30] $\mu_{it} = (SearchSessions_{it})/(Households_{it})$. At the mean, 9.1% of households within a zip code search for retail tariffs at one of our sample comparison platforms, whereas there is substantial variation ranging from 0% to 34.7%.

Instruments. — Data on the share of young households, measured as households with a household head below the age of 40 (*U40*), are obtained from *Acxiom* at a zip code-year resolution. Data on local broadband internet availability (*BBA*), which is the share of households which are offered internet speeds of 16 MBit/s or higher, are obtained from *Breitbandatlas.de*, with the same resolution. *BBA* data are at the municipal level and we aggregate them to the zip code level to match them with our remaining data. As a threshold, we chose a broadband rate of 16 MBit/s, as it provides substantial variation and gives the best first-stage F-statistic of excluded instruments.

Control variables. — We compute a variable reflecting retailers' net costs (excluding VAT) in order to control for spatial and time-variant cost differences. Detailed data on cost components are primarily obtained from *ene't* and include, for example, grid charges, concession fees, renewable energy surcharges ("EEG Umlage"), CHP (combined heat and power) surcharges ("KWK Umlage") and electricity taxes. Grid charges are paid by the electricity provider to the respective system operator and, thus, vary across grid areas (i.e. clusters of zip codes) and time as they are adjusted annually. The concession fee has to be paid by the system operator to the respective municipality for the right to install and operate electric cables on public roads. Hence, the concession fees vary at the municipality level and also over time. The remaining cost

[28]It should also be noted that a search session only contains the current search activity of an individual household and we cannot distinguish whether the same household starts a new search session on another date. Therefore, we treat each search session as conducted by an individual household.

[29]Nevertheless, our results are fully robust to the inclusion of eco-label searches.

[30]Since we observe some extreme outliers in some zip codes, apparently resulting from price comparing software " bots" or data crawling researchers, we truncate 2% of the upper bound of the sample distribution of our consumer information measure.

Table 1: Summary statistics

	Unit, data source	Mean	SD	Min	Max
Dependent variables					
Incumbent Base (P_H^I)	€/a, *ene't*	1007.51	77.63	800.00	1204.00
Incumbent Cheapest (P_L^I)	€/a, *ene't*	931.04	85.80	716.00	1117.00
Overall Cheapest (P^E)	€/a, *ene't*	808.32	58.62	617.00	903.00
Price Dispersion ($P_H^I - P^E$)	€/a, *ene't*	199.21	38.77	77.00	357.00
Price Discrimination ($P_H^I - P_L^I$)	€/a, *ene't*	76.51	41.32	0.00	282.00
Online Price Dispersion ($P_L^I - P^E$)	€/a, *ene't*	122.66	45.72	0.00	289.00
Variables of interest					
Search (μ)	%, *ene't*	9.63	7.11	1.39	52.42
Instruments					
U40	%, *Acxiom*	24.63	5.06	8.38	55.05
BBA	%, *breitbandatlas.de*	63.85	32.02	0.00	100.00
Control variables					
No. competitors	#, *ene't*	133.27	24.81	55.00	192.00
No. households	#, *Acxiom*	4.66	4.31	0.20	29.60
Costs	€/a, *ene't* & *EEX*	6.52	0.06	6.33	6.71
Average HH size	Integer, *Acxiom*	2.11	0.18	1.52	2.52
Purchase Power	K €/HH, *Acxiom*	43.34	7.53	21.03	110.34
Income <25k €/a	%, *Acxiom*	39.17	7.45	2.43	70.61
Income 25-50k €/year	%, *Acxiom*	32.24	2.55	13.76	52.92
Lower class social status	%, *Acxiom*	19.56	23.56	0.00	98.87
Middle class social status	%, *Acxiom*	55.14	21.96	0.00	100.00
High Urbanization	%, *Acxiom*	20.52	32.24	0.00	341.6
Unemployed	%, *Acxiom*	33.71	42.64	0.00	312.4
Population density	K inhibitants/km^2, *Acxiom*	1.00	2.42	0.00	30.81
Social insurance	%, *Acxiom*	71.38	4.03	46.14	99.17
Switching rate	%, *Verivox*	1.98	0.98	0.00	6.61
Obs.	24,175				

Notes: "Obs" are zip code-year observations. €/a refers to an annual electricity consumption of 3.5 MWh.

components only vary over time but not spatially. Moreover, we also add the one-year ahead future prices at the EEX spot market to our cost variable to proxy for the costs of wholesale electricity, as this one-year ahead price presents the standard purchasing strategy for retailers.[31]

To measure competition within a zip code, we use the number of electricity retail suppliers, as provided by *ene't*. The number of competitors in a zip code varies between 55 and 192.[32] Moreover, as can be inferred from Table B1 in the Appendix, the within-zip-code standard deviation is 19.8, suggesting that there is also a significant fluctuation of retailers within a zip code over time.

Other control variables refer to structural household characteristics, which we obtained from *Acxiom*. These variables are the purchase power, the share of households earning less than 25,000 Euro per year, the share of households earning between 25,000 and 50,000 Euro per year, the local unemployment rate, the share of socially insured people, the share of

[31]Even if retailers purchased electricity through other channels than via the power exchange (e.g. bilateral contracts, OTC, etc.), the price from the power exchange still represents the opportunity cost.

[32]These numbers may seem high but correspond well with BNetzA (2015) and may be the consequence of low entry barriers in the electricity retail market in Germany.

households classified as having a low social status, those classified a middle social status, the average household size, the population density, the number of households as well as the degree of urbanization.

Finally, we received data from *Verivox*, another major price comparison website, on the fraction of consumers in a zip code in a given year that switched their supplier. We include this variable lagged by one year in order to capture the general "turbulence" in a local market.

5 Results

In Table 2, we present the results of our IV estimations for the three retail prices of interest, P_H^I, P_L^I and P^E. As we use a log-log specification the coefficients can be interpreted as elasticities.[33] The instruments are sufficiently strongly correlated with the endogenous variable as shown by the high values of the first-stage effective F-test suggested by Olea and Pflueger (2013).[34] Results from the first stage estimation are reported in Table B2 in the Appendix. Also, the Durbin-Wu-Hausman test for endogeneity (Davidson and MacKinnon, 1993) suggests that the consumer search intensity μ should indeed be treated as endogenous as the null hypothesis of consumer search being an exogenous regressor is clearly rejected. Furthermore, the Hansen J statistic (Hansen, 1982) suggests that – conditional on at least one instrument being actually valid – our instruments are jointly exogenous.[35]

The corresponding OLS estimates are reported next to each IV estimate to give a sense of the endogeneity bias. Even though the sign and the significance are similar, the magnitudes of the OLS estimates are much lower, suggesting that neglecting endogeneity leads to a substantial underestimation of the impact of consumer search on prices.

Coming to the results, column 2 of Table 2 provides evidence that the incumbent reacts to a higher search intensity by increasing its baseline tariff. For a change in consumer search intensity by 10%, the incumbent raises its tariff by 0.46%. Column 4 shows that the incumbent reacts to more search activity in its zip code by reducing its cheaper online tariff considerably. For a 10% increase in search activity, the incumbent decreases its cheapest tariff by 1.14%.

[33]In Table G4 in the Online Appendix we show that the results are robust to a level-level specification.

[34]We apply the threshold for the 5 percent critical value for testing the null hypothesis that the two-stage least squares bias exceeds 10 percent of the ordinary least squares bias. In our case with two instruments, this threshold is 6.95. The effective first stage F-statistics and thresholds were calculated using Pflueger and Wang (2015), Stata command "weakivtest".

[35]Table G3 in the Online Appendix shows the results of a robustness check where we either use only instruments based on broadband internet availability or instruments only based on local age characteristics.

Moreover, column 6 reveals that the overall cheapest tariff in the market provided by an entrant supplier also decreases with more consumer search, whereas the effect is less pronounced than for the incumbents' cheapest tariffs. For every 10% increase in search intensity in a zip code the overall cheapest tariff in the market decreases by 0.31%. Thus, the incumbents' cheapest tariffs react more strongly to consumer search than the overall cheapest tariff.

The empirical effects can be explained along the lines of Proposition 2. If there are more low search cost consumers, then there will be more competition online and thus lower prices. To prevent too many consumers from switching to the entrant, the incumbent has to decrease its online price more aggressively than the entrants: the incumbent would lose a larger markup when losing a customer as the incumbents' cheapest price is still higher than the overall cheapest price offered by an entrant. At the same time, if there is still a considerable fraction of consumers with high enough search costs, the incumbent has an incentive to increase the margin on its baseline tariff as it will not lose too many consumers by doing so relative to the loyal consumers. Hence, the incumbency advantage can be exploited by price discriminating between loyal consumers and consumers who search online but have a brand attachment (loyalty premium).

A back-of-the-envelope calculation may show the reasonableness and economic importance of our estimates. Our estimates from Table 2 imply that the incumbent increases its base tariff by 8.6 euro if search intensity in a zip code increases by one within-zip-code standard deviation (which is 5.1 percentage points), taking as starting points the mean values of prices and search intensity (i.e. 1007 euro and 9.6%, respectively). Moreover, the incumbent decreases its online tariff due to the increased search activity in the zip code by 19.6 euro (mean value is 931 euro). The cheapest entrant decreases its tariff by a further 4.7 euro (mean value is 808 euro). Thus, we would expect from our estimates that price discrimination increases by 28.2 euro on average (which is 36.8% calculated from the mean value of price discrimination of 76.5 euro) due to a one standard deviation increase in search intensity within a zip code. Thus, increased search activity appears to be a substantial part of the explanation of why incumbents price discriminate in liberalized markets.

We now briefly discuss the impact of other control variables. We estimate the cost pass-through to the end-user retail tariffs, which is much higher in the competitive segments of the electricity retail market. For the incumbents' baseline tariffs, we estimate a pass-through of only around 21%, whereas 46% of cost increases are passed on to consumers for the incumbents'

Table 2: IV and OLS estimates of the impact of consumer search on prices (log-log)

	ln Incumbent base P^I_H		ln Incumbent cheapest P^I_L		ln Overall cheapest P^E	
	OLS (1)	IV (2)	OLS (3)	IV (4)	OLS (5)	IV (6)
ln Search (μ)	0.0032***	0.0460***	-0.0039***	-0.1138***	-0.0007**	-0.0313***
	(0.0005)	(0.0080)	(0.0014)	(0.0210)	(0.0004)	(0.0054)
No. competitors	0.0006***	0.0004***	0.0040***	0.0046***	0.0000	0.0002***
	(0.0000)	(0.0001)	(0.0001)	(0.0002)	(0.0000)	(0.0000)
No. households	-0.0000	0.0037***	0.0063**	-0.0032	0.0004	-0.0022**
	(0.0011)	(0.0014)	(0.0026)	(0.0034)	(0.0007)	(0.0009)
ln Costs	0.2359***	0.2140***	0.4023***	0.4585***	0.5006***	0.5165***
	(0.0085)	(0.0111)	(0.0190)	(0.0264)	(0.0076)	(0.0090)
Average HH size	0.0475***	0.0408***	-0.0025	0.0142	-0.0155***	-0.0106**
	(0.0057)	(0.0069)	(0.0151)	(0.0185)	(0.0041)	(0.0050)
ln Purchase power	0.0007	0.0062	-0.0261**	-0.0397***	-0.0054**	-0.0092***
	(0.0046)	(0.0063)	(0.0109)	(0.0128)	(0.0023)	(0.0031)
Income <25k €/a	-0.0000	-0.0001	-0.0018***	-0.0015***	-0.0003***	-0.0002***
	(0.0001)	(0.0001)	(0.0002)	(0.0003)	(0.0001)	(0.0001)
Income 25-50k €/year	-0.0002	-0.0003**	-0.0013***	-0.0008**	-0.0002***	-0.0001
	(0.0001)	(0.0002)	(0.0003)	(0.0003)	(0.0001)	(0.0001)
Lower class social status	0.0001***	0.0001***	-0.0001*	-0.0000	-0.0001***	-0.0001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Middle class social status	0.0001***	0.0001***	0.0001***	0.0001***	0.0000***	0.0000*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
High Urbanization	0.0081**	0.0178***	-0.0330***	-0.0580***	-0.0054**	-0.0124***
	(0.0039)	(0.0046)	(0.0087)	(0.0110)	(0.0026)	(0.0033)
Unemployed	0.0079**	0.0170***	0.0826***	0.0592***	0.0148***	0.0084***
	(0.0038)	(0.0043)	(0.0077)	(0.0099)	(0.0023)	(0.0029)
Population density	0.0002**	0.0003**	0.0002	-0.0001	0.0001	0.0000
	(0.0001)	(0.0001)	(0.0004)	(0.0004)	(0.0001)	(0.0001)
Social insurance	0.0008***	0.0010***	0.0007*	0.0000	0.0002	-0.0000
	(0.0002)	(0.0002)	(0.0004)	(0.0005)	(0.0001)	(0.0001)
ln Lagged switching rate	0.0064***	0.0050***	0.0243***	0.0280***	-0.0001	0.0009*
	(0.0005)	(0.0007)	(0.0015)	(0.0021)	(0.0004)	(0.0005)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Zip code FE	Yes	Yes	Yes	Yes	Yes	Yes
First stage effective F stat.		33.96		33.96		33.96
Hansen J stat. (p-value)		0.27		0.45		0.87
Durbin-Wu-Hausman test		0.00		0.00		0.00
Obs.	24,175	24,175	24,175	24,175	24,175	24,175

Notes: Standard errors clustered at the zip code level in parentheses. Instruments for μ in the IV estimations are *U40* and *BBA*. $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.

cheapest tariffs and 52% for the cheapest entrants' tariffs. These pass-through patterns are in line with Duso and Szücs (2017), who investigate pass-through in the German electricity retail markets and also find that incumbents pass-through costs to a lesser extent. Possibly surprisingly, the number of competitors positively affects all prices. However, the estimated effect of an additional competitor on prices is close to zero and thus economically negligible. Given low entry barriers and the already high number of competitors, it seems that additional entry does not have an economically significant impact on prices.[36] The coefficients of many of the remaining control variables related to income, social status, demographics, or wealth are highly collinear and we therefore refrain from an interpretation. The main reason why we include these variables is to make sure that the exclusion restriction holds, so that the coefficient of search is estimated without bias.

Table 3 presents estimates of the impact of consumer search on the three price dispersion measures. Column 1 focuses on *overall price dispersion*, measured as the incumbent's baseline tariff (P^I_H) minus the overall cheapest tariff (P^E). Evidently, price dispersion goes up if more consumers search, since the incumbent slightly increases its baseline tariff and at the same time the overall cheapest price declines with search: more search in a zip code region may indicate a larger relative mass of consumers with low search costs, which prompts the incumbent to increase its baseline tariff, while tariffs in the competitive segment decline due to increased competitive pressure. For every 10% increase in search intensity, the extent of price dispersion goes up by 4.0%, suggesting that consumers' gain from searching increases with the share of searching consumers.

Incumbents react to increased price pressure from consumer search via *price discrimination*, as they offer a cheaper tariff for searching consumers, which is still above the overall cheapest tariff in the market, and a high incumbent baseline tariff for loyal consumers who do not search. Price discrimination becomes more pronounced with increasing search intensity. An increase in the share of searching consumers by 10% widens the gap between the incumbent's baseline tariff and its cheaper tariff by 15.5%. The extent of price discrimination unambiguously increases if a larger share of consumers searches, predominantly because the incumbent decreases its

[36]Regardless of this, these coefficients have to be interpreted with caution, as they may not reflect causal effects: prices and number of competitors may influence each other. In the robustness section, we also present results from estimations where we additionally instrument for the number of competitors using Hausman-type instruments (Hausman, 1996) but the results only change marginally. Moreover, all results stay fully robust if we drop all covariates (the instrumental-variables regression counteracts an omitted-variables bias), as shown in Table G1 in the Online Appendix.

Table 3: IV estimates of the impact of consumer search on dispersion (log-log)

	ln Price Dispersion $P_H^I - P^E$		ln Price Discrimination $P_H^I - P_L^I$		ln Online Price Dispersion $P_H^I - P^E$	
	OLS (1)	IV (2)	OLS (3)	IV (4)	OLS (5)	IV (6)
ln Search (μ)	0.0150***	0.4017***	0.0609***	1.5469***	-0.0386*	-1.6272***
	(0.0030)	(0.0576)	(0.0180)	(0.2999)	(0.0212)	(0.3107)
No. competitors	0.0034***	0.0013***	-0.0357***	-0.0439***	0.0574***	0.0661***
	(0.0002)	(0.0004)	(0.0012)	(0.0022)	(0.0014)	(0.0025)
No. households	-0.0058	0.0275***	-0.1414***	-0.0131	0.1278***	-0.0112
	(0.0069)	(0.0097)	(0.0335)	(0.0450)	(0.0366)	(0.0497)
ln Costs	-0.8782***	-1.0797***	-1.3066***	-2.0781***	-0.9413***	-0.1306
	(0.0504)	(0.0772)	(0.2392)	(0.3447)	(0.2715)	(0.3853)
Average HH size	0.2858***	0.2250***	0.7121***	0.4816*	1.2155***	1.4563***
	(0.0352)	(0.0504)	(0.2149)	(0.2523)	(0.2307)	(0.2778)
ln Purchase power	0.0597**	0.1072**	0.4139***	0.5977***	-0.6785***	-0.8742***
	(0.0247)	(0.0418)	(0.1295)	(0.1684)	(0.1366)	(0.1628)
Income <25k €/a	0.0014**	0.0005	0.0164***	0.0130***	-0.0249***	-0.0212***
	(0.0005)	(0.0008)	(0.0029)	(0.0033)	(0.0037)	(0.0041)
Income 25-50k €/year	-0.0002	-0.0017	0.0114***	0.0056	-0.0111**	-0.0048
	(0.0008)	(0.0012)	(0.0039)	(0.0045)	(0.0053)	(0.0058)
Lower class social status	0.0007***	0.0007***	0.0019***	0.0017***	0.0014***	0.0017***
	(0.0001)	(0.0001)	(0.0004)	(0.0005)	(0.0005)	(0.0006)
Middle class social status	0.0005***	0.0005***	-0.0000	0.0002	0.0022***	0.0019***
	(0.0001)	(0.0001)	(0.0004)	(0.0004)	(0.0004)	(0.0005)
High Urbanization	0.0630**	0.1515***	0.5269***	0.8666***	-0.5065***	-0.8690***
	(0.0275)	(0.0351)	(0.1116)	(0.1465)	(0.1208)	(0.1549)
Unemployed	-0.0450*	0.0367	-0.5554***	-0.2408*	1.6357***	1.2969***
	(0.0257)	(0.0314)	(0.1181)	(0.1451)	(0.1112)	(0.1426)
Population density	0.0005	0.0015*	-0.0002	0.0034	-0.0003	-0.0041
	(0.0007)	(0.0009)	(0.0036)	(0.0040)	(0.0051)	(0.0052)
Social insurance	0.0024**	0.0049***	0.0178***	0.0276***	0.0280***	0.0174**
	(0.0012)	(0.0015)	(0.0055)	(0.0069)	(0.0083)	(0.0088)
ln Lagged switching rate	0.0408***	0.0278***	-0.2182***	-0.2683***	0.3509***	0.4042***
	(0.0033)	(0.0053)	(0.0200)	(0.0277)	(0.0228)	(0.0312)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Zip code FE	Yes	Yes	Yes	Yes	Yes	Yes
First stage effective F stat.		33.96		33.96		33.96
Hansen J stat. (p-value)		0.82		0.92		0.25
Durbin-Wu-Hausman test		0.00		0.00		0.00
Obs.	24,175	24,175	24,175	24,175	24,175	24,175

Notes: Standard errors clustered at the zip code level in parentheses. Instruments for μ in the IV estimations are *U40* and *BBA*. $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.

cheapest tariff significantly as a reaction to consumer search to aggressively prevent existing customers from switching to competitors. This can be explained in line with Proposition 1 of our theoretical model: more searching consumers imply more price discrimination if there are relatively sufficient many consumers left who are loyal and "always" buy at the baseline price of the incumbent.

We also see that *online price dispersion*, measured as the difference between the incumbent's cheapest tariff and the overall cheapest tariff in the market, narrows considerably with search intensity. The more consumers search in a market, the more the incumbent is forced to approach the overall cheapest price. For a 10% increase in search intensity, the "loyalty premium" narrows by 16.3%.

Overall, we find that the loyal consumers who stay with the incumbent's baseline tariff get "milked" when there are more searching consumers in a local market. In contrast, those who are willing to search either get a cheaper incumbent tariff, which includes a brand premium compared to the overall cheapest tariff in the market, or switch to a cheaper entrant tariff. The incumbent reacts to more consumer search with price discrimination by slightly increasing its baseline tariff while at the same time significantly reducing its cheaper online tariff. Entrants react to more search with somewhat lower prices. Intensified consumer search thus increases overall price dispersion and price discrimination, and it leads to a fiercer price competition (i.e. an alignment of incumbent and entrant prices) in the competitive online segment.

Our results are robust to a large number of alternative specifications, such as using alternative instruments (in the spirit of Hausman, 1996, i.e. mean average values of the instruments in the 50 surrounding zip codes), level-level estimation, allowing for a non-linear relationship between search and tariffs, and additionally instrumenting for the number of competitors. We present and discuss these specifications in Sections F and G of the Online Appendix.

6 Conclusion

We provide a theoretical model and an empirical analysis analyzing the pricing patterns in markets with incumbents and search frictions. A typical feature of such markets is a substantial information asymmetry between an incumbent provider and competing firms as consumers in an ongoing relation with a provider get the first price quote for free from their current provider, but have to pay a search cost to get informed of the offers by alternative providers. We show that managers of incumbent firms can effectively use this asymmetry to price discriminate between loyal consumers and searching consumers who consider leaving.

In our empirical analysis of the German retail electricity market, we show that the incumbent baseline price increases when consumers search more. The incentive to increase the baseline tariff arises if a lower price would not keep many consumers from searching and catering to high search cost consumers allows the incumbent to siphon off larger loyalty rents. In contrast, once consumers have shown a willingness to search (e.g. by conducting a price comparison on an online platform), the incumbent has a strong incentive to prevent consumers from switching to an entrant by setting a lower online price. That is, incumbents engage in price discrimination with a high baseline price and a lower online price. The incumbent's cheapest online tariff is still higher, however, than the overall cheapest tariff offered by an entrant competitor, as consumers – presumably – attach a brand or loyalty premium to the incumbent. In this way, the incumbent can simultaneously appropriate surplus from loyal consumers and prevent consumers who indicated a willingness to search from switching to an entrant. As few consumers actually switch, the incumbent appropriates an important share of market revenue. These findings are consistent with the search theoretic model that we develop to analyze the interaction between an incumbent and an entrant.

The results of the paper should also be of relevance to any other market where firms can price discriminate between loyal and searching consumers. After having built up a share of loyal customers themselves, entrants may also follow a similar strategy of price discrimination and increase their prices for their existing consumer base, while simultaneously setting a more competitive price to attract new customers. As entrants are very small in the German electricity markets, such strategy will likely not be of quantitative importance for our study, but may lead to new implications in other markets.

References

A.T. Kearney, 2012. Der Strom- und Gasvertrieb im Wandel: Unabhängige Anbieter am Scheideweg. Presentation, Berlin, December 2012.

Bar-Isaac, H., Caruana, G., Cuñat, V., 2012. Search, design, and market structure. American Economic Review 102, 1140–1160.

Baye, M.R., Morgan, J., Scholten, P., 2006. Information, search, and price dispersion, in: Hendershott, T. (Ed.), Handbook of Economics and Information Systems. Elsevier.

Berry, S., Haile, P., 2015. Identification in Differentiated Products Markets. Technical Report.

Blake, T., Nosko, C., Tadelis, S., 2016. Returns to consumer search: Evidence from eBay, in: Proceedings of the 2016 ACM Conference on Economics and Computation, ACM Conference.

BNetzA, 2013. Monitoring report 2013. Federal Network Agency of Germany.

BNetzA, 2015. Monitoring report 2015. Federal Network Agency of Germany.

Brynjolfsson, E., Smith, M., 2001. Consumer decision-making at an internet shopbot: Brand still matters. Journal of Industrial Economics 49, 541–558.

Brynjolfsson, E., Smith, M., 2014. Frictionless commerce? A comparison of internet and conventional retailers. Management Science 46, 563–585.

Byrne, D.P., De Roos, N., 2017. Consumer search in retail gasoline markets. The Journal of Industrial Economics 65, 183–193.

Byrne, D.P., Martin, L.A., Nah, J.S., 2019. Price discrimination, search, and negotiation in an oligopoly: A field experiment in retail electricity. Working Paper.

Cabral, L., 2016. Dynamic pricing in customer markets with switching costs. Review of Economic Dynamics 20, 43–62.

Cabral, L., 2017. Switching costs, search costs, and market power: Theory and application to energy markets. Closing Address, V International Academic Symposium, Challenges for the Energy Sector, Barcelona, February 7.

Cabral, L., Gilbukh, S., 2020. Rational buyers search when prices increase. Journal of Economic Theory 187, 104998.

Coey, D., Larsen, B.J., Platt, B.C., 2020. Discounts and deadlines in consumer search. American Economic Review 110, 3748–3785.

Davidson, R., MacKinnon, J.G., 1993. Estimation and Inference in Econometrics. Oxford University Press.

Davies, S., Price, C.W., Wilson, C.M., 2014. Nonlinear pricing and tariff differentiation: Evidence from the British electricity market. The Energy Journal 35, 57–77.

Doganoglu, T., 2010. Switching costs, experience goods and dynamic price competition. Quantitative Marketing and Economics 8, 167–205.

Dressler, L., Weiergraeber, S., 2019. Alert the inert! switching costs and limited awareness in retail electricity markets. Mimeo.

Duso, T., Szücs, F., 2017. Market power and asymmetric pass-through in German electricity retail. European Economic Review 98, 354–372.

Giulietti, M., Waterson, M., Wildenbeest, M., 2014. Estimation of search frictions in the British electricity market. The Journal of Industrial Economics 62, 555–590.

Goolsbee, A., Syverson, C., 2008. How do incumbents respond to the threat of entry? Evidence from the major airlines. Quarterly Journal of Economics 123, 1611–1633.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hausman, J., Leonard, G., Zona, J.D., 1994. Competitive analysis with differenciated products. Annales d'Économie et de Statistique , 159.

Hausman, J.A., 1996. Valuation of new goods under perfect and imperfect competition, in: Bresnahan, T.F., Gordon, R.J. (Eds.), The Economics of New Goods. University of Chicago Press, pp. 207–248.

Heim, S., 2021. Asymmetric cost pass-through and consumer search: empirical evidence from online platforms. Quantitative Marketing and Economics 19, 227–260.

Hortacsu, A., Madanizadeh, S., Puller, S., 2017. Power to choose? An analysis of consumer inertia in the residential electricity market. American Economic Journal: Economic Policy 9, 192–226.

Houde, J.F., Allen, J., Clark, R., 2019. Search frictions and market power in negotiated price markets. Journal of Political Economy 127.

Ishibashi, I., Matsushima, N., 2009. The existence of low-end firms may help high-end firms. Marketing Science 28, 136–147.

Janssen, C.W., Moraga-González, J.L., 2004. Strategic pricing, consumer search and the number of firms. Review of Economic Studies 71, 1089–1118.

Koulayev, S., 2014. Search for differentiated products: Identification and estimation. RAND Journal of Economics 45, 553–575.

Lach, S., Moraga-González, J.L., 2017. Asymmetric price effects of competition. The Journal of Industrial Economics , 767–803.

Nevo, A., 2000. Mergers with differentiated products: The case of the ready-to-eat cereal industry. RAND Journal of Economics 31, 395.

Olea, J.L.M., Pflueger, C., 2013. A robust test for weak instruments. Journal of Business & Economic Statistics 31, 358–369.

Pennersdorfer, D., Schmidt-Dengler, P., Schutz, N., Weiss, C., Yontcheva, B., 2020. Information and price dispersion: Theory and evidence. International Economic Review 61, 871–899.

Perloff, J.M., Suslow, V.Y., Seguin, P.J., 1995. Higher prices from entry: Pricing of brand-name drugs. CUDARE Working Papers 25104, University of California, Berkeley, Department of Agricultural and Resource Economics.

Pflueger, C.E., Wang, S., 2015. A robust test for weak instruments in stata. The Stata Journal 15, 216–225.

Salop, S., 1977. The noisy monopolist: Imperfect information, price dispersion and price discrimination. Review of Economic Studies 44, 393–406.

De los Santos, B., Hortacsu, A., Wildenbeest, M., 2012. Testing models of consumer search using data on web browsing and purchasing behavior. American Economic Review 102, 2955–2980.

Sorensen, A.T., 2000. Equilibrium price dispersion in retail markets for prescription drugs. Journal of Political Economy 108, 833–850.

Stahl, D.O., 1989. Oligopolistic pricing with sequential consumer search. American Economic Review 79, 700–712.

Tang, Z., Smith, M.D., Montgomery, A., 2010. The impact of shopbot use on prices and price dispersion: Evidence from online book retailing. International Journal of Industrial Organization 28, 579–590.

Tappata, M., 2009. Rockets and feathers: Understanding asymmetric pricing. RAND Journal of Economics 40, 673–687.

Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. The MIT Press.

Appendix

A Proof of Proposition

Proposition 1. The effect of an increase in brand loyalty θ is that all prices and online price dispersion increase, while overall price dispersion decreases and price discrimination is unaffected. Price discrimination increases if, and only if, the inverse hazard rate evaluated at the equilibrium values $\frac{1-F(P_H^{I*}-P_L^{I*};z)}{f(P_H^{I*}-P_L^{I*};z)}$ is increasing in z. Moreover, online price dispersion is positively related to P_L^I if the density functions are non-increasing, i.e., $\partial f\left(\frac{P_L^I-P^E}{\theta};z\right)/\partial(P_L^I-P^E)\leq 0$. Finally, online price dispersion and price discrimination are linked by $1+f(P_H^I-P_L^I;z)(P_H^I-P_L^I)=f\left(\frac{P_L^I-P^E}{\theta};z\right)\left(\frac{P_L^I+P^E}{\theta}\right)$.

Proof. The effects of changes in the loyalty parameter θ can be understood as follows. First, it clearly follows from (3) that $P_H^I-P_L^I$, which is the measure of price discrimination, is independent of θ. Recognizing this, it follows from deducting (1) from (2) that $\frac{P_L^I-P^E}{\theta}$ is also independent of θ, or in other words, that $P_L^I-P^E$, which is the measure of online price dispersion, is proportional to θ. From each of the individual equations (1) and (2) it then follows that both P_L^I and P^E have to be proportional to θ. Thus, all prices are increasing in θ. Overall price dispersion is decreasing in θ as P_H^I increases less than proportionally.

To understand the effects of changes in z, we first consider the result on price discrimination. Taking the total differential of (3) with respect to $P_H^I-P_L^I$ and z yields

$$\left[-2f(P_H^I-P_L^I;z)-\frac{\partial f(P_H^I-P_L^I;z)}{\partial(P_H^I-P_L^I)}(P_H^I-P_L^I)\right]d(P_H^I-P_L^I) = \left[\frac{\partial f(P_H^I-P_L^I;z)}{\partial z}(P_H^I-P_L^I)+\frac{F(P_H^I-P_L^I;z)}{\partial z}\right]dz. \tag{6}$$

As profit maximization implies that the second-order condition of (3) with respect to $P_H^I-P_L^I$ is negative, it should be that in an equilibrium,

$$-2f(P_H^I-P_L^I;z)-\frac{\partial f(P_H^I-P_L^I;z)}{\partial(P_H^I-P_L^I)}(P_H^I-P_L^I)<0.$$

On the other hand, the inverse hazard rate $\frac{1-F(P_H^{I*}-P_L^{I*};z)}{f(P_H^{I*}-P_L^{I*};z)}$ is increasing in z if and only if

$$-\frac{\partial f(P_H^I-P_L^I;z)}{\partial z}(1-F(P_H^I-P_L^I;z))-\frac{F(P_H^I-P_L^I;z)}{\partial z}f(P_H^I-P_L^I)>0,$$

which using (3) can be rewritten as

$$-f(P_H^I - P_L^I; z)\left[\frac{\partial f(P_H^I - P_L^I; z)}{\partial z}(P_H^I - P_L^I) + \frac{F(P_H^I - P_L^I; z)}{\partial z}\right] > 0.$$

Thus, if the inverse hazard rate is increasing in z, then in any equilibrium both square bracket terms in (6) are negative, implying $\frac{d(P_H^I - P_L^I)}{dz} > 0$.

To investigate online price dispersion, we take the total differential of (1) with respect to P_L^I and P^E to obtain

$$0 = \frac{1}{\theta}\left[f\left(\frac{P_L^I - P^E}{\theta}; z\right) - f'\left(\frac{P_L^I - P^E}{\theta}; z\right)\frac{P^E}{\theta}\right] dP_L^I$$
$$+ \frac{1}{\theta}\left[-2f\left(\frac{P_L^I - P^E}{\theta}; z\right) + f'\left(\frac{P_L^I - P^E}{\theta}; z\right)\frac{P^E}{\theta}\right] dP^E,$$

where f' is the derivative of the density function with respect to prices. From the second-order condition for profit maximization by the entrant, we know that the second term in square brackets must be negative. If $f'\left(\frac{P_L^I - P^E}{\theta}; z\right) \leq 0$, then the first term in square brackets is positive, and its absolute value is smaller than the first term in square brackets. Thus, $0 < dP^E/dP_L^I < 1$. Therefore, $0 < d(P_L^I - P^E)/dP_L^I < 1$.

Finally, to understand how price discrimination and online price dispersion are related, we substitute (1) and (3) into (2) to get the condition stated in the Proposition.

B Additional Figures and Tables

Table B1: Decomposition of standard deviations between and within zip codes

Variable	Mean	SD overall	SD between	SD within
Incumbents' baseline tariff (P_H^I)	1,007	77.6	41.4	67.7
Incumbents' cheaper online tariffs (P_L^I)	931	85.8	42.5	76.9
Entrants' overall cheapest tariffs (P^E)	808	58.6	25.3	54.3
Overall price dispersion ($P_H^I - P^E$)	199.2	38.7	33.9	19.4
Price discrimination ($P_H^I - P_L^I$)	76.5	41.3	26.1	33.0
Onlince price dispersion ($P_L^I - P^E$)	122.7	45.7	30.9	35.2
Consumer search intensity (μ)	9.6	7.1	4.5	5.1
Head of HH below age of 40 (U40)	24.6	5.1	5.1	0.8
Broadband internet availability (BBA)	63.9	32.0	29.7	12.9
No. Competitors (#)	133.3	24.8	15.8	19.4
Net costs	683.0	42.3	30.0	30.7

Table B2: First-stage regressions of consumer search (μ)

	ln Search (μ) (1)	ln Search (μ) (2)
U40	0.0199***	0.0235***
	(0.0022)	(0.0023)
BBA	0.0004***	0.0003**
	(0.0001)	(0.0001)
No. competitors		0.0053***
		(0.0005)
No. households		-0.0958***
		(0.0146)
ln Costs		0.4892***
		(0.1106)
Average HH size		0.3231***
		(0.0924)
ln Purchase Power		-0.1590***
		(0.0586)
Income <25k €/a		0.0018
		(0.0013)
Income 25-50k €/year		0.0034**
		(0.0017)
Lower class social status		0.0001
		(0.0002)
Middle class social status		-0.0002
		(0.0002)
High Urbanization		-0.2386***
		(0.0449)
Unemployed		-0.1698***
		(0.0466)
Population density		-0.0023
		(0.0021)
Social insurance		-0.0052**
		(0.0024)
ln Lagged switching rate		0.0323***
		(0.0091)
Year FE	Yes	Yes
Zip code FE	Yes	Yes
Obs.	24,175	24,175

Notes: Standard errors clustered at the zip code level in parentheses. $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.

Online Appendix

Incumbency Advantages: Price Dispersion, Price Discrimination and Consumer Search at Online Platforms

Klaus Gugler[†] Sven Heim[‡] Maarten Janssen[§] Mario Liebensteiner[¶]

November 14, 2021

Abstract

This document contains additional results, which are not included in the main paper.

[†]Vienna University of Economics and Business (WU), klaus.gugler@wu.ac.at

[‡]Mines ParisTech, Paris, and ZEW Center for European Economic Research Mannheim, sven.heim@mines-paristech.fr

[§]University of Vienna, National Research University Higher School of Economics Moscow and CEPR, maarten.janssen@univie.ac.at

[¶]Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), mario.liebensteiner@fau.de

C A piece-wise linear search cost distribution

In this part of the online Appendix, we analyze the impact of changes of the search cost distribution on prices and the fraction of searchers for the piece-wise linear search cost distribution. We focus on parameter values such that $\widehat{s}_1 < \widetilde{s}_1 < \widehat{s}_2 < \widetilde{s}_2$, i.e., the consumer that is indifferent between two online offers is in the first interval of the search cost distribution, while the consumer that is indifferent between searching and not searching is in the second interval of the search cost distribution. With this formulation, an increase in z unambiguously leads the search cost distribution to have a larger fraction of consumers with lower search cost and a smaller fraction of consumers with intermediate search cost.

It is important to note that the number of active searchers $F(P_H^{I^*} - P_L^{I^e})$ is endogenously determined by the equilibrium prices. To dethermine the number of active searchers, we first determine the level of price discrimination $P_H^{I^*} - P_L^{I^*}$. Using (3) it is easy to see that for the case where the search cost distribution is piece-wise linear the equilibrium level of price discrimination equals

$$P_H^{I^*} - P_L^{I^*} = \frac{\widetilde{s}_2 - \widetilde{s}_1 - (z-1)\widetilde{s}_1\widetilde{s}_2}{2\left(\widetilde{s}_2 - z\widetilde{s}_1\right)} \tag{7}$$

and thus that the equilibrium fraction of online searchers equals

$$F(P_H^{I^*} - P_L^{I^*}) = \frac{(z-1)\widetilde{s}_1\widetilde{s}_2 + \widetilde{s}_2 - \widetilde{s}_1}{2\left(\widetilde{s}_2 - \widetilde{s}_1\right)}.$$

Applying the piece-wise linear search cost distribution to (1) and (2), it is easy to see that the relation between the equilibrium online prices is given by $P^{E^*} = \frac{1}{2}P_L^{I^*}$ so that

$$P_L^{I^*} = \frac{\widetilde{s}_2 - \widetilde{s}_1 + (z-1)\widetilde{s}_1\widetilde{s}_2}{3z\left(\widetilde{s}_2 - \widetilde{s}_1\right)}\theta,$$

which implies that

$$P^{E^*} = \frac{\widetilde{s}_2 - \widetilde{s}_1 + (z-1)\widetilde{s}_1\widetilde{s}_2}{6z\left(\widetilde{s}_2 - \widetilde{s}_1\right)}\theta$$

and

$$P_H^{I^*} = \frac{\widetilde{s}_2 - \widetilde{s}_1 + (z-1)\widetilde{s}_1\widetilde{s}_2}{3z\left(\widetilde{s}_2 - \widetilde{s}_1\right)}\theta + \frac{\widetilde{s}_2 - \widetilde{s}_1 - (z-1)\widetilde{s}_1\widetilde{s}_2}{2\left(\widetilde{s}_2 - z\widetilde{s}_1\right)}.$$

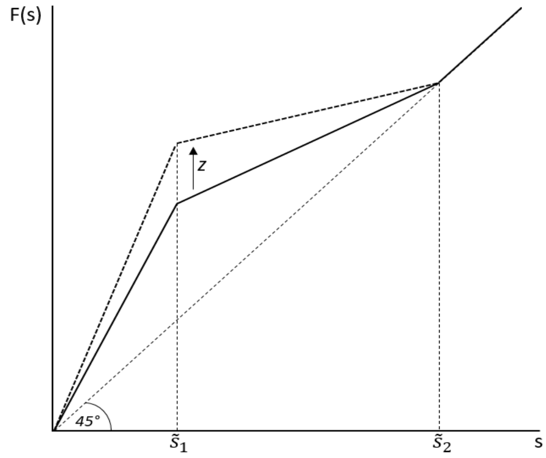
Using Proposition 1 and the fact that for a piece-wise linear distribution $f' = 0$ in the interior of the intervals, online equilibrium prices always change in the same direction and the level of online price dispersion $P_L^{I^*} - P^{E^*}$ positively correlates with both prices. Using the expressions for the different prices, it is easy to see that the condition $\widehat{s}_1 < \widetilde{s}_1 < \widehat{s}_2 < \widetilde{s}_2$ is satisfied if

$$\frac{\widetilde{s}_2 - \widetilde{s}_1 - \widetilde{s}_1\widetilde{s}_2}{\widetilde{s}_1(5\widetilde{s}_2 - 6\widetilde{s}_1)} < z < \frac{\widetilde{s}_2^2 - (\widetilde{s}_2 - \widetilde{s}_1)(1 - \widetilde{s}_2)}{\widetilde{s}_1\widetilde{s}_2}. \tag{8}$$

The proposition below contains the comparative statics properties of our model in terms of price discrimination and dispersion using the piece-wise linear search cost distribution.

Proposition 2 (price levels). If (8) holds, then an increase in the fraction of online searchers $F(P_H^{I^*} - P_L^{I^*})$, initiated by an increase in z, coincides with a decrease in online prices P^{E^*} and $P_L^{I^*}$ if and only if $\widetilde{s}_2 - \widetilde{s}_1 > \widetilde{s}_2\widetilde{s}_1$, while it coincides with an increase in the baseline price $P_H^{I^*}$ if θ is

Figure C1: A piece-wise linear search cost distribution



Notes: An increase in z shifts the piece-wise linear search cost distribution such that there is more mass of consumers with lower search costs.

small enough, z is large enough, or $\widetilde{s}_2 - \widetilde{s}_1$ is small enough.

Proof. It is clear that

$$\frac{\partial(P_H^{I^*} - P_L^{I^*})}{\partial z} = \frac{(\widetilde{s}_2 - \widetilde{s}_1)\widetilde{s}_1(1 - \widetilde{s}_2)}{2\left(\widetilde{s}_2 - z\widetilde{s}_1\right)^2} > 0.$$

From the expressions determining equilibrium prices, it follows that

$$2\frac{\partial P^{E^*}}{\partial z} = \frac{\partial P_L^{I^*}}{\partial z} = -\frac{\theta}{3z^2}\left(\frac{\widetilde{s}_2 - \widetilde{s}_1 - \widetilde{s}_1\widetilde{s}_2}{\widetilde{s}_2 - \widetilde{s}_1}\right),$$

which is clearly negative if and only if $\widetilde{s}_2 - \widetilde{s}_1 > \widetilde{s}_2\widetilde{s}_1$. Also,

$$\frac{\partial P_H^{I^*}}{\partial z} = \frac{\theta}{3z^2}\left(-1 + \frac{\widetilde{s}_1\widetilde{s}_2}{\widetilde{s}_2 - \widetilde{s}_1}\right) + \frac{(\widetilde{s}_2 - \widetilde{s}_1)\widetilde{s}_1(1 - \widetilde{s}_2)}{2\left(\widetilde{s}_2 - z\widetilde{s}_1\right)^2}.$$

As the second term is positive, this is clearly positive if either the first term is small enough (θ is small enough or z is large enough), or the first term is positive ($\widetilde{s}_2 - \widetilde{s}_1$ is small enough). *Q.E.D.*

These results demonstrate that price discrimination maximizes an incumbent firm's profits, as long as it is possible to charge searching and loyal consumers different tariffs. The results can be explained as follows. First, Proposition 1 already stated that price discrimination increases if the inverse hazard condition is satisfied, which is the case for the piece-wise linear distribution. Second, if $\widetilde{s}_2 - \widetilde{s}_1 > \widetilde{s}_2\widetilde{s}_1$, then $\widetilde{s}_1$ is relatively far away from $\widetilde{s}_2$. In this case, if z increases, there are relatively many online consumers that have a relatively low transaction cost to switch away from the incumbent. This gives the incumbent little market power on the online platform, resulting in a lower online price. The result then follows as Proposition 1 already indicated that online price dispersion is positively correlated with the incumbent's online price. Finally, overall price dispersion is closely related to the incumbent's baseline price. That price (and overall price dispersion) is increasing under two broad set of conditions. First, if θ is relatively small, there is fierce competition online and the more consumers search online, the more the incumbent wants to extract surplus from the consumers with high search costs. Second, if z is relatively large, or $\widetilde{s}_2 - \widetilde{s}_1$ is relatively small, there are relatively few consumers that have their decision on whether or not to search be influenced by the base line price, giving the incumbent an incentive to increase its baseline price.

Proposition 3 (price discrimination and dispersion). If (8) holds, then an increase in the fraction of online searchers $F(P_H^{I^*} - P_L^{I^*})$, initiated by an increase in z, coincides with (i) an increase in price discrimination $P_H^{I^*} - P_L^{I^*}$ and (ii) a decrease in online price dispersion $P_L^{I^*} - P^{E^*}$, if and only if $\widetilde{s}_2 - \widetilde{s}_1 > \widetilde{s}_1\widetilde{s}_2$ and (iii) an increase in overall price dispersion $P_H^{I^*} - P^{E^*}$ if θ is small enough, z is large enough, or $\widetilde{s}_2 - \widetilde{s}_1$ is small enough.

Proof. The proof simply follows from calculating the different partial derivatives. As

$$\frac{\partial(P_H^{I^*} - P_L^{I^*})}{\partial z} = \frac{(\widetilde{s}_2 - \widetilde{s}_1)\widetilde{s}_1(1 - \widetilde{s}_2)}{2\left(\widetilde{s}_2 - z\widetilde{s}_1\right)^2} > 0$$

and

$$\frac{\partial F(P_H^{I^*} - P_L^{I^*})}{\partial z} = \frac{\widetilde{s}_1\widetilde{s}_2}{2\left(\widetilde{s}_2 - \widetilde{s}_1\right)} > 0,$$

an increase in the fraction of online searchers, initiated by an increase in z, certainly leads to an increase in price discrimination $P_H^{I^*} - P_L^{I^*}$. As

$$\frac{\partial(P_L^{I^*} - P^{E^*})}{\partial z} = -\frac{\theta}{6z^2}\left(\frac{\widetilde{s}_2 - \widetilde{s}_1 - \widetilde{s}_1\widetilde{s}_2}{\widetilde{s}_2 - \widetilde{s}_1}\right)$$

it leads to a decrease in online price dispersion if $\widetilde{s}_2 - \widetilde{s}_1 - \widetilde{s}_1\widetilde{s}_2 > 0$. Finally, as

$$\frac{\partial(P_H^{I^*} - P^{E^*})}{\partial z} = \frac{\theta}{6z^2}\left(-1 + \frac{\widetilde{s}_1\widetilde{s}_2}{\widetilde{s}_2 - \widetilde{s}_1}\right) + \frac{(\widetilde{s}_2 - \widetilde{s}_1)\widetilde{s}_1(1 - \widetilde{s}_2)}{2\left(\widetilde{s}_2 - z\widetilde{s}_1\right)^2},$$

and the second term is positive, it leads to an increase in price discrimination if either the first term is small enough (θ is small enough or z is large enough), or the first term is positive ($\widetilde{s}_2 - \widetilde{s}_1$ is small enough). *Q.E.D*

D Welfare Analysis of Price Discrimination

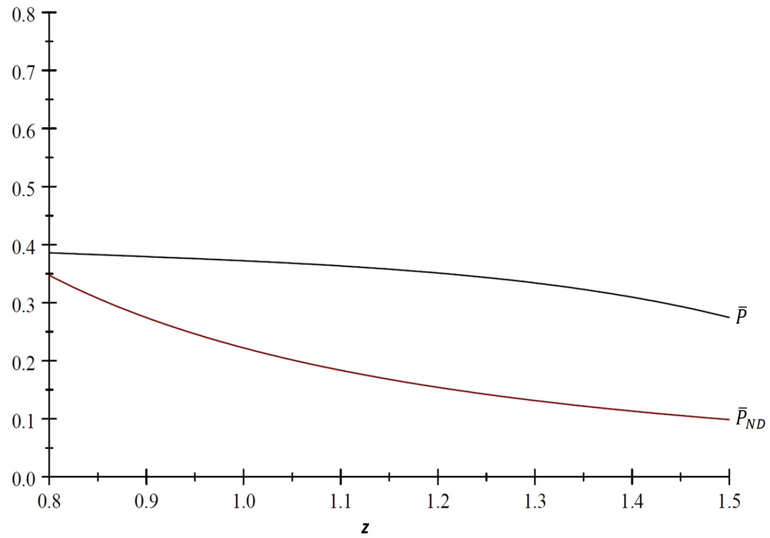
We will now analyze whether consumers are better off without price discrimination. To this end, we simply force $P_H^I = P_L^I$ (and denote this value by P^I) and solve for the equilibrium values, denoting the price choice of the entrant under "no discrimination" by P_{ND}^E (to distinguish it from the price it chooses when the incumbent can price discriminate). As now we have that

$$\pi_E = F(\widehat{s}_1; z)P^E = F\left(\frac{P^I - P_{ND}^E}{\theta}; z\right)P_{ND}^E$$

and

$$\pi_I = [1 - F(\widehat{s}_1; z)]\,P^I = \left[1 - F\left(\frac{P^I - P_{ND}^E}{\theta}; z\right)\right]P^I,$$

Figure D1: Average prices with and without price discrimination



Notes: The figure predicts price changes as a function of z with $\widetilde{s}_2 = 3/5$ and $\widetilde{s}_1 = 1/5$ and $\theta = 2/5$. $\bar{P}$ represents the average tariff with price discrimination; $\bar{P}_{ND}$ represents the average tariff without price discrimination.

it is easy to see that the two F.O.C.s are given by

$$F\left(\frac{P^I - P^E_{ND}}{\theta}; z\right) - f\left(\frac{P^I - P^E_{ND}}{\theta}; z\right)\frac{P^E_{ND}}{\theta} = 0,$$

and

$$1 - F\left(\frac{P^I - P^E_{ND}}{\theta}; z\right) - f\left(\frac{P^I - P^E_{ND}}{\theta}; z\right)\frac{P^I}{\theta} = 0.$$

Note that these conditions are very close to (1) and (2). In particular, it is clear that as $F(P^I_H - P^I_L; z) < 1$ in (2) in equilibrium $P^I_L < P^I$ and that because of the strategic complementarity of the price strategies, $P^E < P^E_{ND}$. Thus, searching consumers are better off with price discrimination. Intuitively, without price discrimination the incumbent has a larger share of "loyal" consumers it serves with the price P^I, compared to when it can price discriminate where P^I_L is meant to compete with the entrant's price and the large share of loyal consumers is "addressed" by P^I_H. Thus, with price discrimination, there is simply more online competition to attract searching consumers.

To compare P^I_H and P^I for the general case (and thus to make an overall comparison of the average price consumers pay[1]) is more difficult. Intuitively, though, it would be natural to have that $P^I_H > P^I$, as under price discrimination the incumbent does not need to directly compete with the entrant's price when setting P^I_H. This is easily confirmed for the uniform distribution of search costs, where $z = 1$. In that case $P^{E^*} = \frac{\theta}{6}, P^{I^*}_L = \frac{\theta}{3}$, and $P^{I^*}_H = \frac{1}{2} + \frac{\theta}{3}$, while $P^{E^*}_{ND} = \frac{\theta}{3}, P^{I^*} = \frac{2\theta}{3}$. As $\theta < 1$ it is easy to see that $P^{I^*} < P^{I^*}_H$.

For the case of the uniform distribution, it is also easy to calculate the average price con-

[1] One can also inquire into how the average price depends on the search intensity. The weighted average price is given by $(1 - (F(\widehat{s}_2))P^{I^*}_H + (F(\widehat{s}_2) - F(\widehat{s}_1))P^{I^*}_L + F(\widehat{s}_1)P^{E^*} = P^{I^*}_H - F(\widehat{s}_2)(P^{I^*}_H - P^{I^*}_L) - F(\widehat{s}_1)(P^{I^*}_L - P^{E^*})$.

sumers pay as $\frac{1}{2}\left(\frac{1}{2}+\frac{\theta}{3}\right)+\frac{1}{3}\frac{\theta}{3}+\frac{1}{6}\frac{\theta}{6}=\frac{1}{4}+\frac{11\theta}{36}$, for the case of price discrimination, while without price discrimination, the average price equals $\frac{2}{3}\frac{2\theta}{3}+\frac{1}{3}\frac{\theta}{3}=\frac{5\theta}{9}$. It follows that as $\theta<1$, on average, the effect of the higher baseline price $P_H^{I^*}$dominates and that consumers are worse off under price discrimination. Figure D1 confirms that, for the piece-wise linear distribution that we have considered above and for the same parameter values ($\widetilde{s}_2=3/5$ and $\widetilde{s}_1=1/5$ and $\theta=2/5$), on average consumers are worse off under price discrimination. This average hides, however, that searching consumers are better off, and that loyal consumers are considerably worse off, under price discrimination. In addition, with price discrimination, there will be fewer consumers switching to entrants than if price discrimination were banned. Note that the fraction of switchers is given by $F\left(\frac{P_L^I-P^E}{\theta};z\right)$ for the case of price discrimination and by $F\left(\frac{P^I-P_{ND}^E}{\theta};z\right)$ when price discrimination is banned. We have argued that $P_L^I<P^I$ and $P^E<P_{ND}^E$. As for the piece-wise linear distribution $P_{ND}^E=P^I/2$ and $P^E=P_L^I/2$, it follows that $\frac{P_L^I-P^E}{\theta}<\frac{P^I-P_{ND}^E}{\theta}$ so that fewer consumers switch under price discrimination.

E Sequential price setting game

When the firms compete in a sequential price setting game, in which the incumbent sets its baseline rate first, the respective profits of the entrant and incumbent do not change and are as given in the main text:

$$\pi_E=F(\widehat{s}_1;z)P^E=F\left(\frac{P_L^I-P^E}{\theta};z\right)P^E$$

and

$$\begin{aligned}\pi_I&=\left[F(\widehat{s}_2;z)-F(\widehat{s}_1;z)\right]P_L^I+(1-F(\widehat{s}_2;z))P_H^I\\&=\left[F(P_H^I-P_L^{I^e};z)-F\left(\frac{P_H^I-P^E}{\theta};z\right)\right]P_L^I+(1-F(P_H^I-P_L^{I^e});z)P_H^I.\end{aligned}$$

In taking the first-order conditions, one has to be careful in this "Stackelberg" environment where, in the second stage, the incumbent sets the online price P_L^I simultaneously with the entrant choosing P^E, and the incumbent chooses the baseline price P_H^I in the first stage. In this case, when setting online prices, both players have to take the number of consumers who search online, i.e., $F(P_H^I-P_L^{I^e})$, as given. Thus, if (as explained in the main text) both online prices react to the incumbent baseline price, the F.O.C.s (evaluated at the equilibrium where $P_L^{I^e}=P_L^{I^*}$) for the online prices (for given P_H^I), do not change either, so that they are given by:

$$F\left(\frac{P_L^I-P^E}{\theta};z\right)-f\left(\frac{P_L^I-P^E}{\theta};z\right)\frac{P^E}{\theta}=0$$

and

$$F(P_H^I-P_L^I;z)-F\left(\frac{P_L^I-P^E}{\theta};z\right)-f\left(\frac{P_L^I-P^E}{\theta};z\right)\frac{P_H^I}{\theta}=0,$$

respectively. This determines the online prices for given $P_H^I: P^E(P_H^I)$ and $P_L^I(P_H^I)$.

However, in determining the baseline price under "Stackelberg", the incumbent and the consumers take these reactions into account. Thus, when observing P_H^I, consumers realize that the second stage prices will be affected by a change in P_H^I. Thus, the incumbent sets P_H^I such that

$$0 = -f(P_H^I - P_L^I; z)(P_H^I - P_L^I)(1 - \frac{\partial P_L^I}{\partial P_H^I}) - \frac{P_H^I}{\theta} f\left(\frac{P_L^I - P^E}{\theta}; z\right) \frac{\partial (P_L^I - P^E)}{\partial P_H^I}$$
$$+ \left[F(P_H^I - P_L^I; z) - F\left(\frac{P_L^I - P^E}{\theta}; z\right)\right] \frac{\partial P_L^I}{\partial P_H^I} + (1 - F(P_H^I - P_L^I; z))$$

This expression has several new terms compared to the F.O.C. for P_H^I in the simultaneous choice model analyzed in the main text as, when setting P_H^I the incumbent (and the consumers) now consider how both online prices and the market shares change in response to changes in P_H^I.

For general distribution functions, it is not possible to solve these three equations in a meaningful way. Thus, in the rest of this appendix we consider the piece-wise linear distribution, where (as in the main text) we consider $\frac{P_L^I - P^E}{\theta} < \widetilde{s}_1 < P_H^I - P_L^I < \widetilde{s}_2$. As in the main text, the solution to (1) yields $P^E = P_L^I/2$, while in combination with (2) we have

$$\left(\frac{3z}{2\theta} + \frac{\widetilde{s}_2 - z\widetilde{s}_1}{\widetilde{s}_2 - \widetilde{s}_1}\right) P_L^I = \frac{\widetilde{s}_2 - z\widetilde{s}_1}{\widetilde{s}_2 - \widetilde{s}_1} P_H^I + \frac{(z-1)\widetilde{s}_1\widetilde{s}_2}{\widetilde{s}_2 - \widetilde{s}_1}.$$

Note from this equation it is clear that online prices are increasing in P_H^I but not to the full extent. In particular, $0 < \frac{\partial (P_L^I - P^E)}{\partial P_H^I} < \frac{\partial P_L^I}{\partial P_H^I} < 1$. Thus, if $\frac{P_L^I - P^E}{\theta} < \widetilde{s}_1 < P_H^I - P_L^I < \widetilde{s}_2$ the incumbent base line price solves

$$0 = -\frac{\widetilde{s}_2 - z\widetilde{s}_1}{\widetilde{s}_2 - \widetilde{s}_1}(P_H^I - P_L^I)\left(1 - \frac{\frac{\widetilde{s}_2 - z\widetilde{s}_1}{\widetilde{s}_2 - \widetilde{s}_1}}{\frac{3z}{2\theta} + \frac{\widetilde{s}_2 - z\widetilde{s}_1}{\widetilde{s}_2 - \widetilde{s}_1}}\right) - \frac{z}{2\theta} \frac{\frac{\widetilde{s}_2 - z\widetilde{s}_1}{\widetilde{s}_2 - \widetilde{s}_1}}{\frac{3z}{2\theta} + \frac{\widetilde{s}_2 - z\widetilde{s}_1}{\widetilde{s}_2 - \widetilde{s}_1}} P_H^I$$
$$+ \left[\frac{(z-1)\widetilde{s}_1\widetilde{s}_2}{\widetilde{s}_2 - \widetilde{s}_1} - \frac{\widetilde{s}_2 - z\widetilde{s}_1}{\widetilde{s}_2 - \widetilde{s}_1}(P_H^I - P_L^I) - z\frac{P_L^I - P^E}{\theta}\right] \frac{\frac{\widetilde{s}_2 - z\widetilde{s}_1}{\widetilde{s}_2 - \widetilde{s}_1}}{\frac{3z}{2\theta} + \frac{\widetilde{s}_2 - z\widetilde{s}_1}{\widetilde{s}_2 - \widetilde{s}_1}}$$
$$+ 1 - \frac{(z-1)\widetilde{s}_1\widetilde{s}_2}{\widetilde{s}_2 - \widetilde{s}_1} - \frac{\widetilde{s}_2 - z\widetilde{s}_1}{\widetilde{s}_2 - \widetilde{s}_1}(P_H^I - P_L^I),$$

or as

$$\left(\frac{3z}{2\theta} + \frac{\widetilde{s}_2 - z\widetilde{s}_1}{\widetilde{s}_2 - \widetilde{s}_1}\right) P_L^I = \frac{\widetilde{s}_2 - z\widetilde{s}_1}{\widetilde{s}_2 - \widetilde{s}_1} P_H^I + \frac{(z-1)\widetilde{s}_1\widetilde{s}_2}{\widetilde{s}_2 - \widetilde{s}_1}$$

we have that

$$0 = -2\left(\frac{3z}{2\theta}P_L^I - \frac{(z-1)\widetilde{s}_1\widetilde{s}_2}{\widetilde{s}_2 - \widetilde{s}_1}\right) - \frac{z}{\theta} \frac{\frac{\widetilde{s}_2 - z\widetilde{s}_1}{\widetilde{s}_2 - \widetilde{s}_1}}{\frac{3z}{2\theta} + \frac{\widetilde{s}_2 - z\widetilde{s}_1}{\widetilde{s}_2 - \widetilde{s}_1}} P_L^I$$
$$+ 1 - \frac{(z-1)\widetilde{s}_1\widetilde{s}_2}{\widetilde{s}_2 - \widetilde{s}_1}\left(1 - \frac{\frac{\widetilde{s}_2 - z\widetilde{s}_1}{\widetilde{s}_2 - \widetilde{s}_1}}{\frac{3z}{2\theta} + \frac{\widetilde{s}_2 - z\widetilde{s}_1}{\widetilde{s}_2 - \widetilde{s}_1}}\right),$$

which can be simplified to

$$\frac{z}{\theta}\left(3+\frac{\frac{\widetilde{s}_2-z\widetilde{s}_1}{\widetilde{s}_2-\widetilde{s}_1}}{\frac{3z}{2\theta}+\frac{\widetilde{s}_2-z\widetilde{s}_1}{\widetilde{s}_2-\widetilde{s}_1}}\right)P_L^I=1+\frac{(z-1)\widetilde{s}_1\widetilde{s}_2}{\widetilde{s}_2-\widetilde{s}_1}\left(1+\frac{\frac{\widetilde{s}_2-z\widetilde{s}_1}{\widetilde{s}_2-\widetilde{s}_1}}{\frac{3z}{2\theta}+\frac{\widetilde{s}_2-z\widetilde{s}_1}{\widetilde{s}_2-\widetilde{s}_1}}\right).$$

Thus, we have that the different equilibrium prices for the incumbent are given by

$$P_L^I=\frac{1+\frac{(z-1)\widetilde{s}_1\widetilde{s}_2}{\widetilde{s}_2-\widetilde{s}_1}\left(1+\frac{\frac{\widetilde{s}_2-z\widetilde{s}_1}{\widetilde{s}_2-\widetilde{s}_1}}{\frac{3z}{2\theta}+\frac{\widetilde{s}_2-z\widetilde{s}_1}{\widetilde{s}_2-\widetilde{s}_1}}\right)}{\frac{z}{\theta}\left(3+\frac{\frac{\widetilde{s}_2-z\widetilde{s}_1}{\widetilde{s}_2-\widetilde{s}_1}}{\frac{3z}{2\theta}+\frac{\widetilde{s}_2-z\widetilde{s}_1}{\widetilde{s}_2-\widetilde{s}_1}}\right)}$$

so that

$$P_H^I=\frac{1+\frac{(z-1)\widetilde{s}_1\widetilde{s}_2}{\widetilde{s}_2-\widetilde{s}_1}\left(1+\frac{\frac{\widetilde{s}_2-z\widetilde{s}_1}{\widetilde{s}_2-\widetilde{s}_1}}{\frac{3z}{2\theta}+\frac{\widetilde{s}_2-z\widetilde{s}_1}{\widetilde{s}_2-\widetilde{s}_1}}\right)}{\frac{z}{\theta}\left(3+\frac{\frac{\widetilde{s}_2-z\widetilde{s}_1}{\widetilde{s}_2-\widetilde{s}_1}}{\frac{3z}{2\theta}+\frac{\widetilde{s}_2-z\widetilde{s}_1}{\widetilde{s}_2-\widetilde{s}_1}}\right)}\left(1+\frac{3z}{2\theta}\frac{\widetilde{s}_2-\widetilde{s}_1}{\widetilde{s}_2-z\widetilde{s}_1}\right)-\frac{(z-1)\widetilde{s}_1\widetilde{s}_2}{\widetilde{s}_2-z\widetilde{s}_1}.$$

For the parameter values we considered before, where $\theta=2/5$, $\widetilde{s}_1=3/10$ and $\widetilde{s}_2=3/5$, this results in

$$P_L^I=\frac{1+\frac{.18(z-1)}{.3}\left(1+\frac{6-3z}{\frac{9z}{0.8}+6-3z}\right)}{\frac{5z}{2}\left(3+\frac{6-3z}{\frac{9z}{0.8}+6-3z}\right)}$$

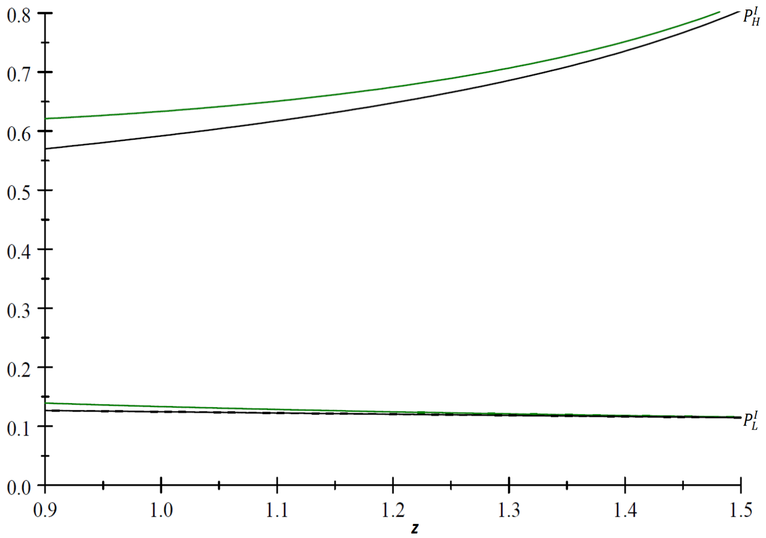
and

$$P_H^I=\frac{1+\frac{.18(z-1)}{.3}\left(1+\frac{6-3z}{\frac{9z}{0.8}+6-3z}\right)}{\frac{5z}{2}\left(3+\frac{6-3z}{\frac{9z}{0.8}+6-3z}\right)}\left(1+\frac{.9z}{0.8(.6-.3z)}\right)-\frac{.18(z-1)}{.6-.3z}.$$

Figure F1 plots these prices under sequential price setting as a function of z together with the corresponding prices for the simultaneous move game analyzed in the main text. The figure shows that the two different analyses (simultaneous versus sequential choice of offline and online prices) show that equilibrium outcomes are very close to each other. The reason is twofold. First, as indicated above, for given and identical P_H^I, the online market is governed by the same incentives and F.O.C.s. Second, if in the sequential setting the incumbent wants to increase its baseline tariff compared to the simultaneous choice setting, the incumbent not only gains because all prices will increase, but also loses as more consumers will switch to the entrant instead of buying from the online incumbent's price. These opposing forces are such that the net effect is that the baseline price is almost identical in the two cases.

The figure shows that if online prices react to the baseline price, the same pattern with respect to changes in z emerges, namely that if z increases (and therefore, more consumers search online), online prices decrease, while the incumbent's baseline price increases. Thus, price discrimination between loyal and searching consumers increases and online price disper-

Figure F1: Price patterns: simultaneous versus sequential game



Notes: The figure predicts price changes under sequential price setting (green) and simultaneous price setting (black) as a function of z with $\widetilde{s}_2 = 3/5$ and $\widetilde{s}_1 = 1/5$ and $\theta = 2/5$. Since the entrants' online tariffs (P^E) are half of the incumbents' cheaper online tariffs (P_L^I), P^E is not shown for better clarity.

sion decreases.

F Additional Figures and Tables

Figure F1: Price zones of "Envia Mitteldeutsche Energie GmbH"

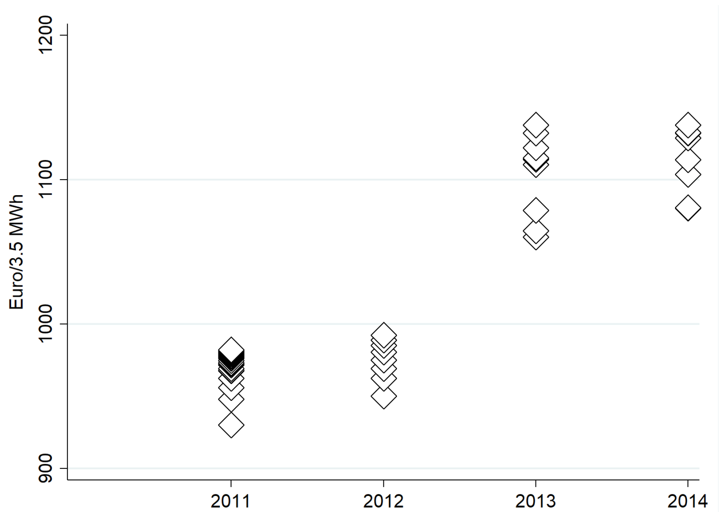
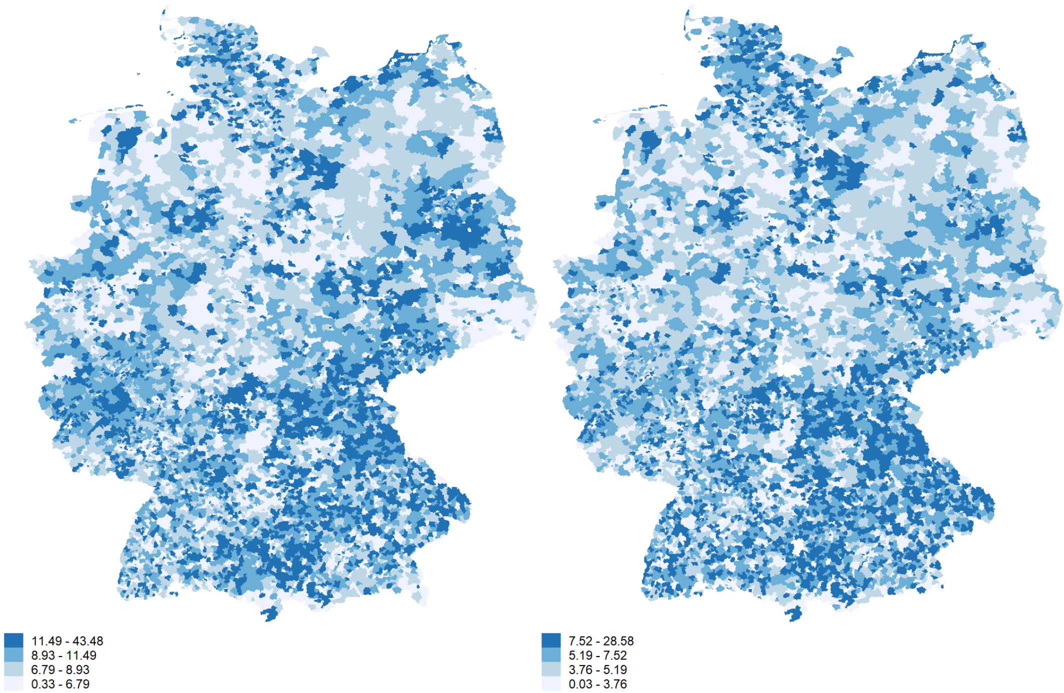
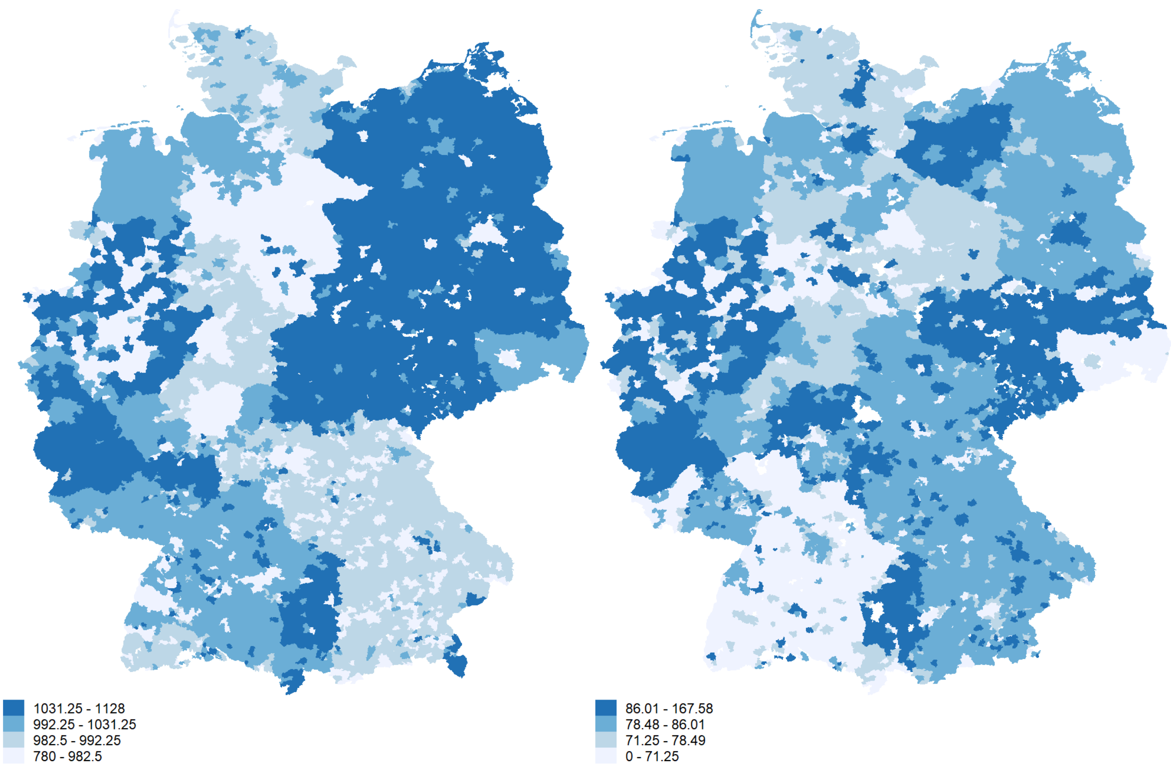


Figure F2: Between and within variation of consumer search intensity



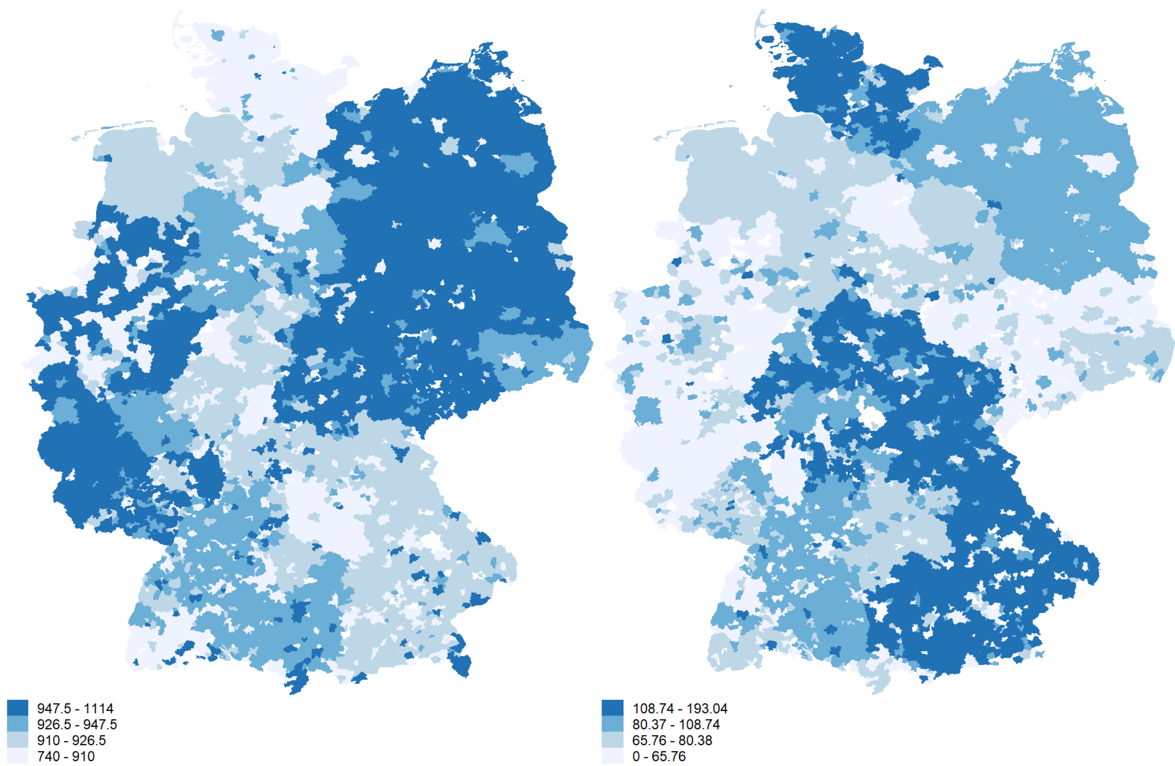
Notes: The left panel presents the between variation in consumer search intensity computed as the average search intensity per zip code during our observation period. The right panel presents the within variation per zip code computed as the standard deviation per zip code during our observation period.

Figure F3: Between and within variation of incumbents' base prices



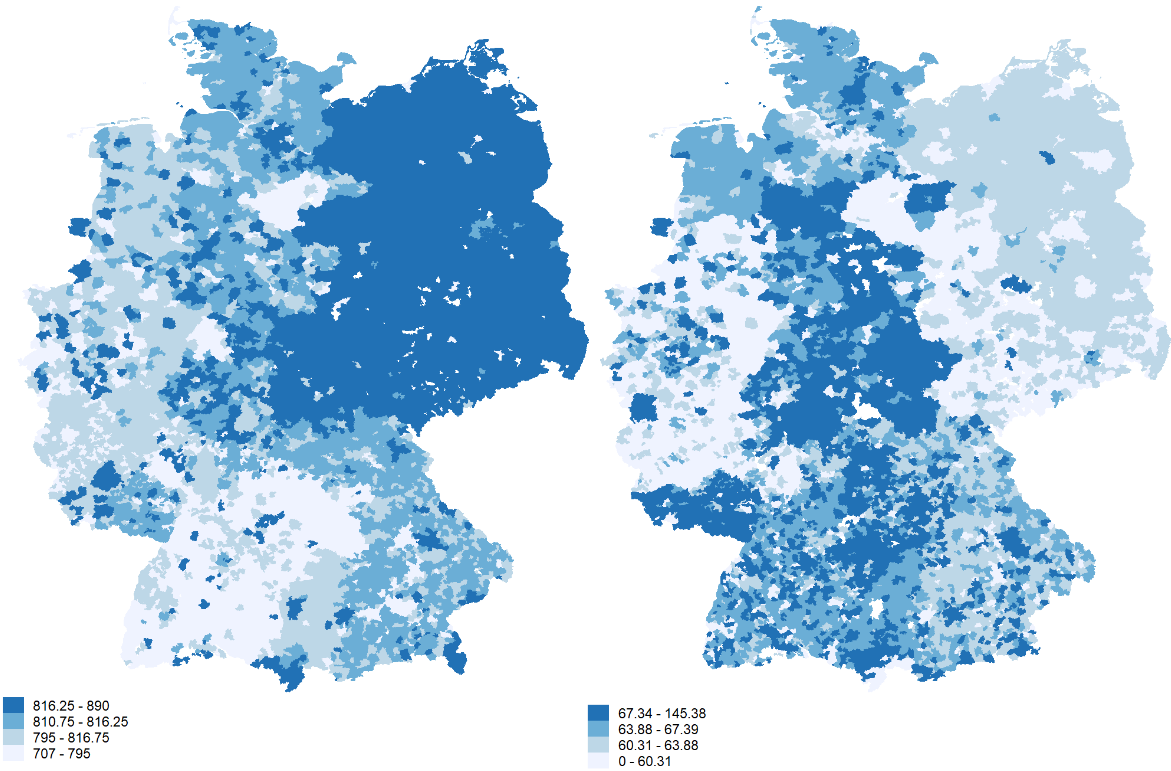
Notes: The left panel presents the between variation in incumbent base prices computed as the average incumbent base price per zip code during our observation period. The right panel presents the within variation per zip code computed as the standard deviation per zip code during our observation period.

Figure F4: Between and within variation of incumbents' cheapest prices



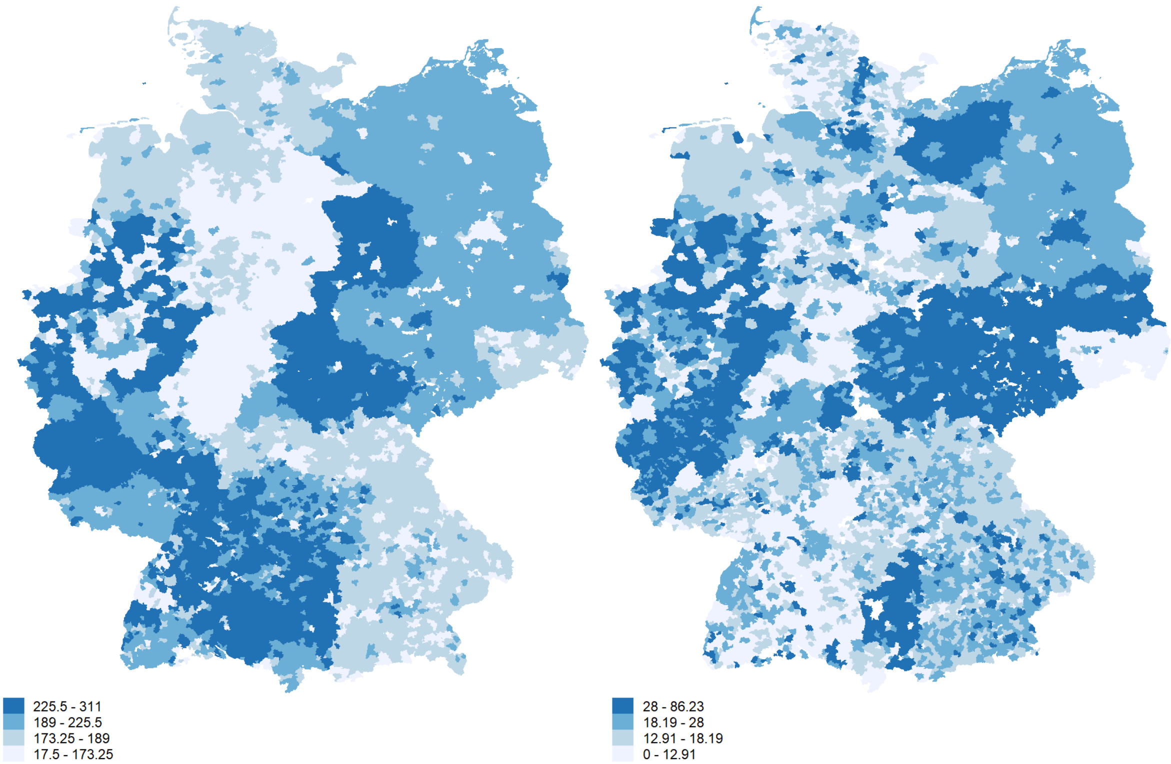
Notes: The left panel presents the between variation in incumbents' cheapest prices computed as the average incumbents' cheapest price per zip code during our observation period. The right panel presents the within variation per zip code computed as the standard deviation per zip code during our observation period.

Figure F5: Between and within variation of the overall cheapest prices



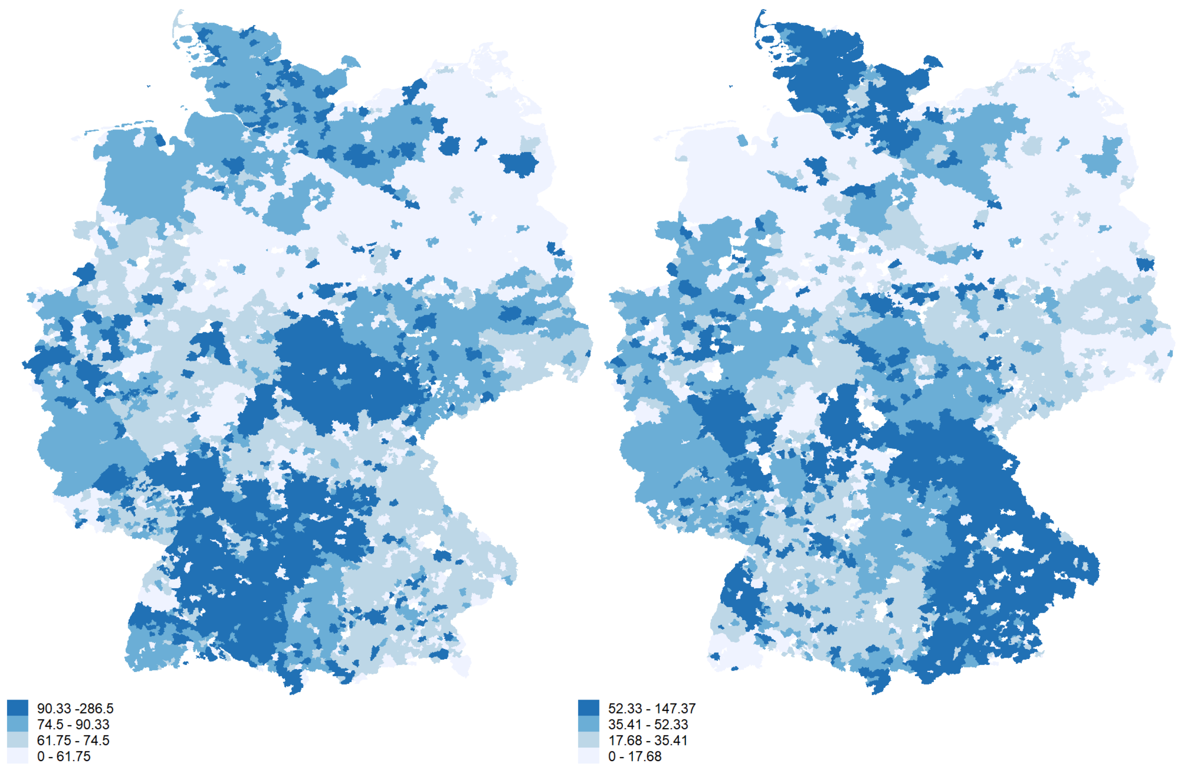
Notes: The left panel presents the between variation in the overall cheapest prices computed as the average cheapest price per zip code during our observation period. The right panel presents the within variation per zip code computed as the standard deviation per zip code during our observation period.

Figure F6: Between and within variation of price dispersion



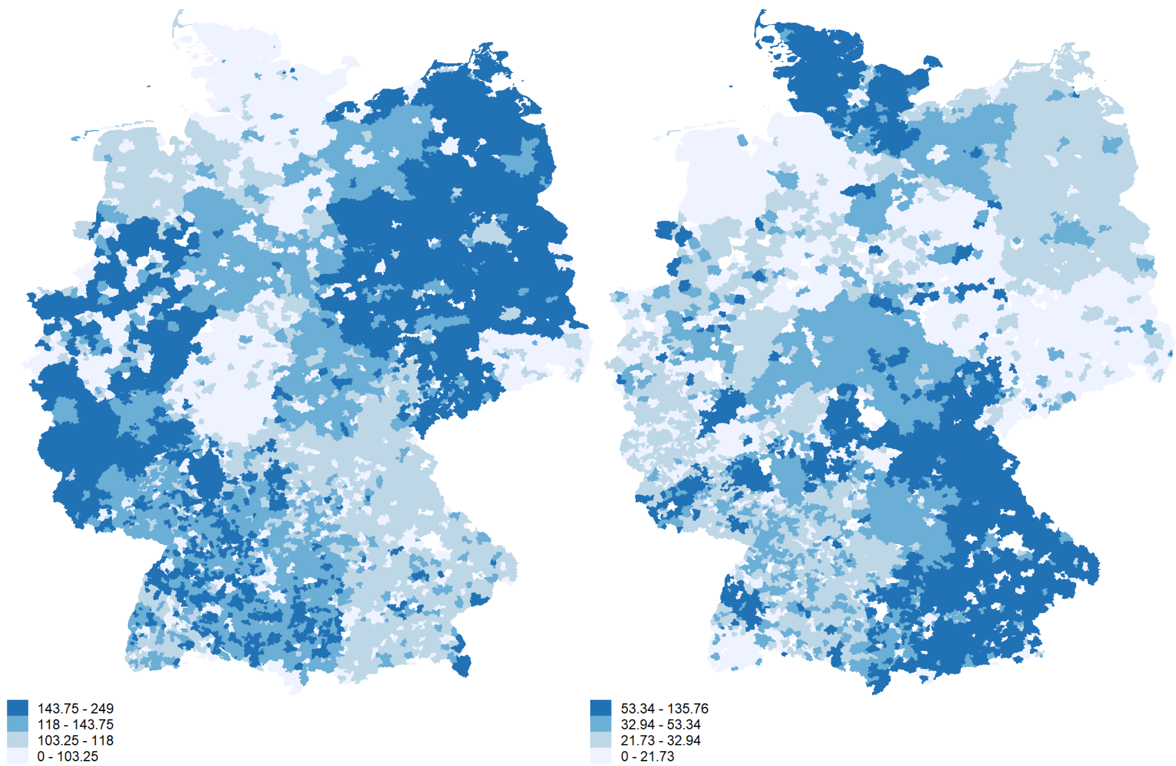
Notes: The left panel presents the between variation in the price dispersion computed as the average price dispersion per zip code during our observation period. The right panel presents the within variation per zip code computed as the standard deviation per zip code during our observation period.

Figure F7: Between and within variation of price discrimination



Notes: The left panel presents the between variation in the price discrimination computed as the average price discrimination per zip code during our observation period. The right panel presents the within variation per zip code computed as the standard deviation per zip code during our observation period.

Figure F8: Between and within variation of online price dispersion



Notes: The left panel presents the between variation in the online price dispersion computed as the average online price dispersion per zip code during our observation period. The right panel presents the within variation per zip code computed as the standard deviation per zip code during our observation period.

G Robustness

Estimation without covariates — All results stay robust when we drop all covariates, as shown in Table G1. That is, the instrumental-variables regression counteracts an omitted-variables bias.

Alternative Instruments I — Our two classes of instruments, broadband availability (BBA) and the share of young households (U40) in a zip code, arguably fulfill the exclusion restriction conditional on the included explanatory variables, since one may argue that these variables directly affect search intensity but do not directly affect prices – conditional on our control variables. To further increase the credibility of our identification strategy, we also apply Hausman-type instruments, as inspired by Hausman (1996) (see also Berry and Haile, 2015; Hausman et al., 1994; Nevo, 2000). Thus, we take averages of BBA and U40 in the 50 *surrounding* zip codes as instruments for search intensity in the focal zip code. In addition, we only include those surrounding zip codes when their prices differ from those in the focal zip code. The logic behind the validity of these instruments is that variation in these two sets of instruments proxy for changes in search intensity unrelated to electricity prices in the focal zip code. For example, if their variation is geographically correlated – so that broadband expansion and/or the age distribution are correlated across zip codes – the instruments are correlated with search intensity in the focal zip code.[2] The key assumption on excludability is that the search intensity in the focal zip code is (mean) independent of the error terms of the price equation conditional on the exogenous control variables. This would only fail if the price shocks in the focal zip code are correlated with broadband expansion or the age distribution in *surrounding* zip codes, which we deem highly unlikely.[3] Also, there is sufficient correlation between our Hausman-type instruments and our main instruments: 0.55 for BBA and 0.60 for U40. We find that the results stay robust to these alternative instruments, as shown in Table G2. Figure G1 compares the estimates of the baseline IV with the Hausman-type estimates and shows that there is significant overlap in the confidence intervals.

Alternative Instruments II — It may be that the IV results are largely driven by one of the instruments, either the broadband internet availability or the age structure. In a robustness exercise we employ either only broadband internet or age based instruments in the estimations. The results stay qualitatively robust and are reported in Table G3.

Level-Level Estimation — Our results remain also fully robust if we run level-level (instead of log-log) specifications of the models, as shown in Table G4.

[2]There is ample reason to believe that this is the case, e.g. firms expand their broadband network across many zip codes at once due to economies of scale, scope, and density of broadband expansion.

[3]We call these instruments "Hausman-type" and not "Hausman" instruments, because there is a difference in application which further strengthens our identification strategy. "Hausman instruments" in our context would be, for example, the average search intensity in surrounding ZIP codes. In this case, one may argue that – if there are problems of reverse causality, measurement error or omitted variables – that they also manifest themselves in the geographically related variables. For example, if an electricity retailer sets prices for several ZIP codes simultaneously, and if there is reverse causality, so that consumers change their search behavior as a reaction to a price change, the two error terms in the first stage and the outcome equation may be correlated and "Hausman" instruments may not fulfill the exclusion restriction. Our requirements are much more modest, since it is highly unlikely that a price change in the focal ZIP code change broadband expansion or the age distribution in surrounding ZIP codes.

Non-Linear Effect of Search on Prices — We also allow for a non-linear relationship between search and prices, by adding a μ^2 in Equation 4. We instrument for μ^2 by using the square of the first-stage estimate of μ from Equation 5 as the instrument for μ^2 (see Wooldridge (2010, p. 262) on this approach). The results remain robust and are reported in Table G5 in the Online Appendix.

Alternative Outcome Variable — We also estimate models using *Lerner Indices* as the dependent variables (Table G6). They are computed as the ratio of markups (i.e.the differences between (net) prices and costs) to prices. The results when using Lerner Indices as the dependent variables are as one would expect from the results of the price estimations.

Additionally Instrumenting for the Number of Competitors — The prices and the number of competitors may affect each other. In order to test if potential endogeneity of the number of competitors affects our results, we also instrument for the number of firms in a zip code by using the average number of firms in the surrounding 50 zip codes in the spirit of Hausman (1996) (for further details, see the discussion in Subsection "Alternative Instruments"). There is sufficient variation regarding the number of firms in the zip code itself and the surrounding 50 zip codes: we observe an absolute difference between the former and the latter of 7 and also a standard deviation of 7. The results only change marginally, as shown in Table G7 suggesting robustness of our results to instrumenting for the number of competitors.

Alternative Clustering of Standard Errors — Many incumbents operate only locally and 46% of the incumbents only have a single zip code in their incumbency area. These small incumbents are mostly municipal utilities. However, larger incumbents often have several zip codes in their incumbency area and charge locally differing baseline tariffs. The different *price zones* are not necessarily at the zip code level (see Section 2 for more details). Hence, as a robustness check, we cluster standard errors in two alternative ways. In the first version, we allow the residuals to correlate within a prices zone and cluster standard errors at that level, instead of the zip code level (see Table G8). In a second version, the clustering is at the incumbency area level (see Table G9). In both cases the results remain fully robust.

Table G1: IV estimates of the impact of consumer search on prices (log-log) – estimations without covariates

	ln Incumbent Base (P^I_H) (1)	ln Incumbent Cheapest (P^I_L) (2)	ln Overall Cheapest (P^E) (3)
ln Search (μ)	0.0431*** (0.0087)	-0.1151*** (0.0261)	-0.0296*** (0.0069)
Year FE	Yes	Yes	Yes
Zip code FE	Yes	Yes	Yes
First stage effective F stat.	26.37	26.37	26.37
Obs.	24,175	24,175	24,175

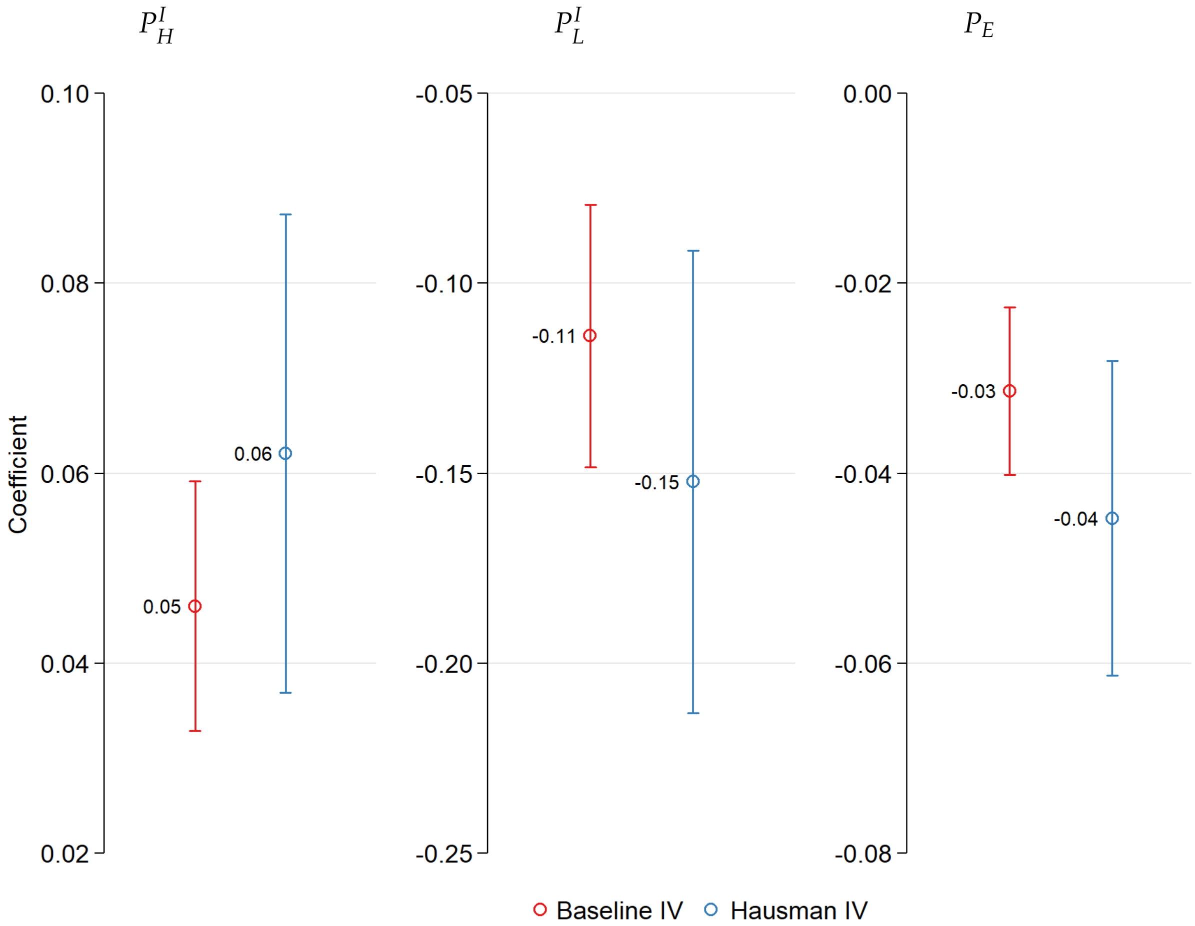
Notes: Standard errors clustered at the zip code level in parentheses. Estimation is by GMM. Instrumented for μ by *U40* and *BBA*. $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.

Table G2: IV estimates of the impact of consumer search on prices (log-log) – Hausman-type instruments for search

	ln Incumbent Base (P_H^I) (1)	ln Incumbent Cheapest (P_L^I) (2)	ln Overall Cheapest (P^E) (3)
ln Search (μ)	0.0621***	-0.1522***	-0.0447***
	(0.0153)	(0.0370)	(0.0101)
No. competitors	0.0003***	0.0049***	0.0003***
	(0.0001)	(0.0002)	(0.0001)
No. households	0.0052***	-0.0065	-0.0032***
	(0.0019)	(0.0044)	(0.0012)
ln Costs	0.2054***	0.4770***	0.5221***
	(0.0142)	(0.0316)	(0.0103)
Average HH size	0.0390***	0.0124	-0.0091*
	(0.0082)	(0.0203)	(0.0054)
ln Purchase Power	0.0067	-0.0459***	-0.0116***
	(0.0075)	(0.0144)	(0.0034)
Income <25k €/a	-0.0002	-0.0015***	-0.0002***
	(0.0001)	(0.0003)	(0.0001)
Income 25-50k €/year	-0.0004**	-0.0008**	-0.0001
	(0.0002)	(0.0004)	(0.0001)
Lower class social status	0.0001***	-0.0000	-0.0001***
	(0.0000)	(0.0000)	(0.0000)
Middle class social status	0.0001***	0.0001**	0.0000
	(0.0000)	(0.0000)	(0.0000)
High Urbanization	0.0216***	-0.0672***	-0.0162***
	(0.0057)	(0.0135)	(0.0039)
Unemployed	0.0204***	0.0523***	0.0068**
	(0.0053)	(0.0125)	(0.0034)
Population density	0.0004**	-0.0002	-0.0000
	(0.0001)	(0.0004)	(0.0001)
Social insurance	0.0012***	-0.0002	-0.0002
	(0.0002)	(0.0005)	(0.0002)
ln Lagged switching rate	0.0044***	0.0285***	0.0014**
	(0.0009)	(0.0024)	(0.0006)
Year FE	Yes	Yes	Yes
Zip code FE	Yes	Yes	Yes
First stage effective F stat.	12.37	12.37	12.37
Obs.	24,175	24,175	24,175

Notes: Standard errors clustered at the zip code level in parentheses. Estimation is by GMM. Instrumented for μ by *U40* and *BBA* in the 50 surrounding zip codes. $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$. $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.

Figure G1: Comparison of estimates from baseline IV and Hausman-type IV estimations



Notes: the red circles represent point estimates from our baseline IV regressions while the blue circles those from the Hausman-type IV regressions. Red and blue vertical lines represent the corresponding 90% CIs.

Table G3: Estimates of the impact of consumer search on prices (log-log) – instruments either only based on broadband availability or on age characteristics

	ln Incumbent base P_H^I		ln Incumbent cheapest P_L^I		ln Overall cheapest P^E	
	Age (1)	Internet (2)	Age (3)	Internet (4)	Age (5)	Internet (6)
ln Search (μ)	0.0478***	0.0598***	-0.1044***	-0.1533***	-0.0333***	-0.0149**
	(0.0086)	(0.0121)	(0.0227)	(0.0340)	(0.0062)	(0.0066)
No. competitors	0.0004***	0.0004***	0.0046***	0.0050***	0.0002***	0.0001***
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0000)	(0.0000)
No. households	0.0037***	0.0046***	-0.0028	-0.0072*	-0.0022**	-0.0014
	(0.0014)	(0.0016)	(0.0035)	(0.0043)	(0.0010)	(0.0009)
ln Costs	0.2106***	0.2235***	0.4560***	0.5079***	0.5242***	0.5317***
	(0.0114)	(0.0130)	(0.0264)	(0.0324)	(0.0093)	(0.0085)
Average HH size	0.0407***	0.0382***	0.0156	0.0194	-0.0064	-0.0127***
	(0.0070)	(0.0083)	(0.0183)	(0.0225)	(0.0053)	(0.0045)
ln Purchase Power	0.0047	0.0103	-0.0356***	-0.0450***	-0.0099***	-0.0055**
	(0.0064)	(0.0079)	(0.0128)	(0.0165)	(0.0033)	(0.0028)
Income <25k €/a	-0.0001	-0.0002	-0.0016***	-0.0014***	-0.0002***	-0.0003***
	(0.0001)	(0.0001)	(0.0003)	(0.0003)	(0.0001)	(0.0001)
Income 25-50k €/year	-0.0003**	-0.0004**	-0.0009**	-0.0006	-0.0001	-0.0002*
	(0.0002)	(0.0002)	(0.0003)	(0.0004)	(0.0001)	(0.0001)
Lower class social status	0.0001***	0.0001***	-0.0000	-0.0000	-0.0001***	-0.0001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Middle class social status	0.0001***	0.0001***	0.0001***	0.0001**	0.0000*	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
High Urbanization	0.0185***	0.0221***	-0.0559***	-0.0663***	-0.0131***	-0.0079**
	(0.0047)	(0.0054)	(0.0111)	(0.0135)	(0.0035)	(0.0031)
Unemployed	0.0181***	0.0205***	0.0633***	0.0476***	0.0094***	0.0124***
	(0.0044)	(0.0051)	(0.0101)	(0.0127)	(0.0030)	(0.0028)
Population density	0.0003***	0.0003**	-0.0001	-0.0002	0.0000	0.0001
	(0.0001)	(0.0001)	(0.0004)	(0.0004)	(0.0001)	(0.0001)
Social insurance	0.0011***	0.0011***	0.0001	-0.0002	0.0000	-0.0000
	(0.0002)	(0.0002)	(0.0005)	(0.0006)	(0.0001)	(0.0001)
ln Lagged switching rate	0.0049***	0.0046***	0.0275***	0.0297***	0.0010*	0.0004
	(0.0007)	(0.0009)	(0.0021)	(0.0026)	(0.0006)	(0.0005)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Zip code FE	Yes	Yes	Yes	Yes	Yes	Yes
First stage effective F stat.	34.88	10.27	34.88	10.27	34.88	10.27
Obs.	24,175	23,910	24,175	23,910	24,175	23,910

Notes: Standard errors clustered at the zip code level in parentheses. Instruments for μ are either the local coverage of different broadband internet speeds (16mb/s, 6mb/s as well as mobile internet availaibility of 6mb/s and 1mb/s) or the share of young households (household head younger than 40) and middle aged households (household head between 40 and 60 years old). $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.

Table G4: IV estimates of the impact of consumer search on prices (level-level)

	Incumbent Base (P_H^I)	Incumbent Cheapest (P_L^I)	Overall Cheapest (P^E)
	(1)	(2)	(3)
Search (μ)	2.7979***	-9.4572***	-1.7719***
	(0.7256)	(1.9140)	(0.4170)
No. competitors	0.4122***	4.3446***	0.2551***
	(0.0748)	(0.1907)	(0.0423)
No. households	2.4428	-12.4749***	-2.9812***
	(1.6704)	(4.3756)	(0.9868)
ln Costs	187.3952***	551.0023***	433.6734***
	(16.5061)	(41.9002)	(10.5287)
Average HH size	36.9131***	-43.1948**	-16.9805***
	(6.6625)	(17.7897)	(4.2414)
ln Purchase Power	-1.7343	-15.7611	-4.2896*
	(5.4960)	(13.5147)	(2.5991)
Income <25k €/a	0.2528**	-2.3856***	-0.3234***
	(0.1162)	(0.3046)	(0.0726)
Income 25-50k €/year	0.0833	-1.6828***	-0.2409***
	(0.1342)	(0.3552)	(0.0867)
Lower class social status	0.1205***	-0.0629	-0.0518***
	(0.0149)	(0.0397)	(0.0089)
Middle class social status	0.0829***	0.1117***	0.0155**
	(0.0123)	(0.0347)	(0.0077)
High Urbanization	23.4521***	-83.8896***	-14.2943***
	(6.0162)	(16.2818)	(3.8276)
Unemployed	-1.4004	90.4368***	11.8443***
	(4.1647)	(9.5017)	(2.3258)
Population density	0.3171***	-0.0360	0.0339
	(0.1060)	(0.3881)	(0.0786)
Social insurance	0.8248***	-0.3585	-0.0963
	(0.1891)	(0.4753)	(0.1119)
ln Lagged switching rate	3.8982***	28.8753***	1.5850***
	(0.9072)	(2.7828)	(0.5714)
Year FE	Yes	Yes	Yes
Zip code FE	Yes	Yes	Yes
First stage effective F stat.	12.04	12.04	12.04
Obs.	24,175	24,175	24,175

Notes: Standard errors clustered at the zip code level in parentheses. Estimation is by GMM. Instrumented for μ by *U40* and *BBA*. $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.

Table G5: IV estimates of a non-linear impact of consumer search on prices (log-log)

	ln Incumbent Base (P_H^I) (1)	ln Incumbent Cheapest (P_L^I) (2)	ln Overall Cheapest (P^E) (3)
ln Search (μ)	0.0389***	-0.1534***	-0.0373***
	(0.0085)	(0.0224)	(0.0056)
$Search^2$	0.0000	0.0134***	0.0032***
	(0.0004)	(0.0010)	(0.0003)
No. competitors	0.0004***	0.0044***	0.0001***
	(0.0001)	(0.0002)	(0.0000)
No. households	0.0030**	0.0016	-0.0007
	(0.0013)	(0.0032)	(0.0008)
ln Costs	0.2172***	0.4059***	0.5030***
	(0.0105)	(0.0242)	(0.0085)
Average HH size	0.0418***	0.0272	-0.0075
	(0.0066)	(0.0181)	(0.0048)
ln Purchase Power	0.0051	-0.0447***	-0.0099***
	(0.0060)	(0.0118)	(0.0028)
Income <25k €/a	-0.0001	-0.0011***	-0.0001*
	(0.0001)	(0.0003)	(0.0001)
Income 25-50k €/year	-0.0003**	-0.0005	-0.0001
	(0.0001)	(0.0003)	(0.0001)
Lower class social status	0.0001***	-0.0000	-0.0001***
	(0.0000)	(0.0000)	(0.0000)
Middle class social status	0.0001***	0.0001***	0.0000*
	(0.0000)	(0.0000)	(0.0000)
High Urbanization	0.0163***	-0.0444***	-0.0082***
	(0.0044)	(0.0100)	(0.0030)
Unemployed	0.0155***	0.0472***	0.0067**
	(0.0044)	(0.0100)	(0.0028)
Population density	0.0003**	-0.0001	0.0000
	(0.0001)	(0.0003)	(0.0001)
Social insurance	0.0010***	0.0002	0.0000
	(0.0002)	(0.0005)	(0.0001)
ln Lagged switching rate	0.0052***	0.0263***	0.0004
	(0.0007)	(0.0019)	(0.0005)
Year FE	Yes	Yes	Yes
Zip code FE	Yes	Yes	Yes
Kleibergen-Paap F-stat.	28.29	28.29	28.29
Durbin-Wu-Hausman test	0.00	0.00	0.00
Obs.	24,175	24,175	24,175

Notes: Standard errors clustered at the zip code level in parentheses. Estimation is by GMM. Instrumented for μ by *U40* and *BBA* and for μ^2 by $\hat{\mu^2}$ – the square of the first stage predictions of μ. $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.

Table G6: IV estimates of the impact of consumer search on Lerner Indices (LI)

	LI Incumbent Base (1)	LI Incumbent Cheapest (2)	LI Overall Cheapest (3)
ln Search (μ)	0.0359***	-0.0925***	-0.0278***
	(0.0058)	(0.0164)	(0.0048)
No. competitors	0.0003***	0.0034***	0.0002***
	(0.0000)	(0.0001)	(0.0000)
No. households	0.0032***	-0.0033	-0.0023***
	(0.0010)	(0.0026)	(0.0008)
ln Costs	-0.5378***	-0.4016***	-0.3991***
	(0.0080)	(0.0209)	(0.0081)
Average HH size	0.0296***	0.0114	-0.0127***
	(0.0050)	(0.0143)	(0.0045)
ln Purchase Power	0.0051	-0.0290***	-0.0086***
	(0.0046)	(0.0097)	(0.0028)
Income <25k €/a	-0.0001	-0.0011***	-0.0001**
	(0.0001)	(0.0002)	(0.0001)
Income 25-50k €/year	-0.0002**	-0.0006**	-0.0001
	(0.0001)	(0.0003)	(0.0001)
Lower class social status	0.0001***	-0.0000	-0.0001***
	(0.0000)	(0.0000)	(0.0000)
Middle class social status	0.0001***	0.0001***	0.0000
	(0.0000)	(0.0000)	(0.0000)
High Urbanization	0.0137***	-0.0476***	-0.0113***
	(0.0034)	(0.0088)	(0.0030)
Unemployed	0.0102***	0.0411***	0.0038
	(0.0032)	(0.0076)	(0.0026)
Population density	0.0002**	-0.0001	0.0000
	(0.0001)	(0.0003)	(0.0001)
Social insurance	0.0007***	-0.0000	-0.0001
	(0.0001)	(0.0004)	(0.0001)
ln Lagged switching rate	0.0037***	0.0230***	0.0009*
	(0.0005)	(0.0016)	(0.0005)
Year FE	Yes	Yes	Yes
Zip code FE	Yes	Yes	Yes
First stage effective F stat.	33.96	33.96	33.96
Obs.	24,175	24,175	24,175

Notes: Standard errors clustered at the zip code level in parentheses. Estimation is by GMM. Instrumented for μ by *U40* and *BBA*. $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.

Table G7: IV estimates of the impact of consumer search on prices (log-log) – additionally instrumented for #Competitors

	ln Incumbent Base (P_H^I) (1)	ln Incumbent Cheapest (P_L^I) (2)	ln Overall Cheapest (P^E) (3)
ln Search (μ)	0.0414***	-0.1265***	-0.0309***
	(0.0079)	(0.0224)	(0.0054)
No. competitors	0.0010***	0.0065***	0.0001**
	(0.0001)	(0.0002)	(0.0001)
No. households	0.0040***	-0.0021	-0.0022**
	(0.0013)	(0.0036)	(0.0009)
ln Costs	0.2319***	0.5129***	0.5146***
	(0.0115)	(0.0291)	(0.0091)
Average HH size	0.0354***	-0.0014	-0.0101**
	(0.0068)	(0.0194)	(0.0050)
ln Purchase Power	-0.0004	-0.0586***	-0.0085***
	(0.0065)	(0.0138)	(0.0032)
Income <25k €/a	-0.0000	-0.0012***	-0.0002***
	(0.0001)	(0.0003)	(0.0001)
Income 25-50k €/year	-0.0003*	-0.0006*	-0.0001
	(0.0001)	(0.0004)	(0.0001)
Lower class social status	0.0001***	-0.0001***	-0.0001***
	(0.0000)	(0.0000)	(0.0000)
Middle class social status	0.0001***	0.0001**	0.0000**
	(0.0000)	(0.0000)	(0.0000)
High Urbanization	0.0190***	-0.0545***	-0.0125***
	(0.0045)	(0.0114)	(0.0033)
Unemployed	0.0130***	0.0479***	0.0088***
	(0.0043)	(0.0107)	(0.0029)
Population density	0.0003**	-0.0001	0.0000
	(0.0001)	(0.0004)	(0.0001)
Social insurance	0.0010***	-0.0001	-0.0000
	(0.0002)	(0.0005)	(0.0001)
ln Lagged switching rate	0.0058***	0.0302***	0.0009
	(0.0007)	(0.0022)	(0.0005)
Year FE	Yes	Yes	Yes
Zip code FE	Yes	Yes	Yes
Kleibergen-Paap stat.	24.96	24.96	24.96
Obs.	24,175	24,175	24,175

Notes: Standard errors clustered at the zip code level in parentheses. Estimation is by GMM. Instrumented for μ by *U40* and *BBA*. Instrumented for #Competitors by #Competitors in the 50 surrounding zip codes. $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$. $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.

Table G8: IV estimates of the impact of consumer search on prices (log-log) – alternative clustering of standard errors I

	ln Incumbent Base (P_H^I) (1)	ln Incumbent Cheapest (P_L^I) (2)	ln Overall Cheapest (P^E) (3)
ln Search (μ)	0.0486***	-0.1130***	-0.0315***
	(0.0119)	(0.0271)	(0.0082)
No. competitors	0.0004***	0.0046***	0.0002**
	(0.0001)	(0.0004)	(0.0001)
No. households	0.0037	-0.0033	-0.0022
	(0.0023)	(0.0042)	(0.0015)
ln Costs	0.2099***	0.4559***	0.5168***
	(0.0192)	(0.0450)	(0.0168)
Average HH size	0.0404***	0.0128	-0.0108
	(0.0100)	(0.0236)	(0.0069)
ln Purchase Power	0.0075	-0.0391***	-0.0092**
	(0.0076)	(0.0149)	(0.0043)
Income <25k €/a	-0.0001	-0.0015***	-0.0002**
	(0.0001)	(0.0003)	(0.0001)
Income 25-50k €/year	-0.0004**	-0.0008**	-0.0001
	(0.0002)	(0.0003)	(0.0001)
Lower class social status	0.0001***	-0.0000	-0.0001***
	(0.0000)	(0.0000)	(0.0000)
Middle class social status	0.0001***	0.0001***	0.0000
	(0.0000)	(0.0000)	(0.0000)
High Urbanization	0.0186***	-0.0574***	-0.0124***
	(0.0049)	(0.0125)	(0.0036)
Unemployed	0.0168***	0.0582***	0.0083*
	(0.0056)	(0.0136)	(0.0044)
Population density	0.0003**	-0.0001	0.0000
	(0.0001)	(0.0004)	(0.0001)
Social insurance	0.0010***	0.0000	-0.0001
	(0.0002)	(0.0006)	(0.0002)
ln Lagged switching rate	0.0049***	0.0278***	0.0009
	(0.0009)	(0.0025)	(0.0008)
Year FE	Yes	Yes	Yes
Zip code FE	Yes	Yes	Yes
First stage effective F stat.	42.71	42.71	42.71
Obs.	24,175	24,175	24,175

Notes: Standard errors clustered at the incumbents' price zone level in parentheses. Estimation is by GMM. Instrumented for μ by *U40* and *BBA*. $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.

Table G9: IV estimates of the impact of consumer search on prices (log-log) – alternative clustering of standard errors II

	ln Incumbent Base (P_H^I) (1)	ln Incumbent Cheapest (P_L^I) (2)	ln Overall Cheapest (P^E) (3)
ln Search (μ)	0.0515*	-0.0930**	-0.0322*
	(0.0276)	(0.0396)	(0.0175)
No. competitors	0.0004	0.0036***	0.0002
	(0.0003)	(0.0010)	(0.0003)
No. households	0.0045	-0.0006	-0.0024
	(0.0033)	(0.0084)	(0.0024)
ln Costs	0.2020***	0.4984***	0.5145***
	(0.0663)	(0.1256)	(0.0575)
Average HH size	0.0424	-0.0035	-0.0101
	(0.0254)	(0.1321)	(0.0197)
ln Purchase Power	0.0089	-0.0357	-0.0095
	(0.0099)	(0.0303)	(0.0076)
Income <25k €/a	-0.0001	-0.0012**	-0.0002
	(0.0002)	(0.0006)	(0.0002)
Income 25-50k €/year	-0.0004	-0.0007	-0.0001
	(0.0002)	(0.0005)	(0.0002)
Lower class social status	0.0001*	-0.0000	-0.0001***
	(0.0000)	(0.0001)	(0.0000)
Middle class social status	0.0001***	0.0001	0.0000
	(0.0000)	(0.0001)	(0.0000)
High Urbanization	0.0195**	-0.0503***	-0.0127**
	(0.0080)	(0.0126)	(0.0053)
Unemployed	0.0179**	0.0449	0.0085
	(0.0088)	(0.0337)	(0.0087)
Population density	0.0003**	0.0001	0.0000
	(0.0001)	(0.0007)	(0.0001)
Social insurance	0.0011***	-0.0001	-0.0000
	(0.0004)	(0.0016)	(0.0003)
ln Lagged switching rate	0.0056***	0.0262***	0.0009
	(0.0017)	(0.0077)	(0.0016)
Year FE	Yes	Yes	Yes
Zip code FE	Yes	Yes	Yes
First stage effective F stat.	10.97	10.97	10.97
Obs.	24,175	24,175	24,175

Notes: Standard errors clustered at the incumbency area level in parentheses. Estimation is by GMM. Instrumented for μ by *U40* and *BBA*. $^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$.