# DISCUSSION PAPER SERIES

DP15707

## Gains from Convenience and the Value of E-commerce

Yufeng Huang and Bart Bronnenberg

**INDUSTRIAL ORGANIZATION**

ISSN 0265-8003

# Gains from Convenience and the Value of E-commerce

*Yufeng Huang and Bart Bronnenberg*

Discussion Paper DP15707
Published 24 January 2021
Submitted 07 January 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

# Gains from Convenience and the Value of E-commerce

## Abstract

Why do consumers value shopping online? We decompose the value of e-commerce to individual consumers and highlight the role of convenience, i.e., the avoidance of transportation costs. We complement household purchase panel data with precise locations of consumers and stores, and show that travel distance is a strong driver of consumer store choice and the substitution to the online channel. Using a structural model of retailer and channel choice, we report that during 2016-2018 the total value from e-commerce to consumers is equivalent to a 23% discount on all prices. Of this value, a quarter comes from convenience in the form of lower transportation costs, a quarter from intensified price competition, and the remaining half from new online retailers and online channels of existing offline retailers. We further demonstrate that consumer gains are heterogeneous. Consumers far from offline stores or experienced in online shopping will benefit more from e-commerce, whereas consumers who likely do not shop online still benefit indirectly from price competition. Finally, our results show that, as consumers gain more online shopping experience, substantial additional gains from e-commerce are yet to materialize in the future.

JEL Classification: D12, L81, M31

Keywords: E-commerce, retail, convenience, Transportation Costs, online/offline

Yufeng Huang - yufeng.huang@simon.rochester.edu
*University of Rochester*

Bart Bronnenberg - bart.bronnenberg@uvt.nl
*Tilburg University, Dhr and CEPR*

Acknowledgements
The authors thank Bernadette van Ewijk at AIMARK and Alfred Dijs at GfK for their extensive help with the data used in this study. Bronnenberg and Huang are grateful to the Netherlands Foundation for Scientific Research (NWO Vici Grant) for financial support. Huang thanks the Marketing Science Institute for financial support. The authors are grateful for comments and suggestions from Tomomichi Amano, Bryan Bollinger, Kristina Brecko, Paul Ellickson, Mitch Lovett, Adam Smith, Scott Shriver, Takeaki Sunada, and conference and seminar participants at the European Quant Marketing Seminar, London Business School, the New Economic School in Moscow, Tilburg University, the Virtual Quantative Marketing Seminar, the University of Rochester and the 2018 and 2019 Marketing Science Conference. All errors are the authors'.

# Gains from Convenience and the Value of E-commerce*

Yufeng Huang† and Bart J. Bronnenberg‡

September 22, 2020

**Abstract**

Why do consumers value shopping online? We decompose the value of e-commerce to individual consumers and highlight the role of convenience, i.e., the avoidance of transportation costs. We complement household purchase panel data with precise locations of consumers and stores, and show that travel distance is a strong driver of consumer store choice and the substitution to the online channel. Using a structural model of retailer and channel choice, we report that during 2016-2018 the total value from e-commerce to consumers is equivalent to a 23% discount on all prices. Of this value, a quarter comes from convenience in the form of lower transportation costs, a quarter from intensified price competition, and the remaining half from new online retailers and online channels of existing offline retailers. We further demonstrate that consumer gains are heterogeneous. Consumers far from offline stores or experienced in online shopping will benefit more from e-commerce, whereas consumers who likely do not shop online still benefit indirectly from price competition. Finally, our results show that, as consumers gain more online shopping experience, substantial additional gains from e-commerce are yet to materialize in the future.

Keywords: e-commerce, retail, convenience, transportation costs, online/offline.

JEL: D12, L81, M31

*The authors thank Bernadette van Ewijk at AIMARK and Alfred Dijs at GfK for their extensive help with the data used in this study. Bronnenberg and Huang are grateful to the Netherlands Foundation for Scientific Research (NWO Vici Grant) for financial support. Huang thanks the Marketing Science Institute for financial support. The authors are grateful for comments and suggestions from Tomomichi Amano, Bryan Bollinger, Kristina Brecko, Paul Ellickson, Mitch Lovett, Adam Smith, Scott Shriver, Takeaki Sunada, and conference and seminar participants at the European Quant Marketing Seminar, London Business School, the New Economic School in Moscow, Tilburg University, the Virtual Quantative Marketing Seminar, the University of Rochester and the 2018 and 2019 Marketing Science Conference. All errors are the authors'.

†University of Rochester Simon Business School. yufeng.huang@simon.rochester.edu

‡Tilburg School of Economics and Management and CEPR. bart.bronnenberg@tilburguniversity.edu

# 1 Introduction

E-commerce has expanded rapidly across a broad set of geographic markets and product categories. For example, the share of online spending in the Dutch retail apparel market has risen from 5% in 2007 to 27% in 2018. This rapid growth, in line with findings in the recent literature (Bronnenberg and Ellickson, 2015; Hortaçsu and Syverson, 2015), suggests that e-commerce provides considerable value to consumers. However, our knowledge about *in what ways* e-commerce provides value is still limited. Online shoppers might value the enormous product variety offered by new pure-play online retailers, prompting traditional retailers to enlarge their assortments. At the same time, shoppers might value the convenience of shopping without having to travel, prompting retailers to relocate their stores or offer convenient retail services. Further, all consumers –even those who do not shop online– might indirectly benefit from lower prices due to intensified competition. Moreover, these benefits may differ in magnitude and relative importance across consumers.

While the bulk of the empirical literature has focused on consumer gains from getting access to an expanded set of product varieties (Brynjolfsson et al., 2003; Quan and Williams, 2018), much less is known about the value of convenience (Bronnenberg, 2015; Goldfarb and Tucker, 2019), i.e., shopping without transportation costs. This lack of emphasis on transportation costs sharply contrasts with the folk wisdom that “the three most important drivers of success in the retail business are location, location, and location.” We address this lacuna in the literature and show that convenience is an important source of value of e-commerce.

Measuring and decomposing the value of e-commerce poses three challenges. The first challenge is that consumers substitute across chains and between channels. This substitution behavior is crucial for understanding the value of e-commerce (and the part of this value that comes from convenience). Whereas most existing studies on omnichannel retail focus on data from one retailer, we assemble a new individual-level panel dataset in the Dutch retail apparel market that offers complete coverage across all major retail chains. This dataset allows us to construct and estimate a model that quantifies the value of the e-commerce channel, and subsequently, allows us to decompose this value into different mechanisms.

The second challenge is that estimating transportation costs requires precise measures of consumer and store locations. Often, conventional data only provide approximate location measures, e.g., at the US zipcode or county level. In our data, we observe exact store locations. We also observe precise consumer locations in cells with a surface area of 1.3 square kilometers on average or approximately 0.5 square miles (for reference, a US zip code averages about 90 square miles). The granularity of these locations allows us to measure the distance between stores and consumer residences with high accuracy.

The third challenge is the identification of the sensitivity to distance. Whereas store entry and exit decisions create variation in the distance between consumers and chains, these decisions are likely strategic at the level of a market area and endogenous to unobserved demand changes in that market (Li, 2019). To identify the treatment effect of distance on consumer store choices, we exploit that stores can target a geographic area but not each consumer in it. When a new store enters a market, consumers located closer to it face a shorter travel distance to the chain compared to observationally similar consumers who are further away. We demonstrate that the entry of a new store indeed affects consumer chain choices and that the impact depends on the consumer's exact location within the market. We also show that consumers at different locations are similar in observed demographics and display parallel trends in their shopping behavior towards a chain. We show that our identification strategy of consumers' sensitivity to distance is a variant of "spatial difference-in-differences" (Ellickson and Grieco, 2013), which, in our case, leverages granular individual-level panel data.

We start by describing consumer shopping behavior and its sensitivity to distance. We first show considerable growth of the online channel, and that consumer experience with e-commerce plays an important role in this growth. Next, we estimate that consumer store choices are highly sensitive to distance. Moving the closest store of a chain from 0 to 1 kilometer (km) away from the consumer leads to a 9.9% reduction in store-visit incidence. We also show that substitution effects across chains depend on their proximity to the consumer, suggesting that consumer transportation costs make chains compete locally on a spatially-differentiated market. Finally, whereas we find no complementarity between offline and online channels within a chain (cf. Wang and Goldfarb, 2017, Bell et al., 2017, among others), we demonstrate that offline and online channels are substitutes for existing shoppers to a chain.

To quantify the impact of e-commerce, we construct and estimate a structural demand model to measure consumer preferences for shopping at each retail chain and in each channel. In the model, the consumer chooses the chain-channel combination that maximizes her short-term utility. She considers distances to- and prices of each option and is also affected by her past shopping history. In particular, we flexibly characterize consumer learning about online shopping technology. Such learning makes it possible for us to forecast the long-run market outcome when the expansion of e-commerce reaches a steady state. We further include rich observed and unobserved consumer preference heterogeneity, including a set of demographic variables and random coefficients. We estimate our model using simulated maximum likelihood, and we report plausible distance and price sensitivities and the degree of heterogeneity across demographic groups. While the demand model serves as the primary framework for quantifying and decomposing the gains from e-commerce, we also estimate a simple static pricing game to assess the impact of e-commerce on equilibrium prices.

We find that the gain from e-commerce is equivalent to a 23% difference in current prices during 2016-2018, or about €2.6 billion of annual retail value at final prices.[1] Of this total gain, the elimination of transportation costs accounts for 25%, equivalent to a 5.6% decrease in all prices. In addition, the emergence of new online retailers and the online channel of existing retailers benefit consumers significantly, equivalent to a 5.4% and 5.6% discount in prices, respectively. We interpret such benefits as gains from store and product variety (Brynjolfsson et al., 2003; Quan and Williams, 2018), noting that the (pure-play) online retailers also offer a wider selection. We also demonstrate that consumers farther away from the cluster of stores or with more online-shopping experiences derive higher gains from convenience and variety. Finally, the rise of e-commerce reduces local market power, resulting in 5.9% lower prices for an average consumer and yielding a significant indirect benefit for all consumers regardless of whether or not they shop online.

The 2016-2018 market has not yet fully adopted online channels. Leveraging the estimated learning effect, we simulate the long-run stationary market where consumers have fully learned about online-shopping technology. We predict significant additional gains from e-commerce as consumers learn and gain more experience with the new channel.

### 1.1 Contribution and related literature

**Consumer gains from e-commerce.** Our primary contribution to the literature is that we quantify the consumer gains from reducing transportation costs by shopping online – an important source

[1] The annual revenue of the apparel market in the Netherlands is estimated between €10.7-12.5 billion during 2008-2018 (see Footnote 7).

of the gains from e-commerce, which the empirical literature has largely overlooked. Brynjolfsson et al. (2003) and Quan and Williams (2018) study the gains from increased product variety from online retailers. Closely related to our paper, Dolfen et al. (2020) separately estimate consumer gains from online channels of offline retailers and new online chains, and find that the latter contribute to most of the consumer value of e-commerce. Our paper contributes by carefully illustrating the identification and estimation of consumer transportation costs (and the heterogeneity across consumers), which is the central primitive in driving the value of convenience. Our framework also allows us to separately quantify the impact of eliminating consumer transportation costs, price competition due to changes in spatial market power, and, as in Dolfen et al. (2020), the addition of new online channels and new online chains. We find substantial gains from eliminating transportation costs alone,[2] complementing existing works.[3]

**Price competition between online and offline retailers.** As a second contribution to the literature, we extend the question of spatial competition beyond the offline retail context (Smith, 2004, 2006; Houde, 2012; Ellickson et al., 2020) and study the impact of e-commerce on the competitive structure in markets with consumer transportation costs.[4] Specifically, we study how e-commerce reduces frictions from travel and intensifies retail competition, and how the entry of pure-play online retailers affects this competition. The wide coverage of chains and channels in the data enables us to study this question.

Also, our study on the competitive effect of the internet complements the literature on how the internet facilitates price transparency (Cavallo, 2017; Jo et al., 2019).[5] While this literature primarily

[2]Dolfen et al. (2020) refers to the convenience value as the total value of adding the online channel for existing chains. Our measure of this value, under their definition, is half of the total value from e-commerce.

[3]Broadly, our study is also related to valuing different aspects of digitization, where various literatures highlight different mechanisms. Related works include Yoganarasimhan (2013); Bai (2016); Xin (2018) on the value of reputation, Frechette et al. (2019) on reduction of search and matching frictions, and Zervas et al. (2017); Farronato and Fradkin (2018) on flexibility of supply by platform participants.

[4]Smith (2006) estimates consumer demand for supermarkets using individual-level store choice data with granular measures of location, and shows that regulations of store location and size offer little improvement of consumer welfare. Houde (2012) examines the spatial market structure of the Quebec retail gasoline market and demonstrates that carefully modeling consumer commuting routes is crucial to characterizing competition between gasoline chains and simulating the merger effect on the prices they charge. Ellickson et al. (2020) characterize spatial demand of US retailers combining store-level revenue data with detailed demographic data at the census tract level, and show the model produces reasonable diversion ratio estimates and reasonable recommendations for retail merger cases.

[5]Cavallo (2017) documents the uniformity of online and offline prices and suggests that e-commerce facilitates price transparency. Jo et al. (2019) shows that the entry of e-commerce suppresses the rate of price increase for goods sold online and closes the intercity gaps in offline prices of these goods.

focuses on how e-commerce mitigates information friction, our focus is on how the new channel affects spatial friction and the competitive effect therein.

**Substitution or complementarity between online and offline.** Finally, we provide new evidence on the substitution or complementarity of online and offline retail. There is a long-standing debate on whether online and offline purchase options are substitutes (Gentzkow, 2007; Forman et al., 2009; Pozzi, 2013) or complements (Bell et al., 2017; Wang and Goldfarb, 2017; Li, 2019; Shriver and Bollinger, 2020).[6] In particular, Wang and Goldfarb (2017) propose a potential reconciliation of this debate: Offline stores might provide new information about the retailer, but this information spillover effect only exists among new consumers, who are not yet familiar with the chain. Shriver and Bollinger (2020) structurally characterize consumer demand for fashion goods across channels within a retailer. They use the model to quantify the tradeoff between demand expansion effects from the online channel and cannibalization between online and offline channels. Leveraging detailed consumer panel data across many retailers (whereas most existing papers in this literature focus on one retailer), we find robust evidence supporting this tradeoff. Specifically, we demonstrate that online and offline channels are neither *net* substitutes nor *net* complements on average across all consumers, but are substitutes within the set of customers who have previously shopped at the chain.

The rest of the paper is organized as follows. Section 2 presents the context and data, and describes stylized facts about the apparel industry. Section 3 presents identification and descriptive estimates of consumers' sensitivity to distance, and distance-driven substitution patterns between and within chains. Section 4 and 5 presents the model and estimation results of the model. Section 6 shows the simulated welfare impact of e-commerce in counterfactual simulations. Section 7 concludes.

[6] Gentzkow (2007) exploits consumers' accessibility to high-speed internet as an instrument and finds that the online Washington Post substitutes its print counterpart. Forman et al. (2009) show that local brick-and-mortar book retailers are substitutes to online bookstores. Pozzi (2013) shows that a new online grocery channel substitutes the offline channel of the same grocery chain. In contrast, Bell et al. (2017) demonstrate that the opening of a showroom increases the sales near the showroom for an online retailer, and Li (2019) confirms the directional result but shows that part of the complementarity effect comes from a selection bias. Wang and Goldfarb (2017) and Shriver and Bollinger (2020) use data from a US retail chain and show that offline stores can complement online sales.

# 2 Context, data, and descriptive statistics

## 2.1 Data

Our study focuses on the retail apparel industry in the Netherlands – a retail market for clothing, footwear, and accessories. The annual revenue in this industry is estimated between €10.7-12.5 billion during 2008-2018.[7]

Our primary purchase data come from GfK's JURY panel in the Netherlands. The panel covers 2,267,772 purchases of apparel items for 29,284 consumers, covering the period between January 2007 and June 2018. Purchases are recorded using an online diary. For each purchased item, this diary lists the time and location of purchase, type and brand of the product, quantity and expenditure, as well as whether the transaction occurred online. The data set also contains information about household demographics, including residential location measured in Dutch 5-digit zipcode.

We supplement the primary data with two additional data sets. We collect addresses, opening dates, and closing dates for each branch of each chain, from the BvD Orbis database.[8] This data set allows us to measure the store location and entry and exit dates precisely. The second dataset is the Geo Suite database, which provides the mapping between zipcode centroids and coordinates (latitude and longitude).

## 2.2 Sample construction

**Consumer location and the distance measure.** The 6-digit zipcode, such as "5042AB," is the most granular zipcode in the Netherlands. Each 6-digit zipcode contains up to 8 home addresses. Our consumer panel data cover panelists' residential locations at the level of 5-digit zipcodes (e.g. "5042A"). We then assign the panelist to the centroid of the 5-digit zipcode, using the mapping provided by Geo Suite data. We discuss the precision of the 5-digit zipcode in the next section. Similarly, we locate each store at the centroid of its 6-digit zipcode.

[7]Source: Centraal Bureau voor de Statistiek (CBS). We combine two sources of aggregate statistics to arrive at this number. First, aggregate turnover in Year 2013 of SIC sectors 4771 (shops selling clothing) and 4772 (shops for footwear and leather goods) are available from `https://opendata.cbs.nl/statline/#/CBS/en/dataset/81161ENG/table?ts=1578943816612`. Second, turnover growth is available from `https://opendata.cbs.nl/statline/#/CBS/en/dataset/83868Eng/table?dl=219F9`. Note that these reported revenues might exclude online clothing and footwear sales, which is reported in 4791 as online retail for all categories.

[8]https://orbis.bvdinfo.com/

We measure the distance between each panelist and each chain using the distance to the closest store of that chain at any given point in time. To convert location differences (in coordinates) between consumers and stores into distances, we use the standard "great circle" distance formula. For each consumer $i$ at time $t$ and a given store $s$ distance is equal to:

$$D_{ist} = r \cdot \arccos\left(\sin(\text{lat}_{it}) \cdot \sin(\text{lat}_s) + \cos(\text{lat}_{it}) \cdot \cos(\text{lat}_s) \cdot \cos\left(\text{lon}_{it} - \text{lon}_s\right)\right). \tag{1}$$

We use $r = 6,371$ as the radius of the earth to measure distance in kilometers. Given the distance between each consumer and each store, we compute the distance between $i$ and the closest store of each chain

$$D_{ijt} = \min_{s \in j(t)} D_{ist}. \tag{2}$$

Note that the location of a store can be taken as fixed (if a store moves, this is coded as one store closing and another opening). The distance between consumer $i$ and chain $j$ can vary because different sets of stores operate at different points in time, or because of changes in consumer residential locations.

**Price index.** We construct prices indices at the chain-month level to measure overall price level within the chain.[9] Because we only observe the price conditional on purchase, averaging these purchase prices in a given chain-month will lead to a selection problem of *who* purchases *which* product. We circumvent this difficulty by leveraging the fact that we observe *whether* a purchased product is on a price discount, along with its price. Specifically, we assume that, given rich observed and unobserved characteristics of the consumer, discounts are not targeted to unobserved demand shocks.[10] With this assumption, we project prices onto discount dummies to obtain time-varying discount frequency and discount depth, by chain and controlling for consumer and product characteristics. We construct the price index of each chain using the projected discount frequency and depth. Appendix Section A documents the detail of this construction.

We demonstrate in Appendix Section B that, conditional on the same fixed effects as our main

[9] Constructing chain-month level price indices assumes that prices are equal across channels. This assumption has been tested elsewhere. For instance, Cavallo (2017) finds that fashion and clothing have a very high degree of price uniformity across channels, with 92% of fashion products carrying identical prices online and offline within chain.

[10] Thus, this allows for targeting based on the distribution of observed consumer and product characteristics, composition of consumers with different time-invariant heterogeneity, and common time effects such as seasonality.

analysis, whether a transaction contains products on discount is unrelated to the demographics of the consumer or that of the local area. This result is consistent with the assumption that discounts are untargeted to local or consumer-level demand shifters, supporting our construction of the price index.

**Construction of the final sample.** From the full sample of 2,267,772 purchase records, we construct the estimation sample by excluding data with missing chain identities or missing consumer locations, or where the consumer has moved in the sample. In particular, 470,958 observations (21%) have missing retailer identities and 1,893 (0.1%) have missing consumer locations. Further, 2,912 consumers (accounting for 384,030 purchase records) have relocated within the sample. As we will explain and motivate in Section 3, we leverage store entry and exit, instead of consumer relocation, to identify consumer's distance sensitivity. To keep the identifying variation clean, we drop the sample of households that relocate. These sample-selection criteria lead us to a sub-sample of 1,482,298 transaction records, representing a total value of €37.2 million. To analyze our panel data, we collapse purchases to the consumer-chain-channel-month level. To study the role of distance and price on purchase incidence, we complete this data set by including months without purchase incidence. In this expanded dataset, we further focus on consumers who ever at least shopped at 4 chains in the entire sample, excluding about 11% observations. This step leads to the final data set, containing 135,183,166 observations.

## 2.3 Descriptive statistics

This section presents summary statistics that are important for the analysis. First, we demonstrate that the sample of panelists is representative of Dutch households. Next, we demonstrate that our measure of consumer location is granular, revealing precise spatial distribution patterns of where consumers shop. We then present the extent of expansion of online expenditure over time, plus a decomposition of this expansion that suggests that consumer experience and learning are important drivers of e-commerce growth. Finally, we show that store entry and exit is common in our sample period, which is important in identifying consumer transportation costs.

Table 1: Comparison of demographics between the sample and the LISS panel

| | sample: mean | sd | LISS: mean | sd |
|---|---|---|---|---|
| respondent is female | 0.574 | 0.494 | 0.568 | 0.495 |
| age of the respondent | 50.707 | 16.387 | 51.322 | 16.111 |
| education: beyond secondary | 0.390 | 0.488 | 0.400 | 0.490 |
| currently employed | 0.532 | 0.499 | 0.591 | 0.492 |
| monthly household income net of taxes | 2456.142 | 1033.050 | 2702.159 | 1275.969 |
| observations | 80,294 | | 58,172 | |

**Notes:** This table reports mean and standard deviation of demographic variables between the JURY panel (our main sample) and the LISS panel. An observation is a consumer/household-year.

**Representativeness of the sample demographics.** We first describe the distribution of demographic profiles among the panelists. The first column of Table 1 presents the mean and standard deviation of gender, age, education, employment, and monthly income of the panelists. In the next two columns, we contrast these measures to the distribution of a representative household panel, the Longitudinal Internet Studies for the Social sciences (LISS) panel.[11]

We first note that the JURY data contain one respondent per household, and 57% of them are female. In constrast, the LISS panel surveys each household member separately. To make the two samples comparable, we draw "primary respondents" from each household in the LISS panel, such that 57% of them are female.[12] We compare demographics between the two datasets and find that the mean and variance of age and education are very close between the two samples, and the JURY panelists are about 10% less employed and earn 10% lower income. Overall, this comparison suggests that the JURY panel, except for gender, is not far from a representative panel of the Dutch population.

**Precision of the location measure.** Recall that we define the consumer residential location at the centroid of her 5-digit zipcode. We now examine the precision of this measure. For each 5-digit zipcode, we compute the distance to its closest neighbor measured in inter-centroid distance. In

[11]The LISS panel is administered by CentER Data (Tilburg University, The Netherlands). It is a a representative sample of Dutch individuals who participate in monthly Internet surveys. Households that could not otherwise participate are provided with a computer and Internet connection by CentER Data. A longitudinal survey is fielded in the panel every year, covering a large variety of domains including work, education, income, housing, time use, political views, values and personality. More information about the LISS panel can be found at: `www.lissdata.nl`.

[12]We include all households where the household head does not have a spouse. Then, for households with both the household head and his/her spouse, we randomly draw 64% female and 36% male as the primary respondent from the household. This procedure leads to 57% female overall, matching the JURY panel.

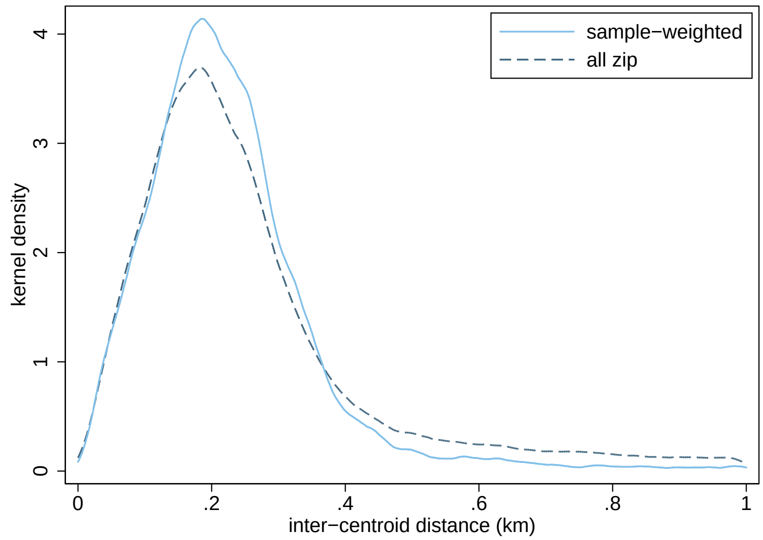



Figure 1: Distribution of inter-centroid distances

**Note:** Kernel density of inter-centroid distances. The dashed line represents the distribution among all 5-digit zipcodes, each to its nearest zipcode. The solid line weights the previous distribution by panelists in the sample.

Figure 1, the solid line measures the inter-centroid distances weighted by the panelist distribution. The modal inter-centroid distance is about 0.2 km, which implies a distance between the centroid and border of the zipcode at 0.1 km (110 yards). This distance bounds the modal measurement error of consumer location.

**Spatial distribution of offline purchases.** As the first step to understanding the role of consumer transportation costs, we examine the spatial distribution of consumer expenditures and document that offline expenditures concentrate within close proximity to a consumer's residence. In particular, Figure 2 displays the share of total panel expenditures as a function of panelists' travel distance to their closest store. The sum of the first 3 bars implies that 36% of a panelist's total expenditures occur offline *and* within 3 km of the consumer's residence. Further, 28% of expenditure occurs between 3 and 10 km and 11% between 11 and 20 km. Such a high concentration of expenditure in the local market provides initial support for the presence of sizable consumer transportation costs.

This figure also demonstrates the need for our granular location data. The modal offline-shopping distance is about 2 km, suggesting that studying spatially-differentiated retail markets (Smith, 2004; Zheng, 2016; Ellickson et al., 2020) and the cost of transportation requires precise measurement and characterization of consumer and store locations. Our metric of consumer location at the 5-digit

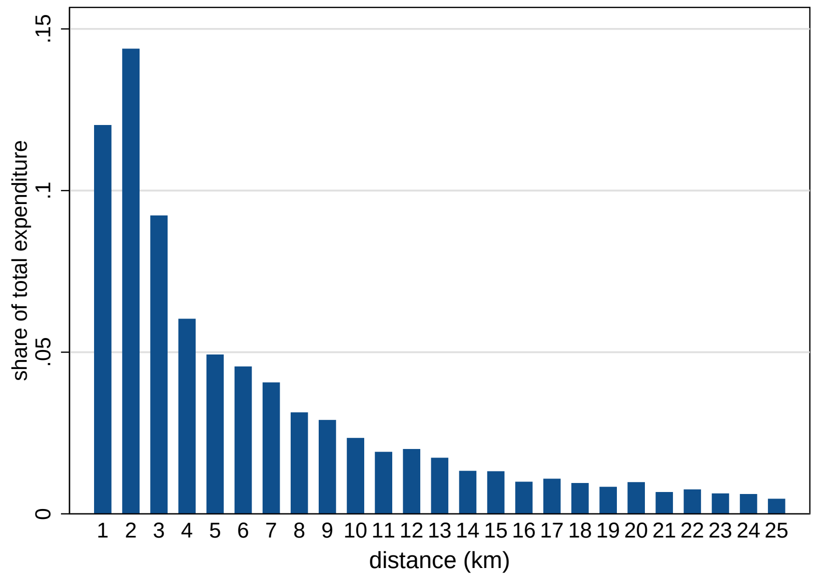



Figure 2: Distribution of expenditure across distance

**Note:** To provide detail in the short distances and avoid cluttering, the figure shows the spatial distribution of expenditure shares only for shopping trips $\leq 25$ km. This covers a 80% of offline expenditures (which is also shown by the sum of the bars).

zipcode has a modal measurement error of 0.1 km, which is small compared to the distance variation in the sample.

**The growth of online retail.** Figure 3 Panel A shows total quarterly panel expenditure (in log-scale) across all chains and channels in the data (solid line), and total expenditure from online shopping across these chains (dashed line). The share of online revenue is noted in the same figure also. While total revenue stays within a 20% band over the entire sample period, online retail expenditure expands very quickly. In 2007, only 5% of overall sales takes place online. In contrast, by 2018 it is 26%.

What explains the rapid growth of e-commerce? One explanation might be the changes in consumer and chain composition: that younger, more "tech-savvy" consumers enter the sample at a later point, or new online chains enter in the latter half of the sample. Another explanation is that consumers learn and develop habits for shopping online. A final explanation is that existing chains improve the quality of their online shopping experience over time. We decompose the growth of online-expenditure share into the three components. Specifically, we estimate

$$\text{expd\_share}_{ijt} = \alpha_{ij} + \sum_{\tau} \beta_{\tau} \mathbb{I}_{\#\text{online trips}_{it-1}=\tau} + \delta_{jt} + \varepsilon_{ijt}. \tag{3}$$

In this equation, the first term, $\alpha_{ij}$, captures time-invariant consumer and chain factors that correlate with the preference for online shopping. Note that while this term is time-invariant, changes in the set of consumers $i \in I_t$ and chains $j \in J_t$ over time $t$ partly capture the growth of e-commerce. The second term in equation (3), $\sum_{\tau} \beta_{\tau} \mathbb{I}_{\#\text{online trips}_{it-1}=\tau}$, captures the effect of past online trips, or the consumer's experience in online shopping. The third term captures chain-time effects that are common across consumers, which might represent changes in website design, product assortments, or a general online-shopping culture. We plot the average of the estimated time effect, $\bar{\delta}_t = \frac{1}{||J_t||} \sum_j \delta_{jt}$, with and without controlling for the composition effect and the experience effect.

Panel B of Figure 3 presents this decomposition. We find that about half of the growth is explained by the composition of consumers and chains, i.e., consumers who enter the sample at a later point have a stronger inclination to shop online, and online chains enter in the latter part of the sample. In addition, consumers' past online-shopping experience explains about 20% of the growth, and time trends the remaining 30%. To characterize these different mechanisms, our demand model captures rich consumer and chain heterogeneity, accommodates the effect of accumulating online-shopping experience, and allows for a flexible time trend.

**Store entry and exit.** Store entry and exit is essential to our identification of consumer transportation costs. We use the entry and exit data to calculate the number of unique stores for each chain in each year from 2000 until 2018. To summarize the evolution of chains, we focus on a subset of top-100 chains that (1) match with Orbis store location data and (2) survive from 2000 to 2018. We then normalize the number of stores for each chain by its number of stores in 2007 (the start of our main sample), and present the pattern in Figure 4. We find that within this set, chains generally continue to grow from 2000 to 2012. Beginning in 2013, with the rise of online revenue, the expansion of stores was no longer universal, and some chains reduced their number of stores, while the overall number of chains remained stable in the market. Beyond this figure, we find that 11 chains with offline stores entered after 2000 (3 of them entered after 2007), eight online chains entered in this

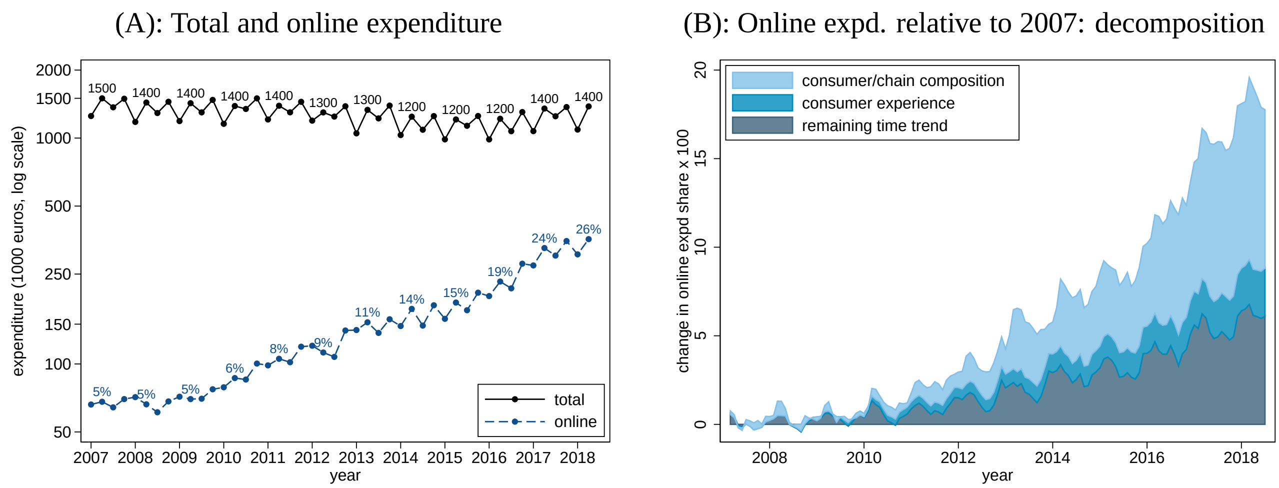



Figure 3: Growth of e-commerce in the apparel industry

**Note:** Left figure: Quarterly revenue (solid) and revenue from online sales (dashed). The solid lines depict quarterly total sales levels in our sample in 1000 Euros. The dashed lines represent the same for quarterly online sales only. The labels of the dashed lines reflect online sales as a percentage of total recorded sales. Right figure: regression coefficients $\bar{\delta}_t = \frac{1}{||J_t||}\sum_j \delta_{jt}$ from Equation (3) (with $\delta_{jt}$ normalized to 0 at $t = 2007$), controlling for both composition and experience (dark blue), controlling for only composition but not experience (blue), and not controlling for either (light blue). Note that $t$ is a month and the estimates are smoothed using a 3-month moving window.

period, and five chains exited prior to 2018.[13] In total, we document 2,284 store entry occasions and 446 store exit occasions during this period among the 100 largest chains. From this we conclude that store entry and exit is common in this industry.

# 3 Sensitivity to distance: identification and descriptive evidence

## 3.1 Identification and preliminary analysis

**Identifying variation.** A key parameter throughout this paper is consumers' transportation cost. This structural parameter captures the impact of travel distance on consumers' chain- and channel choice. Frequent entry and exit of stores (which characterizes this industry, as shown by Figure (4)) affects a consumers' minimum travel distance to a given chain. We use this variation for identification. In contrast, we do not focus on consumer relocation, and omit consumer migration from our analysis, because of the concern that relocation is driven by important changes in life (such as graduation, marriage, childbirth, and retirement) that are correlated with changes in shopping behavior.

[13] Also, 21 chains have no match in the Orbis data. These chains account for 8% of expenditure out of the top-100 chains.

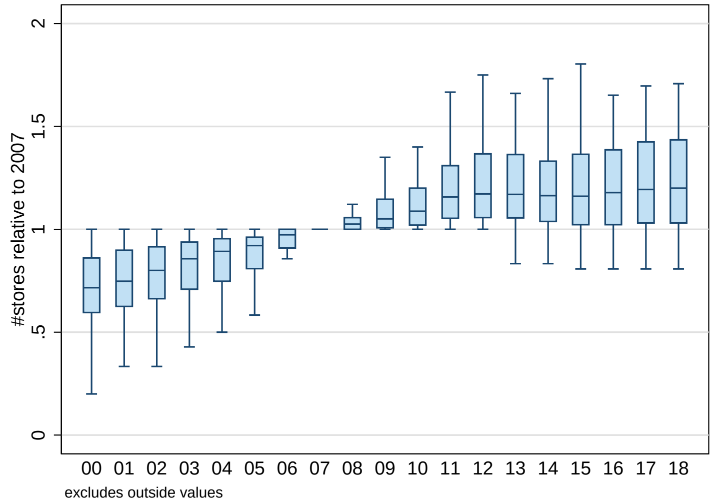



Figure 4: Distribution of the number of stores per chain (relative to year 2007)

**Note:** The figure shows the distribution (across chains) of the number of stores for each chain-year, divided by the number of stores for that chain in 2007. The horizontal bars represent the median, boxes represent the inter-quartile range (IQR), and the outer bars represent 1.5 times the IQR on either side.

However, one might still be concerned that store entry is a strategic decision made by the chain and that the chain tends to build stores in areas with a high density of customers who prefer the chain relative to its competition. As a result, entry decisions are potentially correlated with changes in the composition of heterogeneous customers in the local market, resulting in an endogeneity problem in market-level sales and entry data. To identify the causal effect of travel distance, we leverage the fact that a store cannot target *each customer* in the market. Therefore, each location decision creates heterogeneous "treatments" for individuals who live in different locations within the same market but are otherwise similar. Specifically, when a chain opens a new store, consumers close to it face a shorter travel distance. In contrast, consumers in other locations might be less impacted by the entry (for example, some consumers already have access to a different store that is even closer). Given that the new store cannot target each consumer, its entry creates a differential impact on different consumers, allowing us to construct a difference-in-differences strategy.

We illustrate this identification strategy in Figure 5. Panel (1) pictures a spatial market where the same chain operates three stores, A, B, and C. The closest store to the chain for Consumer 1 is Store B, whereas, for Consumer 2, Store A. Panel (2) shows the division of potential customers changes after Store D enters. Consumer 1 faces a drastic change in her distance to the chain because Store

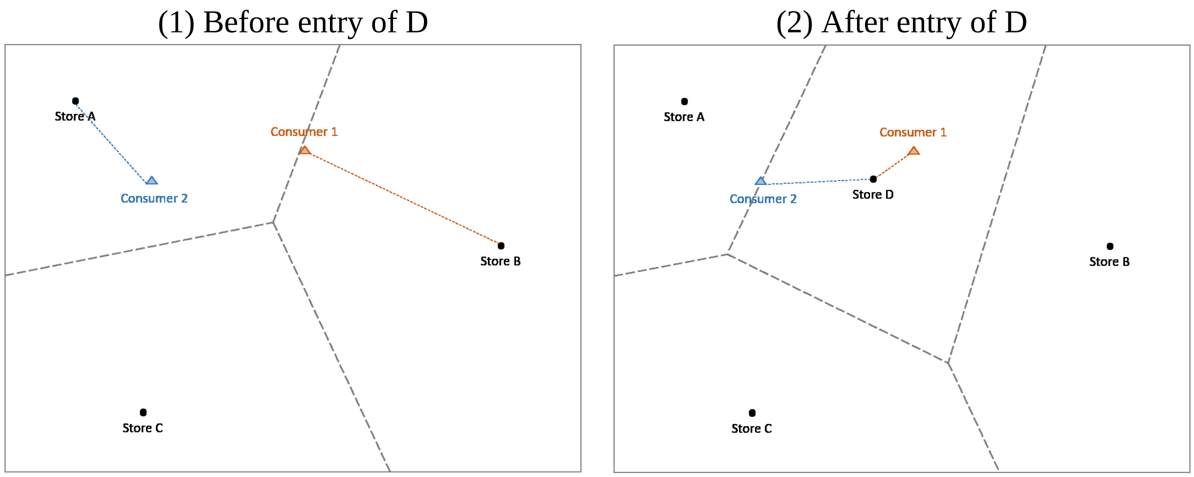



Figure 5: Heterogeneous exposure to the entry of store D

**Notes**: These two figures illustrate the idea behind the identification strategy. Panel A illustrates the division of potential customers to stores A, B and C. Panel B illustrates the division after store D enters.

D is much closer than Store B. In contrast, Consumer 2 barely experiences a change in distance to the chain because she is located at the border between Store A and D's coverage area. Consumers at different locations around store D form treatment and control groups to identify the sensitivity to distance.

**Is the "common time trend" assumption reasonable?** Although a store might selectively enter into areas where the overall customer composition matches the chain's clientele, the difference-in-differences strategy assumes that the store cannot target local consumers within its coverage area *individually*. Specifically, it assumes that these consumers *would change* their shopping behavior in the same way, so that store entry cannot select on *time-varying unobservables* across locations. Is this "common time trend" assumption reasonable in our context?

To begin with, we show that the store cannot (or does not) target the income of local households. Appendix Section B shows that conditional on consumer-chain and chain-year fixed effects (same as our main analysis), stores do not locate in the 5-digit zipcodes where the local consumer income matches with the income of its clientele. Broadening the market to the 4-digit zipcode level, targetting based on local income is still limited. We conclude from this analysis that chains do not target local observed demand shifters when the chain chooses its store locations.

Beyond demographics, one might wonder whether the store selectively enters areas with stronger shopping trends for the focal chain. In this case, one should find a stronger growth of purchase tendency for consumers closer to the store even *before* the store enters. To test for this possibility, we examine how consumer purchase patterns change before and after a new store enters, and how such changes depend on the consumer's location relative to the store. We first focus on consumer-chain pairs for which (1) the consumer lives within 20 km of the chain at the start of the sample,[14] and (2) the chain opens another store closer to the consumer at a later point. Define $\text{postEntry}_{ijt} = 1$ if a new offline store of chain $j$, closer to consumer $i$ than $j$'s existing offline stores, has entered in or before month $t$. We estimate the following equation:

$$\text{purchase}_{ijt} \times 100 = \beta_0 \times \text{postEntry}_{ijt} + \beta_1 \left(1 - \text{postEntry}_{ijt}\right) \times t + \beta_2 \text{postEntry}_{ijt} \times t + \delta_{ij} + \varepsilon_{ijt} \quad (4)$$

where $\text{purchase}_{ijt}$ is an indicator variable of any offline purchases by consumer $i$ at chain $j$ (then multiplied by 100) in month $t$.[15] Fixed effects $\delta_{ij}$ capture consumer $i$'s time-invariant offline-shopping tendency for chain $j$. The parameter $\beta_0$ is the effect of store entry on the incidence of purchase, and $\beta_1$ ($\beta_2$) is the time trend before (after) the opening of the closer store. We estimate Equation (4) separately for consumer-chain pairs whose distance changes no more than 0.1 km (corresponding to Consumer 2 in the example), and those whose distance changes in more meaningful ways. Consistent with our "common time trend" assumption, purchase patterns at different relative distances should not show different time trends before the store enters, i.e., $\beta_1$ should be the same between groups. In addition, we can also leverage the flexibility of Equation (4) to test whether the impact of chain entry is increasing in the change in distance, and whether this entry effect is instantaneous or gradual. Further, we take the sample of consumer-chain pairs without any actual store entry or exit, and randomly pick a hypothetical entry time to estimate Equation (4). From this placebo check, one should expect $\beta_0 = 0$ and the same $\beta_1$ as the set of consumers facing an actual store entry.

We present the estimates of Equation (4) separately for consumer-chain pairs whose distance is *reduced* by (1) 5-20 km, (2) 1-5 km, (3) 0.1-1 km, (4) 0-0.1 km, and finally, (5) the placebo test group without an actual store entry. Table 2 presents these estimates. First, the estimated time trends

[14] The choice of this geographic area is motivated by the observation that the coverage area of a store is typically the size of a city or a town. Specifically, for each store, we take the 95th percentile of the shopping-trip distance as a measure of the store's coverage radius, and report that the median of this radius (across stores) is 6.5 km, with an inter-quartile range of [2.8, 11.7] km.

[15] Note that the chain-channel-time data are rectangular at the consumer level, i.e., include observations with 0 purchases for each chain, channel, and month (while the consumer is an active panelist).

Table 2: Store-entry effect on offline shopping incidence

| | distance change 5-20km | 1-5km | 0.1-1km | 0-0.1km | no store entry (placebo) |
|---|---|---|---|---|---|
| post entry | 0.445*** | 0.347*** | 0.199*** | 0.079 | 0.021** |
| | (0.038) | (0.035) | (0.043) | (0.050) | (0.006) |
| pre trend | -0.004** | -0.003* | -0.005** | -0.007** | -0.008*** |
| | (0.001) | (0.002) | (0.002) | (0.002) | (0.000) |
| post trend | -0.004** | -0.002 | -0.003 | -0.006** | -0.007*** |
| | (0.002) | (0.001) | (0.002) | (0.002) | (0.000) |
| R-squared | 0.15 | 0.15 | 0.15 | 0.15 | 0.16 |
| observations | 1,135,074 | 1,361,960 | 813,099 | 473,821 | 27,299,660 |

**Notes:** This table shows estimates of Equation (4), focusing on consumer-chain pairs where the closest store is within 20 km of the consumer at the start of the sample, and where the chain builds a store and reduces its distance to the customer. We divide the sample into groups where the entry has different impact on the travel distance of the customer. Column 1-4 examines customer-chain pairs where the distance change at entry is 5-20 km, 1-5 km, 0.1-1 km, and 0-0.1 km. Finally, the last column presents a placebo test where we focus on consumer-chain pairs where no store entry or exits are relevant, and we hypothetically assign a "store entry date" for each of such pairs.

both before and after the store entry are all economically small, and the differences in pre-trends between groups are small. For example, the difference in pre-trends between columns (2) and (5) is 0.005 percentage points per month (in offline choice probability), about 1.4% of the entry effect in magnitude. This result is consistent with the assumption that consumers close to the new store location –i.e., those most impacted by the store entry– do not have meaningfully different time-varying shopping patterns than other consumers in the broader area. In other words, consumers at different locations in the neighborhood share a "common time trend," supporting the difference-in-differences strategy that leverages the heterogeneous treatment of store entry on different consumers to identify distance sensitivity.

Second, we find that consumers whose travel distances are most impacted change their behavior most drastically. On average, the offline-shopping incidence at a chain increases by 0.45 percentage points if a chain moves closer to the consumer by 5-20 km (conditional on the chain was within 20 km before the new store entry). In contrast, shopping incidence only changes by 0.08 percentage points if the new store is barely any closer, and the effect is statistically insignificant. In addition, we find that the post-entry trend in purchase tendency is similar to the pre-entry trends, consistent with the hypothesis that the effect of entry is instantaneous and constant during the post-entry period, rather than gradually materializing after entry. Further, our placebo test shows that shopping incidence does not substantively change when there is no actual store entry or exit (although the post-entry dummy estimate is significant due to the very large sample size). Figure 6 plots the im-

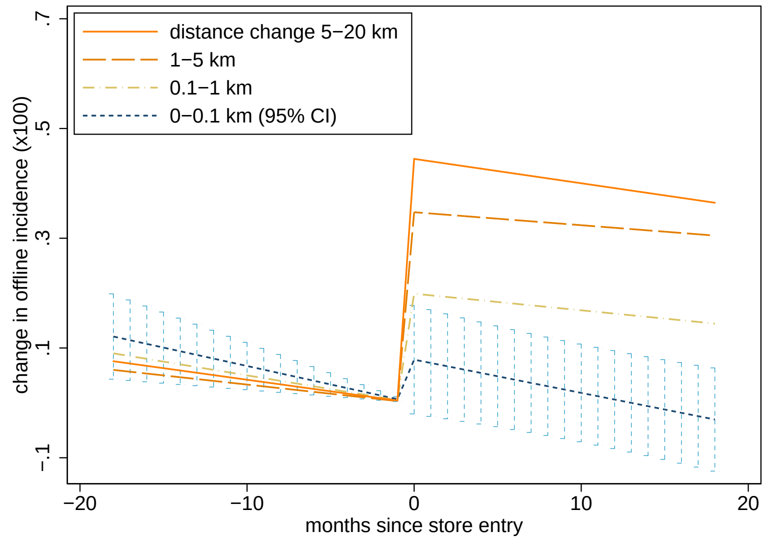



Figure 6: Store-entry effect on offline shopping incidence: illustration

**Notes**: This figure visualizes the estimated time trends in Column 1-4 of Table 2. Whiskers are the 95% confidence interval for the 0-0.1 km group.

plied time trends from the first four columns of Table 2 for better visualization. Finally, we run a similar analysis on individual-chain pairs where the individual faces a store exit. Results are shown in Appendix Table 7.

We also note that this identification strategy is similar to "spatial difference-in-differences" (Ellickson and Grieco, 2013), yet with one crucial difference. We exploit the panel-data structure and allow for heterogeneity across consumers within and across markets, and possible variations in the *composition* of customers across markets – which might be an important driver of store entry. In this sense, our identification strategy is similar to the literature on estimating the effect of geographical distance using individual-level shopping data (Smith, 2004; Wang and Goldfarb, 2017; Shriver and Bollinger, 2020). It is different from the literature using aggregate data (Forman et al., 2009; Ellickson and Grieco, 2013; Ellickson et al., 2020; Li, 2019).

## 3.2 Sensitivity to distance and within-chain substitution to online shopping

Now, we formally examine how the distance to the chain causally influences consumer shopping decisions. Separately for each channel $c = 0, 1$, we estimate a linear probability model of individual

$i$'s purchase decision at chain $j$, in month $t$ as a function of distance to the nearest store, plus covariates. Suppressing $c$ for compactness of notation, we specify

$$\text{purchase}_{ijt} \times 100 = \beta \log\left(D_{ijt}+1\right) + \delta_{ij} + \lambda_{jy(t)} + W_{it}' \cdot \gamma + \varepsilon_{ijt}. \tag{5}$$

where $\log\left(D_{ijt}+1\right)$ is the log distance to the chain,[16] and $\delta_{ij}$, and $\lambda_{jy(t)}$ are consumer-chain and chain-year fixed effects. $\lambda_{jy(t)}$ captures the "common time trend" across all consumers shopping at the chain, regardless of the location of the consumers. We discuss alternative specifications with more flexible time trends below. $W_{it}$ are additional time-varying observables and contain month-of-the-year dummies and demographic variables, including income, work hours, education, age, family size, employment status, retirement status, and home-ownership status.

The first two columns of Table 3 present the main results across all customers and chains using the full sample. We find that offline demand is adversely affected by distance. For example, increasing distance from 0 km to 1 km will reduce the probability of purchasing at the store by $0.338 \times (\log(1+1) - \log(0+1)) = 0.234$ percentage points. The baseline purchase probability if the consumer-chain distance is within 1 km is 2.357 percentage points. Therefore, a 1 km change of distance causes a 9.9% change in incidence. Meanwhile, the 1 km change in distance will also increase online purchase probability by 0.003 percentage points (which equates to 3.3% of the baseline incidence), but this substitution effect is statistically indistinguishable from zero.

To further examine how travel distance impacts demand for on- and offline stores of the same chain, we focus on existing customers, defined as those who have previously purchased from the chain.[17] We re-estimate Equation (5) for this subset of consumer-chain pairs and find that, while changes in the distance to the chain discourage shopping at the store for existing customers, the percentage effects (semi-elasticities) are not larger for existing customers. Specifically, their average propensity of making a purchase is 6.821 percentage points, implying a semi-elasticity of 9.8%. On the other hand, we find that an increase in distance drives existing consumers to buy online. For example, the nearest store moving from 0 to 1 km will increase the customer's tendency to shop online by 0.020 percentage points, a semi-elasticity of 11.1%. Assuming away *direct* effects of

[16]We add one so that $\log(D_{ijt}+1) = 0$ as $D_{ijt}$ goes to zero.

[17]To construct the sub-sample of existing customers, we drop observations at and before the first observed purchase incidence for each individual-chain pair in order to focus on the time periods after the consumer have purchased from the chain.

Table 3: Sensitivity of shopping incidence to distance

| | all customers | | existing customers | |
|---|---|---|---|---|
| | offline | online | offline | online |
| log(distance + 1) | -0.338*** | 0.003 | -0.962*** | 0.030* |
| | (0.012) | (0.002) | (0.084) | (0.015) |
| consumer-chain FE | Yes | Yes | Yes | Yes |
| chain-year FE | Yes | Yes | Yes | Yes |
| month FE | Yes | Yes | Yes | Yes |
| demographics | Yes | Yes | Yes | Yes |
| R-squared | 0.14 | 0.10 | 0.14 | 0.12 |
| observations | 54181026 | 54181026 | 7714817 | 7714817 |

**Notes:** This table presents the sensitivity to distance of consumer's purchase incidence. A unit of observation is an individual-chain-channel-month. Column 1 and 2 presents evidence for the entire sample. Column 3 and 4 focuses on individual-chain pairs where the individual have shopped at the chain before (we cut the sample after the first trip). The number of observations for online and offline purchases are the same by construction, i.e., for each consumer, our data at the chain-channel-month level are rectangular. Distance to offline chains is equal to the distance to the closest offline store of the same chain (if they exist) or the distance to the head quarter of pure-play online chains.

distance on the decision to shop online, this distance effect implies that, on average, offline and online stores are net substitutes for existing customers.

There is a long-standing literature on the extent to which online and offline retail channels are complements or substitutes (Gentzkow, 2007; Forman et al., 2009; Pozzi, 2013; Bell et al., 2017; Wang and Goldfarb, 2017; Shriver and Bollinger, 2020). Our finding is consistent with Wang and Goldfarb (2017), in that offline stores might create an information spillover effect to online sales, but that spillover effect only exists for new consumers. In addition, the Dutch retail apparel industry contains many existing, well-known brands, which potentially explains why one might see a lesser degree of information spillover effect than documented by others.

In Appendix Table 4, we present significant heterogeneity in the sensitivity to distance across retail formats. The consumer is most sensitive to the distance to a discounter or a general merchandizer. This is consistent with the conjecture that discounters and general merchandizers supply less differentiable products or offer little service, compared to branded chains and department stores. In addition, specialty chains are the only format where we find a net complementarity effect from offline to online. Directionally, this finding is consistent with specialty chains being less well-known than a branded chain, and therefore, proximity plays an information role. However, the lack of either a substitution or a complementary effect between channels overall adds to the debate about whether the offline channel is a substitute or a complement to online shopping (Gentzkow, 2007; Forman et al., 2009; Bell et al., 2017; Wang and Goldfarb, 2017; Shriver and Bollinger, 2015; Zhang et al., 2018). While several papers find that offline stores complement online sales, we do not find this effect in a market dominated by large chains selling known brands, despite using an identification strategy similar to the literature.

Our chosen functional form of distance effects, $\log\left(D_{ijt}+1\right)$, implies decreasing marginal sensitivity to each additional kilometer of distance. In Appendix Figure 2, we present estimates of a flexible specification of distance and confirm that the marginal effect of distance decreases with distance.

**Robustness checks.** The primary specification to estimate consumers' sensitivity to distance, Equation (5), assumes common time trends in the shopping incidence to each chain and channel across all consumers. While the estimated pre-trends across consumers in different locations largely support this assumption, we further examine its robustness to alternative specifications that allow for *time-varying* local demand.

First, we further control for store zip-year fixed effects, which represent unobserved demand for consumers in the trade area of a given store. For example, in Panel B of Figure 5, store D's entry might capture changes in demand around the coverage area of store D (the area within the dashed lines around the store). Appendix Table 5 shows that the parameter estimates are very similar to the first two columns of Table 3. This finding suggests that store entry decisions do not target time-varying preferences at the store's trade-area level.

Second, one might wonder whether variations in consumer compositions, such as the expansion of new communities, will affect the preferences of a fixed set of consumers through peer effects. We run a second robustness check, controlling for individual-chain fixed effects and household zip2-year fixed effects. Appendix Table 5 shows that adding household zip2-year fixed effects have no impact on the parameter estimates. These robustness checks further support the main identifying assumption that the heterogeneity in shopping trends across different areas is limited. With these results in mind, in the structural analysis later, we account for individual-level heterogeneity and chain or channel level trends that are common across consumers.

## 3.3 Substitution across chains

The consumer value of convenience not only depends on the sensitivity to travel distance but also the degree to which she substitutes to other chains when this distance varies. In preparation of our structural analysis in Section 4, we directly examine the distance-driven substitution effect across chains. To keep the analysis simple, we focus on the role of distance to consumer $i$'s "primary" chain $k$ and her decision to purchase at stores of rival chains $j \neq k$. We define the primary chain as the one with the highest expenditure for each consumer during her first year in the sample. We estimate the following linear probability model of purchase incidence as a function of distance to the non-primary chain $j$ and to the primary chain $k$,

$$\begin{aligned}\text{purchase}_{ijt} \times 100 = \phi_1 \log\left(D_{ikt}+1\right) \mathbb{I}_{D_{ijt} \leq 4} + \phi_2 \log\left(D_{ikt}+1\right) \mathbb{I}_{D_{ijt} > 4} + \\ \beta \log\left(D_{ijt}+1\right) + \delta_{ij} + \lambda_{jy(t)} + W_{it}' \cdot \gamma + \varepsilon_{ijt}, j \neq k \end{aligned} \tag{6}$$

where, as before, we control for individual-chain, chain-year, month-of-the-year, and demographic fixed effects. This regression drops observations for the primary chain $j = k$, and the first year in the sample for each consumer. We are interested in whether the substitution pattern changes with the distance to the primary chain. As such, we allow purchase incidence to respond to chain $k$'s location differently depending on whether chain $j$ is above or below the median shopping distance to the consumer, i.e., 4 km. The main hypothesis is that chain $j$'s purchase incidence decreases as chain $k$ moves closer to the individual, i.e. $\phi_1, \phi_2 > 0$. We control for the effect of distance between $i$ and $j$.

Table 4 shows that the distance sensitivity $\beta$ is very close to own-chain distance sensitivity reported in Table 3. The cross-chain effects $\phi_1$ and $\phi_2$ suggest that increasing the distance to the primary chain, $D_{ikt}$, will increase the individual's tendency to purchase at a non-primary chain. Yet, in the offline channel, this substitution effect only exists when chain $j$ is within 4 km of the individual. In this case, the cross-chain distance effect is 24% of the magnitude of the own-chain distance effect. These estimates imply meaningful substitution, hence competition, in the offline channel. In the online channel, we no substitution effect from the primary chain's offline store to the non-primary chain's online store.

Table 4: Sensitivity of shopping incidence to distance to other chains

| | purchase at chain j | |
|---|---|---|
| | offline | online |
| log distance to chain j | -0.339*** | 0.001 |
| | (0.017) | (0.002) |
| log distance to chain k (j within 4km) | 0.083*** | -0.006 |
| | (0.025) | (0.004) |
| log distance to chain k (j outside of 4km) | 0.012 | -0.002 |
| | (0.017) | (0.003) |
| consumer-chain FE | Yes | Yes |
| chain-year FE | Yes | Yes |
| month FE | Yes | Yes |
| demographics | Yes | Yes |
| R-squared | 0.13 | 0.08 |
| observations | 32894846 | 32894846 |

**Notes:** This table presents the effect of log distance to the primary chain on purchase incidence from non-primary chains.

## 3.4 The effect of distance on expenditure

We further examine the effect of distance and price (index) on offline expenditure given purchase incidence. We focus on offline purchases and estimate

$$Y_{ijt} = \beta \log\left(D_{ijt}+1\right) + \alpha \log\left(P_{jt}\right) + \delta_{ij} + \lambda_{y(t)} + W_{it}' \cdot \gamma + \varepsilon_{ijt}. \tag{7}$$

where $Y$ is either purchase incidence or log offline expenditure given purchase.[18] Note that Equation (7) controls for individual-chain fixed effects like before, but only year fixed effects instead of chain-year fixed effects. This is because chain-year fixed effects would have absorbed most of the price variation, which is common across individuals.

Table 5 shows that the main effect of both distance and price on purchase incidence is negative. On the contrary, prices have a positive impact ($\hat{\alpha} = 0.490$) on expenditure given purchase, implying that quantity demand given incidence has a price elasticity of $\hat{\alpha} - 1 = -0.510$. Given that the elasticity to prices is mainly at the extensive margin, our supply-side model will characterize pricing decisions where firms consider using prices to attract consumers (and purchase quantity given

[18]We focus on the top 20 chains to estimate Equation (7) because for those chains we have ample observations purchasing price and can reliably compute price indices.

Table 5: Sensitivity of offline shopping incidence and expenditure to distance and price

| | purchase x100 | log expd if purchase |
|---|---|---|
| log(dist + 1) | -0.828*** | 0.005 |
| | (0.034) | (0.009) |
| log(price index) | -4.315*** | 0.490*** |
| | (0.392) | (0.090) |
| consumer-chain, year, month, and demographics FE | Yes | Yes |
| R-squared | 0.15 | 0.54 |
| observations | 14139425 | 324531 |

**Notes:** This table presents the effect of log distance and log price index on offline purchase incidence and on log offline expenditure given purchase incidence. We focus on a subsample of the top-20 chains with ample observations of consumer purchases.

incidence stays fixed).

To summarize, this section describes the sensitivity of a consumer's purchase patterns to her distance to the chain. We first illustrated that store entry creates heterogeneous exposure to distance across otherwise-similar consumers, setting up a "spatial difference-in-differences" identification strategy for the effect of distance. We then demonstrate that increasing the distance to one chain leads to lower offline-purchase incidence from that chain, some substitution to other chains for existing consumers, and substitution to other nearby chains. We also demonstrate that expenditure given purchase has little response to distance (and limited response to price), allowing us to focus on purchase incidence in subsequent analysis. These findings inform the construction of a structural demand model in Section 4.

# 4 Model

## 4.1 Demand

To quantify and decompose the value of e-commerce, we construct and estimate a structural demand model to characterize consumer choices of shopping online or offline, and choices of shopping at chains in different locations. The parameters of interest in our model are transportation costs and price sensitivity, central to quantifying the value of e-commerce.

Our model focuses on the consumer's choice to purchase at a given chain or channel. Motivated by the evidence in Section 3.4, we abstract away from the choice of individual items and purchase quantity given store choice, and from multiple-chain choices within the period. This abstraction keeps the basic model structure simple and accommodates rich set of heterogeneity, at the cost of omitting 22% of total expenditure.[19]

Consumer $i$ in year-month $t$ chooses between focal chains $j = 1,2,\ldots,15$ and other formats $j = 16,\ldots,20$, and the outside option $j = 0$.[20] For each of these inside goods, we allow two transaction channels $c \in \{0,1\}$, where 0 stands for shopping offline and 1 stands for online, except for a few pure-play online sellers who are only available with $c = 1$. A combination of chain $j$ and channel $c$ yields indirect utility

$$u_{ijct} = \alpha_i \log\left(P_{jt}\right) + \beta_i \log\left(\mathbb{I}_{c=0} \cdot D_{ijt} + 1\right) + X_{jc}\gamma_i + S_{ijt}\theta + \delta_{jc} + \left(\lambda_j + \lambda_c\right)t + \varepsilon_{ijct} \tag{8}$$

where $\log\left(P_{jt}\right)$ is the log price index of chain $j$ in month $t$, $\mathbb{I}_{c=0}$ is a dummy for shopping offline, $D_{ijt}$ is the distance (in km) between individual $i$ and the nearest store of chain $j$, $X_{jc}$ are time-invariant chain characteristics, and $S_{ijt}$ captures various consumer states, which includes chain and location state dependence and the effect of consumer experience (see below). Further, $\alpha_i$ and $\beta_i$ capture heterogeneous sensitivities to price and distance, $\gamma_i$ captures heterogeneous consumer preferences for chain-channel characteristics, $\theta$ captures state dependence, $\delta_{jc}$ is a chain-channel fixed effect, and $\lambda_j$ and $\lambda_c$ capture chain- and channel- level time trends. The observed chain characteristics include chain-level normal random coefficients, common across channels, that capture individual-specific tastes for each chain or retail format. It also includes other time-invariant characteristics of the chain or the channel.[21] The consumer can choose not to purchase from any chain-channels and choose the outside option $j = 0$, for which we normalize $u_{i0t} = \varepsilon_{i0t}$.

We next model the random coefficients on observed chain-level characteristics, $\gamma_i$, as functions of demographics $Z_i$ and standard normal random draws $\nu_{i1}$,

$$\gamma_i = \overline{\gamma} + Z_i\gamma_z + \sigma_\gamma \nu_{i1}, \tag{9}$$

[19] In our consumer-month level data, 16% of observations contain at least two trips.

[20] These formats concern fringe sellers collapsed into a single label by format: small branded chains ($j = 16$), discounters (17), general merchandizers (18), online retailers (19), and specialty chains (20). Together, formats $j \in \{16,\ldots,20\}$ account for 46% of recorded expenditure.

[21] These additional observed characteristics include a 1) a dummy variable indicating online channel, 2) a dummy variable indicating that the average price index of the chain is below 10 euros, 3) average concentration of brands sold by the chain, measured by the Herfindahl Index (HHI) of brand shares within the chain, and 4) the average fraction of sales revenue on shoes, as a measure that distinguishes retailers that specialize on shoes.

where the demographic variables $Z_i$ include age bins (cutoff at 25 and 45 years old), gender, income bins (net monthly income of 1000, 2000, and 3000 euros), and an indicator that the consumer chooses to not report income. For the random coefficients on chain dummies, we specify them as independent normal random variables that do not depend on demographics. Further, for random coefficients on log price ($\alpha_i$) and distance ($\beta_i$), we impose a theoretical prior on their sign using the following functional form for $\alpha_i$ (similar for $\beta_i$):

$$\alpha_i = \bar{\alpha} \cdot \exp\left(Z_i \alpha_z + \sigma_\alpha \nu_{i2}\right). \tag{10}$$

The states $S_{ijt} = \left(S^1_{ijt}, S^2_{ijt}, S^3_{izt}, S^4_{it} \cdot \mathbb{I}_{c=1}\right)$ capture various dependencies of choice behavior on the past. First, $S^1_{ijt}$ indicates that customer $i$ is new to chain $j$, defined as $i$ having never purchased from $j$ in $t = 1, ..., t-1$. We interpret such state dependence as an information effect, i.e. new customers might learn some information, such as fit and feel of the clothing items bought, when they purchase from a chain for the first time. We include this variable because the evidence in Section 3 shows that within-chain substitution is present only for existing consumers, implying some degree of complementarity for new consumers. Next, closely following the construction in (Dubé et al., 2009, 2010), $S^2_{ijt}$ represents the consumer's last visited chain $j = 1, ..., 15$. Given chain-channel fixed effects and consumer-chain random coefficients, we interpret this term as a cost of switching away from the previous chain of choice. Third, $S^3_{izt}$ represents whether the switching cost from the location of previous choice $z$, in 4-digit zip code. This term captures the possibility that consumers like a certain shopping area and have a preference for any fashion chain in that location. Finally, $S^4_{it} \cdot \mathbb{I}_{c=1}$ is the consumer's total number of times shopping online in the past across all chains, interacted with the choice of shopping online ($c = 1$). This term captures potential experience effects towards online shopping, knowledge about website layouts or the checkout process, or simply shopping habit. We code $S^4_{it}$ as a categorical variable to estimate its effects non-parameterically.

Finally, the stochastic term $\varepsilon_{ijct}$ captures unobserved tastes. Conditional on the effects of chain, channel, and customer characteristics, we assume that $\varepsilon_{ijct}$ are IID with an type I Extreme Value distribution.

## 4.2 Estimation of demand

For compactness of notation, denote $\Omega_t = \left\{D_{ijt}, P_{jt}, S_{ijt}, X_{jc}\right\}_{i,j}$ as the collection of all observable covariates at time $t$. The extreme value utility shocks $\varepsilon_{ijct}$ imply that individual $i$ chooses chain-channel $(j,c)$ combination with probability

$$l_{it}\left(\Theta_i\right) := \Pr\left(j,c|\Omega_t;\Theta_i\right) = \frac{\exp\left(u_{ijct}\right)}{1+\sum_{j',c'}\exp\left(u_{ij'c't}\right)} \tag{11}$$

where $\Theta_i = \left(\bar{\alpha}, \alpha_z, \sigma_\alpha, \bar{\beta}, \beta_z, \sigma_\beta, \bar{\gamma}, \gamma_z, \sigma_\gamma, \delta, \theta, \lambda\right)$ are parameters of interest. Without knowing individual $i$'s type, the probability of realizing the observed series of choices thus follow the *ex ante* likelihood

$$\Pr\left(\left\{j_t, c_t\right\}_{t=1,\ldots,T_i} |\Omega_t\right) = \int \prod_{t=1}^{T_i} l_{it}\left(\Theta_i\right) dF\left(\Theta_i\right)$$

where $F\left(\Theta_i\right)$ is the probability distribution of the individual-specific coefficients (Kamakura and Russell, 1989).

We estimate the model using simulated maximum likelihood, on a random sub-sample consisting of 10% consumers.[22] For each individual, we take $50 \times (J+6)$ Halton draws $\nu$ to represent the normally-distributed random variables, and obtain consumer-level random coefficients from Equation (9) and (10). We use different draws $\nu$ for different individuals. Then, fixing the set of draws $\nu_\iota$ for $\iota = 1, \ldots, 50$, we maximize the simulated log likelihood function

$$LL = \sum_{i=1}^{N} \log\left(\frac{1}{50}\sum_{\iota=1}^{50}\prod_{t=1}^{T_i} l_{it}\left(\Theta_\iota\right)\right). \tag{12}$$

We take a version of the model without normal random coefficients (i.e. logit demand) and estimate it with multiple starting points, and then use the converged value of the logit demand as starting value for the random coefficient demand model. We compute the numerical Hessian matrix at the parameter estimates and use the Hessian to compute the asymptotic standard error.[23]

[22]We use a random sample because of the high computation burden to evaluate the likelihood. We use a graphical processing unit and the likelihood function takes roughly 0.6 seconds to evaluate once, about 10 times faster than computing using a CPU. We choose this sample size because of the binding graphical memory.

[23]We estimate the logit model without random coefficients 50 times from random starting values drawn from a standard normal distribution. Of these, 46 out of 50 estimates converged to the exact same point. We use this point as the starting value for the random coefficient model. Two sets of initial conditions lead to qualitatively similar estimates but with lower likelihood, and two sets initial conditions lead to implausible estimates with very low likelihood. Although we cannot rule out that our reported parameters are associated with a local maximum of the likelihood function, the

## 4.3 Supply

While the demand model serves as the primary framework for quantifying and decomposing gains from e-commerce, we also estimate a simple static pricing game to infer marginal costs and assess the impact of e-commerce on equilibrium prices. For compactness of notation, denote the expected purchase volume for chain $j$ at $t$ as

$$Q_j(\Omega_t) = \sum_{c=0,1} \int_i q_{ijc}(\Omega_t) \cdot \Pr(j,c|\Omega_t;\Theta_i)\, dF(\Theta_i) \tag{13}$$

where $\Pr(j,c|\Omega_t;\Theta_i)$ is the probability of purchase for consumer-type $\Theta_i$ and $q_{ijc}(\Omega_t)$ is $i$'s purchase quantity (of representative products) *given purchase incidence*. One might argue that in-store purchase quantities might still respond to the overall price level of the chain. From Section 2.3, we show that the price indices mainly affect incidence but have a small effect on purchase quantity given incidence. This finding allows us to focus on the sensitivity of incidence rather than purchase quantity to greatly simplify the model.[24] We therefore assume that quantities $q_{ijc}(\Omega_t) = \bar{q}_{jc}$, i.e., are constant for a given chain-channel, to simplify away the consumer decision of quantity given incidence.[25]

Given this assumption, the supply model boils down to a set of first-order conditions in a static pricing game. Specifically, we assume that retail chains set uniform prices across the two channels and compete in a static Bertrand-Nash game (Berry et al., 1995). In each month $t$, chain $j$ sets prices $p_{jt}$ maximizing its profit:

$$\Pi_{jt} = \left(P_{jt} \cdot (1-\tau) - \mathrm{mc}_{jt}\right) \cdot Q_j(\Omega_t). \tag{14}$$

Where $\tau = 0.09$ is the value-added tax (VAT) for clothing and shoes. We impose that a chain's

convergence to the same estimates is encouraging.

[24] To characterize how quantity responds to prices, one will need to model consumers' choices of product variety and quantity in a structural way. Estimating such a model will require simplifications in other dimensions, as we discussed in Section 4.

[25] In practice, we take the observed average expenditure given purchase incidence divided by the price index as a proxy for $\bar{q}_{jc}$.

marginal costs, $\overline{\mathrm{mc}}_j$, are constant over time and that prices are optimal for each chain *on average*, or

$$\mathbb{E}\left[\frac{\partial \Pi_{jt}}{\partial P_{jt}}\right] = \sum_{c=0,1} \bar{q}_{jc} \cdot \frac{1}{T} \sum_{t=1}^{T} \int \Pr\left(j, c | \Omega_t; \Theta_i\right) dF\left(\Theta_i\right) +$$
$$\left(\bar{P}_j - \frac{\overline{\mathrm{mc}}_j}{1-\tau}\right) \cdot \sum_{c=0,1} \bar{q}_{jc} \cdot \frac{1}{T} \sum_{t=1}^{T} \left(\int \frac{\partial \Pr\left(j, c | \Omega_t; \Theta_i\right)}{\partial P_{jt}} dF\left(\Theta_i\right)\right) = 0. \tag{15}$$

These first-order conditions imply a system of $J$ equations and $J$ unknowns and allow us to exactly solve for the average marginal costs for each chain.

# 5 Estimation results

**Demand-side parameter estimates.** Table 6 presents structural estimates. Panel A presents baseline coefficients for the main model, i.e. $\bar{\alpha}$, $\bar{\beta}$, and other common parameters. Panel B presents estimates for demographic interactions and standard deviation for the random coefficients. Appendix Table 8 presents chain-level parameters, including chain-channel intercepts, chain trends, and chain-specific scale of random coefficients.

The baseline estimates for the sensitivities to price and distance ($\bar{\alpha}$ and $\bar{\beta}$) indicate that consumers are not only averse to high prices but also to high travel distance, i.e., both elements are "costs" of shopping. The average log price coefficient (the average $\alpha_i$) is -1.918.[26] Similarly, the log distance sensitivity averages to -0.329.

We also find considerable heterogeneity in the sensitivities to price and distance across consumers. In particular, younger consumers and men are more sensitive to travel distance, whereas lower-income consumers and men are more sensitive to price. In addition, there is a considerable amount of heterogeneity in $\beta_i$ not explained by observed demographic variables; this heterogeneity is captured by random household-specific components $\nu_i$. Finally, compared to distance sensitivity, there is less heterogeneity in price sensitivity across consumers.

We graphically illustrate the distribution of preference heterogeneity by two hypothetical individuals: Ben, a male student less than 25 years old with net income under 1,000 euros per month, and Amy, a female working professional more than 45 years old with net income over 3,000 euros

[26] Note that the baseline price and distance sensitivities are not the averages of the population effect, because of the exponential function form for the random coefficient and that the mean demographic variables are non-zero.

Table 6: Parameter estimates for the structural model

Panel A: baseline parameters

| | coef. est. | s.e. |
|---|---|---|
| log(distance + 1) | -0.306 | 0.043 |
| log(price) | -1.918 | 0.165 |
| never purchased at chain | -1.685 | 0.024 |
| shop at same chain | 0.635 | 0.019 |
| shop at same location | 0.057 | 0.016 |
| purchased online: once | 0.766 | 0.060 |
| ... twice | 1.043 | 0.066 |
| ... 3-5 times | 1.246 | 0.057 |
| ... 6-15 times | 1.641 | 0.059 |
| ... 16+ times | 2.344 | 0.072 |
| online x year since 2007 | 0.156 | 0.010 |

Panel B: demographics and random coefficients

| | shop online | s.e. | log(distance + 1) | s.e. | log(price) | s.e. | brand conc. | s.e. | share of shoes | s.e. |
|---|---|---|---|---|---|---|---|---|---|---|
| std. of random coef. | 0.351 | 0.027 | 0.844 | 0.037 | 0.141 | 0.013 | 0.410 | 0.024 | 0.029 | 0.065 |
| age<=25 | 0.408 | 0.100 | -0.459 | 0.151 | 0.039 | 0.018 | -0.060 | 0.080 | 0.262 | 0.229 |
| 25<age<=45 | 0.730 | 0.058 | -0.338 | 0.096 | 0.042 | 0.010 | 0.168 | 0.041 | 0.625 | 0.120 |
| female | 0.001 | 0.058 | -0.018 | 0.071 | -0.097 | 0.013 | 0.393 | 0.042 | -0.523 | 0.120 |
| 1000<inc.<=2000 | -0.191 | 0.122 | -0.098 | 0.125 | -0.043 | 0.021 | -0.043 | 0.093 | 0.092 | 0.291 |
| 2000<inc.<=3000 | -0.258 | 0.124 | 0.072 | 0.126 | -0.065 | 0.023 | -0.010 | 0.093 | -0.155 | 0.288 |
| inc.>3000 | -0.003 | 0.128 | -0.444 | 0.141 | -0.042 | 0.022 | -0.024 | 0.094 | 0.164 | 0.295 |
| inc. missing | 0.450 | 0.075 | -0.093 | 0.123 | -0.047 | 0.012 | 0.138 | 0.055 | -0.429 | 0.162 |

**Notes:** Panel A presents mean coefficients estimates, and Panel B presents interactions with demographics and standard deviations for random coefficients. To avoid cluttering, we omitted from the table chain-channel intercepts and chain-time trends; these parameters are reported in Table 8, and month-of-the-year fixed effects. Standard errors are in parenthesis, obtained from the diagonal of the inverse Hessian matrix. The number of observations is 88,519 at the unit of individual-month. The log likelihood of our model is -148,750.

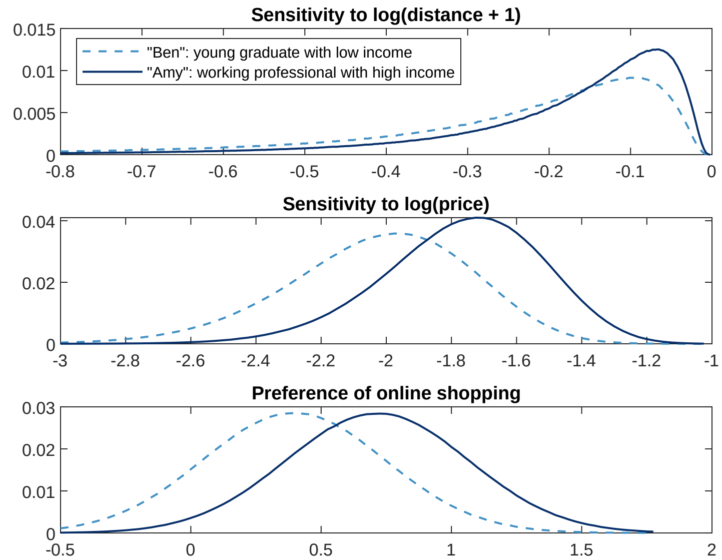



Figure 7: Heterogeneity across consumers: two hypothetical examples

**Notes**: Distribution of random coefficients based on two hypothetical examples. The dashed lines represent "Ben", who is male, 25 years old, income below 1,000 euros per month. The solid lines represent "Amy", who is female, 45 years old, and earns an income above 3,000 euros per month.

per month. Figure 7 shows the distribution of distance and price sensitivity as well as consumer-specific tastes towards online shopping: Ben is more sensitive to distance and price, but values online-shopping (directly) less. Beyond differences across demographic groups, we find significant heterogeneity from random coefficients.

The state dependence terms capture the individual's tendency to revisit the chain or location where she last shopped. Location state dependence can potentially generate spillover effects across chains within the same 4-digit zip code, but we find such effect to be small (yet statistically significantly different from 0). We find significant chain-level state dependence both in the sense that consumers tend to choose chains they have (ever) shopped at in the past, and in the sense that consumers tend to revisit the chain she shopped last time. This state-dependence pattern is consistent with limited awareness present in the market. Indeed, because consumers do not possess information about all the chains in the market, they tend to revisit chains they bought from recently or (to a lesser extent) in the more distant past. Although the spillover effects generate some degree of complementarity between channels, we find in simulations that offline and online channels are still net substitutes, in line with our descriptive evidence.

Finally, the tendency to shop online depends on the number of online-shopping trips in the past.

Consistent with the descriptive evidence, our structural estimates confirm that such an experience effect is important in explaining the growth of e-commerce. Shopping online once increases the utility of shopping online in subsequent periods. This effect is equivalent to about 5.5 years of the growth of e-commerce in magnitude (from the online $\times$ year coefficient). Further online-purchase experience has decreasing marginal effects. Given that we estimate this experience effect while controlling for rich heterogeneity, we interpret it as reminiscent of consumer learning. Past online-shopping experiences might help consumers gain confidence in the quality of service or the ability to pick products of the right fit. With learning, the growing shares of e-commerce have not yet reached a stationary level within the sample. These learning effect estimates allow us to simulate the long-run, stationary market where learning has completed, and assess the full long-run impact of e-commerce.

**Price elasticities.** Our model estimates imply that own-price elasticities are between -1.62 and -1.72 across chains. Cross elasticities vary across chains and are generally higher for popular chains with a higher number of stores. Appendix Table 9 presents the full set of elasticities.

We conjecture that cross elasticities are driven by the proximity of the chains to each customer, and we further quantify this dependence on distance. For each pair of chains, we bin consumers into those located within 1 km to *both* chains, between 1 and 2 km to *both* chains, and so on. Figure 8 presents the distribution of cross-price elasticities across distance bins and finds that, as distance increases, cross-price elasticities drop in distribution. Substitution between chains, and hence competition, is driven by the spatial distribution of stores in the market.

**Supply-side estimates.** We back out the average marginal costs for each chain from the supply-side model, which also allows us to infer profit margins for each chain. All marginal costs are positive. Implied gross profit margins (before tax) range from 51% to 60%. Anecdotal evidence suggests that some Dutch entrepreneurs in the clothing business target margins between 57% to 67%.[27] Our estimates are in line with these anecdotes. Appendix Table 10 reports the average marginal costs and profit margins for the sample period of 2016-2018.

[27] Source: https://www.higherlevel.nl/forums/topic/26371-factor-marge-in-de-kledingbranche-hoe-zit-dat-nu-precies/, extracted in October 2019.

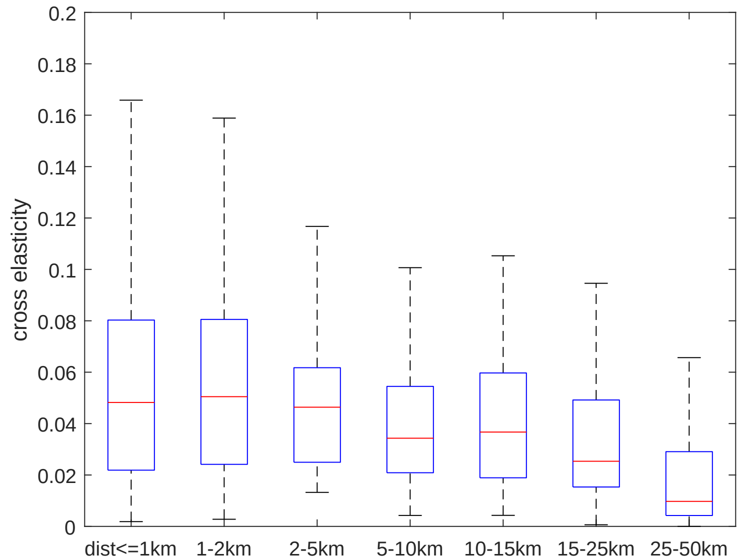



Figure 8: The distribution of cross-price elasticities as a function of distance

**Notes**: Each box shows the distribution of cross-price elasticities between pairs of top-fifteen chains, conditional on the stores of the two chains to a given consumer being located within a distance bin. For example, the first box shows the distribution of cross elasticities among consumer-chain pairs where each chain's closest store is within 1 km to the consumer.

# 6 The consumer gains from e-commerce

## 6.1 Consumer gains from e-commerce

What do these estimates imply for the size and nature of consumer gains from e-commerce? We compute the inclusive value (i.e., the expected maximum utility from the set of available store choices) under various counterfactual scenarios, as listed in Table 7. We use an equivalent variation approach to measure consumer gains from e-commerce. We compute the price increase from the baseline scenario (scenario 1 in Table 7) that makes each consumer indifferent to the counterfactual scenario without e-commerce (with new equilibrium prices, scenario 5 in Table 7). We express these price increases in percent terms relative to the current prices. In other words, the equivalent variation for the counterfactual market without e-commerce (scenario 5) would be the percentage increase from observed prices that would equate a consumer's inclusive value,

$$\Delta_{it}^{1-5} = \arg\min_{\Delta} \left| \underbrace{\log\left(\sum_{j,c} \exp\left(u_{ijct}\left(P_{jt}\cdot(1+\Delta)\right)\right)\right)}_{\text{inclusive value scenario 1 at equivalent prices}} - \underbrace{\log\left(\sum_{j,c\neq 1} \exp\left(u_{ijct}\left(P_{jt}^{*}\right)\right)\right)}_{\text{inclusive value scenario 5}} \right|, \tag{16}$$

where $P_{jt}$ are observed prices and $P_{jt}^{*}$ are equilibrium prices charged in absence of e-commerce. Note that $\Delta_{it}^{1-5}$ is individual-time specific. We then take the mean of $\bar{\Delta}_{i}^{1-5} = \frac{1}{T_i}\sum_t \Delta_{it}^{1-5}$ for each individual, and report the value of e-commerce as sample statistics on $\bar{\Delta}_{i}^{1-5}$.

Next, we decompose the welfare gain from e-commerce (or the welfare loss from the lack of e-commerce) into four terms. The first term focuses on the gains from eliminating the transportation costs. We examine in scenario 2 a counterfactual world where the consumer retains all online-shopping options, but needs to incur transportation costs purchasing from them. This scenario retains other attributes of e-commerce including new online retailers, but isolates the gains from convenience in the form of avoiding transportation costs. To implement this counterfactual experiment, we take equation (16) and replace utilities in the second inclusive value term by utilities in the counterfactual world where consumers derive disutility on distance when shopping from any alternative.[28] This solves for $\Delta_{it}^{1-2}$.

The second term focuses on gains from high-variety online retailers. In our context, we expect that the gains from variety stem predominantly from the addition of new online retailers (large online retailers in this market are multi-brand retailers with a large set of brands and items). Starting from scenario 2 in Table 7, we take away these two retailers as well as "other online retailers" (which collectively also sell a large assortment), and examine the additional welfare loss. We next calculate the equivalent price difference $\Delta_{it}^{2-3}$.

Third, e-commerce presents different shopping experiences that consumers might value (or might not value) for traditional fashion chains. These shopping experiences might include interfaces such as design of the webpage and email notifications, or consumers' uncertainty of the product due to the lack of fit and feel, or perhaps due to increased product variety from an existing omni-channel retail chain. Our model accounts for these benefits using retail-channel fixed effects, trends, as well

[28] This includes shopping at online-only retailers, of which the distance is set at the average distance to all branded chains for a given consumer.

Table 7: Counterfactual scenarios and decomposition of value.

| |
|---|
| 1. Baseline: Keep current online retailers. Keep prices as observed. Distance to online chains equal to 0. |
| ↑ Value from the elimination of travel costs, $\Delta_{it}^{1-2}$ ↓ |
| 2. No savings of travel costs: Keep all online retailers. Keep prices as observed. Set distance to online chains equal to closest offline store of that chain. |
| ↑ Value from high-variety online retailers, $\Delta_{it}^{2-3}$ ↓ |
| 3. Remove only pure-play online retailers. Keep prices as observed. Set distance to other online chains equal to closest offline store of that chain. |
| ↑ Value from other online retailers, $\Delta_{it}^{3-4}$ ↓ |
| 4. Remove all e-commerce options. Keep chain level prices as observed. |
| ↑ Value from price competition, $\Delta_{it}^{4-5}$ ↓ |
| 5. Remove all e-commerce options. Recalculate prices to pure offline Bertrand-Nash equilibrium. |

as logit errors.[29] To quantify the extent to which consumers value these attributes, we start from scenario 3 and remove all online remaining online stores. We next compute the price differences $\Delta_{it}^{3-4}$.

Fourth, in addition to these three sources of gains, e-commerce also alters spatial market power and prices. We use the static pricing game from Section 4.3 to simulate the counterfactual price equilibrium when e-commerce does not exist. Specifically, we hold both demand and marginal cost estimates $\text{mc}_{jt}$ fixed, we start with scenario 4 and iterate the first-order conditions (15) until convergence at the new price equilibrium.[30] Taking these prices as a point of departure, we next solve for the price discount that makes consumers indifferent between scenario 4 and 5, $\Delta_{it}^{4-5}$. This

[29] As is well-known in the literature (Petrin, 2002; Ackerberg and Rysman, 2005; Quan and Williams, 2018), the existence of IID type-1 extreme value errors *alone* will create welfare gains because the addition of e-commerce almost doubles the number of independent error draws. These errors might capture unobserved product characteristics, but might also represent idiosyncratic reasons to visit a webpage or a store location. The following decomposition includes the contribution of such logit errors.

[30] We solve for the new average prices across all consumers and time periods, and compute percent price changes relative to the baseline.

term should be similar to the average price difference between current prices and recalculated prices to reflect the absence of e-commerce.

## 6.2 The total gains from e-commerce

We now present the results from these counterfactual contrasts. Because e-commerce grows rapidly over time in the sample period, one should expect that consumers value e-commerce differently over time. As such, we present the gains from e-commerce over three time periods. We start with computing our counterfactuals exclusively for the period from 2007 to 2009. Next, we compute the value of e-commerce from 2016 to 2018. The difference between these two cases consists of more supply of e-commerce and different consumer experiences towards it (reflected in the state variables). In addition to the two, we note that online shopping is still on its expansion path, and our estimates have shown significant consumer learning. We thus simulate the gains from e-commerce in a stationary market where consumers have fully learned about e-commerce.[31]

Table 8 presents the average, standard deviation, and quartiles for the gains from e-commerce. We find that consumer gains from e-commerce is still modest during 2007-2009: taking away e-commerce (and allowing prices to adjust to the new equilibrium) is payoff-equivalent of a 8.7% price increase in the current situation.[32] In contrast, the gains from e-commerce are much higher during 2016-2018, valued at 22.6% of the price. In other words, consumers would prefer the presence of e-commerce over a counterfactual world with only offline stores and are willing to tolerate a price increase of up to 22.6% from all retailers. Finally, in the hypothetical world where all consumers have learned about e-commerce, taking away the online channel would have been equivalent to a 65.9% price increase, much higher than even the most recent sample period. This difference is consistent with the limited degree of participation and experience-given-participation on the online channel by 2018 (see, e.g., Bronnenberg and Ellickson, 2015 for independent support).

[31]Specifically, we take the data during 2016-2018 but set the past number of online-shopping trips to 16 so that the experience effect falls into the highest category in our model.

[32]To evaluate this counterfactual's precision, we bootstrap the standard errors in these reported gains and find that our estimates are precise. In particular, we draw demand-side parameters from a normal distribution with the mean at the parameter estimates and the variance-covariance matrix computed as the inverse Hessian. For each parameter draw, we use the model to compute the equivalent variation. Finally, we take standard deviation of the equivalent variation across 25 parameter draws as a measure of the standard error of the counterfactual.

Table 8: Gains from e-commerce and decomposition

Panel A: 2007-2009

| | mean surplus | (std err) | 25% | (std err) | 75% | (std err) |
|---|---|---|---|---|---|---|
| equiv. variations: remove all online | 0.087 | 0.010 | 0.051 | 0.006 | 0.102 | 0.011 |
| ... from convenience | 0.030 | 0.004 | 0.011 | 0.002 | 0.036 | 0.004 |
| ... from online-only rtl. | 0.024 | 0.003 | 0.010 | 0.001 | 0.030 | 0.003 |
| ... from online channel of existing rtl. | 0.009 | 0.001 | 0.004 | 0.000 | 0.010 | 0.001 |
| ... from price changes | 0.024 | 0.003 | 0.022 | 0.003 | 0.026 | 0.003 |

Panel B: 2016-2018

| | mean surplus | (std err) | 25% | (std err) | 75% | (std err) |
|---|---|---|---|---|---|---|
| equiv. variations: remove all online | 0.226 | 0.026 | 0.116 | 0.015 | 0.277 | 0.032 |
| ... from convenience | 0.056 | 0.006 | 0.020 | 0.002 | 0.068 | 0.007 |
| ... from online-only rtl. | 0.054 | 0.005 | 0.016 | 0.002 | 0.073 | 0.007 |
| ... from online channel of existing rtl. | 0.056 | 0.005 | 0.019 | 0.002 | 0.069 | 0.007 |
| ... from price changes | 0.059 | 0.011 | 0.053 | 0.010 | 0.065 | 0.012 |

Panel C: Counterfactual – fully learned about e-commerce

| | mean surplus | (std err) | 25% | (std err) | 75% | (std err) |
|---|---|---|---|---|---|---|
| equiv. variations: remove all online | 0.659 | 0.061 | 0.494 | 0.068 | 0.867 | 0.069 |
| ... from convenience | 0.143 | 0.017 | 0.086 | 0.011 | 0.176 | 0.021 |
| ... from online-only rtl. | 0.164 | 0.019 | 0.084 | 0.010 | 0.222 | 0.027 |
| ... from online channel of existing rtl. | 0.229 | 0.017 | 0.123 | 0.015 | 0.311 | 0.025 |
| ... from price changes | 0.123 | 0.013 | 0.114 | 0.019 | 0.156 | 0.026 |

**Notes:** These tables present the decomposition of welfare gains from e-commerce. Mean consumer surplus is the average equivalent variation, $\bar{\Delta}_i$, for each $i$ in the sample period. Panel A focuses in the period of 2007-2009, panel B the period of 2016-2018, and panel C simulates the counterfactual world where consumers have fully learned about e-commerce.

## 6.3 Decomposition of the gains

Next, we decompose the gains from e-commerce into gains from (1) avoidance of travel costs (convenience), (2) new pure-play retailers, (3) online channels from existing retailers, and (4) price competition. First, how much do consumers directly benefit from the convenience of shopping online, in that this way of shopping does not incur transportation costs? We find that during 2016-2018, the gains from convenience are equivalent to 5.6% of the price, an order of magnitude larger than the previous finding. Convenience accounts for 25% of the total gains from e-commerce. This finding suggests that an important value of the online channel is to reduce shopping costs for the many consumers who live at a non-negligible distance from their favorite stores – a shortcoming of the traditional brick-and-mortar format.

In addition, how much do consumers benefit from variety? We find that consumers gain from the presence of new online retailers by an equivalent variation of 5.4% of the price. Such gains could be a result of these new online retailers (usually multi-brand) carrying much larger assortments than traditional ones and allowing consumers access to brands that are not available in traditional single-brand chains (such as H&M). Therefore, we interpret this benefit as a form of gains from variety. Further, consumers also benefit from the addition of online channels in existing chains (on top of convenience) by an equivalent variation of 5.6% of the price. This gain might capture the direct value of using the website, for instance, as a more efficient way of searching for products. This gain also includes the choice-set expansion effect present in the logit model. Overall, our findings for the gains from variety accord directionally with the literature on the gains from product variety (Brynjolfsson et al., 2003; Quan and Williams, 2018) and retailer variety (Dolfen et al., 2020), albeit that we report not nearly the same amount of benefit from variety as Brynjolfsson et al. (2003).

Finally, e-commerce expands the effective radius of purchase for consumers and facilitates price competition between retailers, thus bringing consumers an additional benefit from having lower prices. We simulate counterfactual market equilibirum prices without all online-shopping options. For an average consumer, prices would have been 5.9% higher without e-commerce.[33]

[33] Table 11 further demonstrates the differential impact on prices across retailers.

### 6.4 Distribution of the gains from convenience

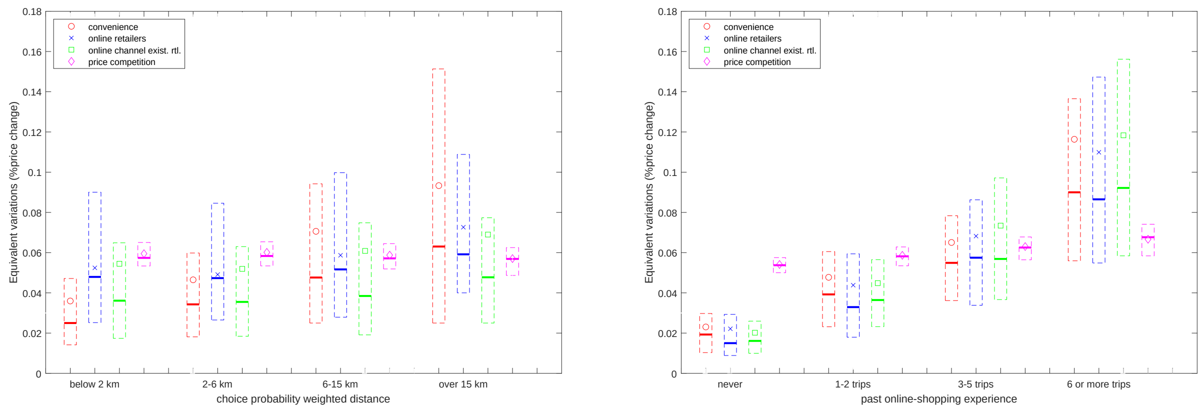
Finally, do different consumers value e-commerce differently, and if so, what do they value? We cluster consumers with similar characteristics and examine, for each group of consumers, the distribution of the gains from convenience, online retailers, online channels of existing retailers, and price competition.

We first examine whether e-commerce benefits urban consumers differently than those in less-urbanized areas. Panel A of Figure 9 shows the mean, median, and inter-quartile range of $\overline{\Delta}_i$ for each of the four components, and separately for consumers with a different weighted average distance to stores. We use choice probabilities to compute weighted average distances so that we put higher weights on frequently-visited chains. We find that the gains from convenience are higher, the further the distance between consumers and stores. For example, consumers within 2 km to chains value convenience at about 2% of the price, whereas those beyond 15 km away value convenience more than twice as much. This contrast suggests that local market conditions play an important role in the value consumers place on the new channel. We also find significant heterogeneity in the gains conditional on the geographic group, plausibly from heterogeneity in online-shopping experiences or further differences in local choice sets.[34] Gains from online retailers also increases with the distance between consumers and stores, but compared to the convenience component, it is more similar across geographic groups. On the other hand, gains from competition are close to homogeneous across consumers both across and within each location bin (the remaining heterogeneity comes from differences in price responses across retailers). This finding implies that consumers who do not yet value e-commerce much (i.e. those at the bottom of the distribution) would mostly benefit from the expansion of e-commerce through price changes. This point can also be seen from Panel B of the figure. There, we show the distribution of the gains by consumers with different online-shopping experience. Those with zero online-shopping experience tend to shop offline. As a result, a large fraction of the gains from e-commerce come from lower prices. On the other hand, those who have shopped online many times benefit much more from both convenience and variety, relative to their gain from the same price decline. Therefore, gains from price competition is a preference externality from those who value shopping online to those who do not (yet) value it.

[34] We also examine heterogeneity across demographic groups such as age and gender, and find limited heterogeneity there.

Figure 9: Heterogeneity in the gains from convenience, variety, and price competition

Panel A: heterogeneity across locations Panel B: across online-shopping experiences




**Notes**: Distribution of $\overline{\Delta}_i$ for each of the four components: convenience, value of online retailers, value of online channel of existing retailers, and gains from price competition. This figure reports on gains from e-commerce between 2016 and 2018. For each group, the marker represents the mean, the horizontal bar represents the median, and the dashed box represents the interquartile range.

# 7 Conclusions

We study how travel distance to brick-and-mortar stores affects consumers' choices of chains and channels and quantify the value of convenience (as a reduction in transportation costs) provided by e-commerce. We leverage granular individual-level panel data in the Dutch retail apparel market from 2007 to 2018, which allows us to observe consumer locations to narrowly-defined geographic areas. To measure consumers' sensitivity to distance, we exploit a spatial difference-in-differences identification strategy at the individual level. We combine this strategy with a model of consumer choices of chains and channels to measure and decompose the value of e-commerce.

Our main result shows that convenience accounts for a large component of the value of e-commerce. The gains from convenience are heterogeneous across consumers, who differ in preferences, locations, demographics, and shopping experience. This result implies considerable value in providing convenience to consumers (such as building stores or offering convenient retail services), and also that the value of these strategies might differ across customer segments. In addition, the convenience offered by e-commerce reduces local market power, and therefore consumers also benefit from intensified price competition among retailers. We demonstrate that the benefit of lower prices affects all consumers, not just those who shop online.

For future research, a limitation of the paper is that we hold the distribution of store locations fixed in our counterfactual analysis. In the sample period, most of the leading chains have built more stores rather than pulling back from the market, which seems to justify our approach to not focus on store closings due to e-commerce in our simulations. Nevertheless, casual observations suggest that during 2018-2019, the US retail market has seen considerable store closing or even chain-level exits. The impact of e-commerce and how it is influenced by the endogenous entry and exit of stores remains an open question for future studies.

# References


Ackerberg, Daniel A and Marc Rysman (2005), 'Unobserved product differentiation in discrete-choice models: estimating price elasticities and welfare effects', *RAND Journal of Economics* **36**(4), 771–789.

Bai, Jie (2016), 'Melons as lemons: Asymmetric information, consumer learning and seller reputation'.

Bell, David R, Santiago Gallino and Antonio Moreno (2017), 'Offline showrooms in omnichannel retail: Demand and operational benefits', *Management Science* **64**(4), 1629–1651.

Berry, Steven, James Levinsohn and Ariel Pakes (1995), 'Automobile prices in market equilibrium', *Econometrica: Journal of the Econometric Society* pp. 841–890.

Bronnenberg, Bart J (2015), 'The provision of convenience and variety by the market', *The RAND Journal of Economics* **46**(3), 480–498.

Bronnenberg, Bart J. and Paul B. Ellickson (2015), 'Adolescence and the path to global retail', *Journal of Economic Perspectives* **29**(4), 113–134.

Brynjolfsson, Erik, Yu Hu and Michael D Smith (2003), 'Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers', *Management Science* **49**(11), 1580–1596.

Cavallo, Alberto (2017), 'Are online and offline prices similar? evidence from large multi-channel retailers', *American Economic Review* **107**(1), 283–303.

Dolfen, Paul, Liran Einav, Peter J Klenow, Benjamin Klopack, Jonathan D Levin, Laurence Levin and Wayne Best (2020), 'Assessing the gains from e-commerce'.

Dubé, Jean-Pierre, Günter J Hitsch and Peter E Rossi (2009), 'Do switching costs make markets less competitive?', *Journal of Marketing research* **46**(4), 435–445.

Dubé, Jean-Pierre, Günter J Hitsch and Peter E Rossi (2010), 'State dependence and alternative explanations for consumer inertia', *The RAND Journal of Economics* **41**(3), 417–445.

Ellickson, Paul B and Paul LE Grieco (2013), ‘Wal-mart and the geography of grocery retailing’, *Journal of Urban Economics* **75**, 1–14.

Ellickson, Paul B, Paul LE Grieco and Oleksii Khvastunov (2020), ‘Measuring competition in spatial retail’, *The RAND Journal of Economics* **51**(1), 189–232.

Farronato, Chiara and Andrey Fradkin (2018), The welfare effects of peer entry in the accommodation market: The case of airbnb, Technical report, National Bureau of Economic Research.

Forman, Chris, Anindya Ghose and Avi Goldfarb (2009), ‘Competition between local and electronic markets: How the benefit of buying online depends on where you live’, *Management science* **55**(1), 47–57.

Frechette, Guillaume R, Alessandro Lizzeri and Tobias Salz (2019), ‘Frictions in a competitive, regulated market: Evidence from taxis’, *American Economic Review* **109**(8), 2954–92.

Gentzkow, Matthew (2007), ‘Valuing new goods in a model with complementarity: Online newspapers’, *The American Economic Review* **97**, 713–744.

Goldfarb, Avi and Catherine Tucker (2019), ‘Digital economics’, *Journal of Economic Literature* **57**(1), 3–43.

Hortaçsu, Ali and Chad Syverson (2015), ‘The ongoing evolution of us retail: A format tug-of-war’, *Journal of Economic Perspectives* **29**(4), 89–112.

Houde, Jean-François (2012), ‘Spatial differentiation and vertical mergers in retail markets for gasoline’, *American Economic Review* **102**(5), 2147–82.

Jo, Yoon J, Misaki Matsumura and David E Weinstein (2019), ‘The impact of e-commerce on relative prices and consumer welfare’.

Kamakura, Wagner A and Gary J Russell (1989), ‘A probabilistic choice model for market segmentation and elasticity structure’, *Journal of marketing research* **26**(4), 379–390.

Li, Kathleen T (2019), ‘Statistical inference for average treatment effects estimated by synthetic control methods’, *Journal of the American Statistical Association* pp. 1–16.

Petrin, Amil (2002), ‘Quantifying the benefits of new products: The case of the minivan’, *Journal of political Economy* **110**(4), 705–729.

Pozzi, Andrea (2013), ‘The effect of internet distribution on brick-and-mortar sales’, *RAND Journal of Economics* **44**(3), 569–583.

Quan, Thomas W and Kevin R Williams (2018), ‘Product variety, across-market demand heterogeneity, and the value of online retail’, *The RAND Journal of Economics* **49**(4), 877–913.

Shriver, Scott and Bryan Bollinger (2015), ‘A structural model of channel choice with implications for retail entry’.

Shriver, Scott and Bryan Bollinger (2020), ‘Demand expansion and cannibalization effects from retail store entry: A structural analysis of multi-channel demand’.

Smith, Howard (2004), 'Supermarket choice and supermarket competition in market equilibrium', *The Review of Economic Studies* **71**(1), 235–263.

Smith, Howard (2006), 'Store characteristics in retail oligopoly', *The RAND Journal of Economics* **37**(2), 416–430.

Wang, Kitty and Avi Goldfarb (2017), 'Can offline stores drive online sales?', *Journal of Marketing Research* **54**(5), 706–719.

Xin, Yi (2018), 'Asymmetric information, reputation, and welfare in online credit markets'.

Yoganarasimhan, Hema (2013), 'The value of reputation in an online freelance marketplace', *Marketing Science* **32**(6), 860–891.

Zervas, Georgios, Davide Proserpio and John W Byers (2017), 'The rise of the sharing economy: Estimating the impact of airbnb on the hotel industry', *Journal of marketing research* **54**(5), 687–705.

Zhang, Dennis, Hengchen Dai, Lingxiu Dong, Qian Wu, Lifan Guo and Xiaofei Liu (2018), 'The value of pop-up stores in driving online engagement in platform retailing: Evidence from a large-scale field experiment with alibaba'.

Zheng, Fanyin (2016), 'Spatial competition and preemptive entry in the discount retail industry', *Columbia Business School Research Paper* (16-37).

# Appendix

## A Construction of the price index

We construct prices indices at the chain-month level to measure overall price levels across products. One way to construct this index is to simply compute the average purchase price in a given chain-month. However, we only observe price conditional on purchase, and one might worry about a selection problem on *who* purchases the product, as well as on unobserved characteristics of the product purchased. We proceed to construct a price index that is net of these unobserved demand shifters. In particular, we observe *whether* a purchased product is on a price discount, along with its price. We will leverage this data advantage to project price variations into discount frequency and depth variations. The underlying assumption of our approach is that the same discounts are offered to all consumers shopping for products of the same characteristic, and thus, discounts are exogenous to unobserved demand shocks conditional on individual-chain fixed effects and observed demographics and product characteristics. Meanwhile, this assumption is less likely to hold for the purchase price because the price itself is selected by the individual shopper.

To implement this idea, we first estimate two hedonic regressions, of price and discount incidence on year $y$, month-of-the-year $m$, observed product characteristics, consumer demographic variables, and consumer-chain fixed effects. For consumer $i$ who purchases item $r$ at chain $j$ in month $t$, we specify

$$\log\left(\text{price}_{ijrt}\right) = p_{j0} + \tau^0_{jy(t)} + \tau^0_{jm(t)} + \left(\tau^1_{jy(t)} + \tau^1_{jm(t)}\right) \cdot \text{discount}_{ijrt} + x^p_{jr}\beta^p_1 + z^p_{it}\beta^p_2 + \alpha^p_{ij} + \omega_{ijrt} \tag{17}$$

where the dummies $\tau^0_{jy(t)}$ and $\tau^0_{jm(t)}$ capture year and month level variations in the regular price for chain $j$, and the dummies $\tau^1_{jy(t)}$ and $\tau^1_{jm(t)}$ capture year- and month variations in discount depth. Note that these effects are net of observed product characteristics $x^p_{jr}$ (brand and product type), observed demographics $z^p_{it}$, and consumer-chain fixed effects $\alpha^p_{ij}$. Similarly, we also estimate

$$\text{discount}_{ijrt} = d_{j0} + \eta_{jy(t)} + \eta_{jm(t)} + x^d_{jr}\beta^d_1 + z^d_{it}\beta^d_2 + \alpha^d_{ij} + \upsilon_{ijrt} \tag{18}$$

to obtain $\eta_{jy(t)}$ and $\eta_{jm(t)}$ as year- and month-of-the-year variations in the discount frequency. Estimating both Equation (17) and (18), we then construct the price index as

$$\log\left(P_{jt}\right) = \hat{p}_{j0} + \hat{\tau}^1_{jy(t)} \times \hat{\eta}_{jy(t)} + \hat{\tau}^1_{jm(t)} \times \hat{\eta}_{jm(t)} \tag{19}$$

where we explicitly concentrate on only the chain-average price level $\hat{p}_{j0}$ and the over-time variations in discount depth and discount frequency.

# B Do store locations and price discounts target local consumer demographics?

To identify consumers' distance sensitivity, the key assumption is that, within the coverage area of a store, consumers are dispersed and the store cannot locate precisely to target individual consumer characteristics or unobserved demand (beyond targeting the demographics and unobserved demand of an area). Recall that we show in Section 3 that those who are close to a newly-entered store do not exhibit a different time trend to shop at the chain, compared to those who are further away from the store, supporting this identifying assumption. We now complement this exercise and demonstrate that the store's location indeed does not target observed demographics beyond the broad demographics in a region.

We first take a 30% random sample of the balanced individual-level data at the individual-chain-month level (i.e. including no-purchase occasions). We combine this sample with 4-digit zipcode level average income and total population (from 2007 to 2014), which we obtain from the census. This exercise brings us to a balanced panel where we can examine whether the presence of the chain in the consumer's local zipcode (5-digit zipcode) is explained by the consumer's demographics or the demographics of a broader area. Denote $\overline{\text{inc}}_{jt}$ as a measure of the average income of the chain's customer base, here constructed as expenditure-weighted average income from the purchase panel. We now estimate

$$100 \times \text{store\_in\_zip5}_{ijt} = b_0 \text{hhinc}_{it} + b_1 \text{hhinc}_{it} \times \overline{\text{inc}}_{jt} + b_2 \text{zipinc}_{m(i)t} + b_3 \text{zipinc}_{m(i)t} \times \overline{\text{inc}}_{jt} + \delta_{ij} + \lambda_{jt} + \varepsilon_{ijt} \tag{20}$$

where $\text{store\_in\_zip5}_{ijt}$ indicates 1 if the closest store of $j$ to a given consumer $i$ is in the same five-digit zipcode as the consumer, $\text{hhinc}_{it}$ is the household income of $i$ and $\text{zipinc}_{m(i)t}$ is the average household income in the 4-digit zipcode of $i$, $m(i)$. If the chain selectively enters into markets (4-digit zipcodes) with local income matching its typical clientele, we should expect $b_3$ to be positive. In addition, if the chain further targets granular clusters of consumers within the 4-digit zipcode, one should expect $b_1$ to be positive.

We present the estimates in the first column of Table 1. We find that given the set of controls, store location is correlated with the interaction between 4-digit zipcode level income and the income of the chain's clientele, but the correlation is small in magnitude. Suppose Chain A caters to customers with an average of €40,000 annual income and Chain B's customers have an income of €20,000. When the average income of a market grows by €1,000, the positive coefficient $\hat{b}_3$ suggests that Chain A will be more likely to enter in this market than Chain B – consistent with the conjecture that chains selectively enter into markets that are similar to their own clientele. However, compared to Chain B, Chain A is $20 \times 1 \times \hat{b}_3 = 20 \times 0.000053 = 0.0011$ percentage points more likely to enter this market, or 0.4% relative to the baseline entry probability at 0.25 percent-

Appendix Table 1: Targeting of store location and price discounts

| | store in zip5 | discount |
|---|---|---|
| household income | 0.001093*** | -0.088312 |
| | (0.000) | (0.093) |
| ... X clientele income | -0.000036*** | 0.001237 |
| | (0.000) | (0.003) |
| average zip4 income | -0.007005*** | -0.271892 |
| | (0.001) | (0.193) |
| ... X clientele income | 0.000053*** | 0.009949* |
| | (0.000) | (0.006) |
| zip4 population | 0.013308*** | 0.132689 |
| | (0.002) | (0.139) |
| individual-chain FE | Yes | Yes |
| chain-year FE | Yes | Yes |
| month FE | Yes | Yes |
| R-squared | 0.96 | 0.37 |
| observations | 15923662 | 566730 |

**Notes:** Column 1 reports regression results of Equation (20). Column 2 focuses on the sample of consumer purchases and reports whether discounts are targeted to local income, using the same set of constrols as Equation (20). The dependent variables are percentage points and the income variables are in thousand euros.

age points. This estimate suggests that store locations do target to the average local income but the degree of targeting is negligible.

More importantly, we find that store locations do not target individual income within the local 4-digit zipcode market. For a given consumer, her income increasing by €1,000 will predict that she is 0.3% *less* likely to be close to Chain A, the high-end chain. Where the sign might be counter-intuitive, we note that the magnitude of this effect is economically negligible. We conclude that we do not find evidence that store locations target to changes in customer income within a 4-digit zipcode.

We further examine whether discounts are targeted to local markets, in a similar way. Specifically, we take the sample of consumer purchases and estimate whether a purchase contains a product on discount, on the same set of variables and fixed effects as Equation (20). We find that an increase in the average income of the 4-digit zipcode is associated with fewer discounts: A €1,000 increase in the average income predicts 0.27 *percentage point* decrease in the share of discount (and it is statistically insignificant), or 0.5% relative to the 47 percentage point baseline discount level. Similarly, the interaction with chain's clientele characteristics also return a small effect. Further, given the average income at the 4-digit zipcode level, individual consumer income and its interaction with the chain's clientele do not predict the share of discount this consumer purchases in a statistically or economically significant way. We conclude that little price discounts were assigned in a targeted fashion.

# C Additional descriptive statistics

**Diversity of retail formats.** Table 2 shows summary statistics at the consumer-chain-month level, using the full (unbalanced) sample and taking into account consumer-months without purchase. We examine, for the overall sample and then by retail format, the frequency of shopping incidence, expenditure given the incidence, frequency of shopping online, and shopping distance if the consumer shops offline. Consumers travel further for branded chains (e.g. H&M) and for specialty stores (e.g. The Shoe Factory). The share of online sales are higher for branded chains and department stores.

Appendix Table 2: Summary of expenditure, variety, channel, and shopping distance

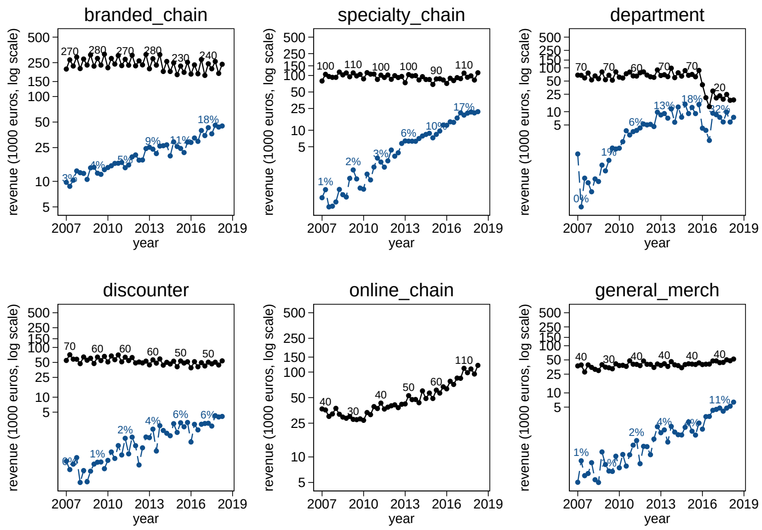
| | branded | department | discounter | general | online | specialty | all formats |
|---|---|---|---|---|---|---|---|
| purchase offline | 0.141 | 0.035 | 0.092 | 0.091 | 0.000 | 0.057 | 0.292 |
| purchase online | 0.016 | 0.003 | 0.002 | 0.003 | 0.032 | 0.005 | 0.055 |
| number of chains purchased from | 4.603 | 0.205 | 1.125 | 0.511 | 1.228 | 2.148 | 9.922 |
| expenditure if purchase | 62.725 | 59.674 | 22.307 | 17.207 | 67.233 | 60.609 | 21.562 |
| distance of offline purchase | 10.521 | 9.329 | 3.226 | 4.564 | 0.000 | 26.081 | 1.573 |
| observations (HH-chain-year-month) | 1,111,402 | 1,111,402 | 1,111,402 | 1,111,402 | 1,111,402 | 1,111,402 | 1,111,402 |

**Notes:** This table reports offline and online incidence at the monthly level, expenditure given shopping incidence, number of distinct chains shopped at, and shopping distance if the consumer shops offline.

**Growth of e-commerce across retail formats.** Figure 1 presents the growth of total and online expenditure across retail formats. One finds that while all formats growth in the total online expenditure (except for department stores, which saw exit of a major player in 2016), the within-format growth rate of online expenditure is lower than the total growth rate. This contrast is explained by the composition change across formats – in particular, online chains take an increasingly significant role.

**Choice of variety.** We examine the number of chains a consumer purchases from, and the composition of expenditure among these chains, in each trip and within various time windows. Table 3 shows the average number of chains a consumer purchases from and the share of expenditure at the top chain, on the unit of analysis of consumer-date, consumer-month, and consumer-quarter. We find that there are limited multi-chain visits within a shopping date: 87.5% trips are only associated with purchasing from one chain and the top chain takes 96% expenditure on that day.

Over a wider time frame, however, we do observe that the consumer purchases from more than one chain: 42% of months and 59% of quarters with positive expenditure are associated with at least two chains. Aggregating across time, only 13% consumers ever bought from only one chain. This indicates significant choice of variety chosen by the consumer, but importantly not on the same date.



Appendix Figure 1: Growth of e-commerce by format

**Note:** See note of Figure 3.

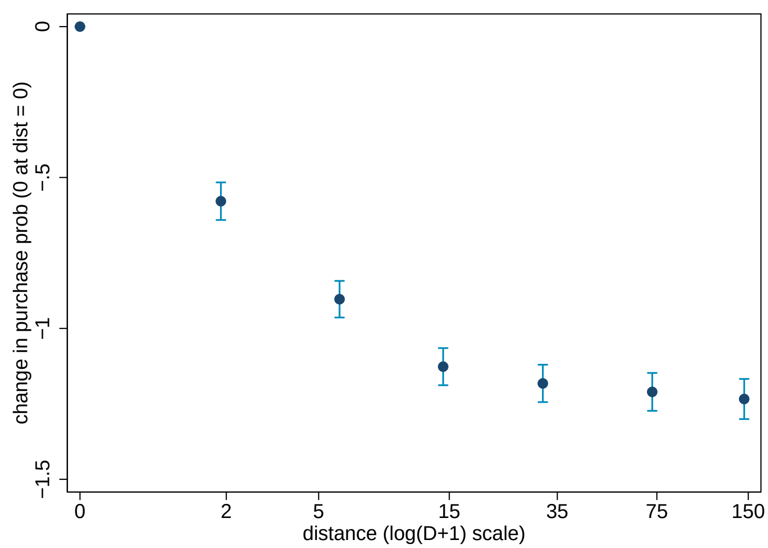
Appendix Table 3: Choice of variety in various time window definitions

| | same day | same month | same quarter | entire sample |
|---|---|---|---|---|
| purchase from 1 chain | 0.868 | 0.575 | 0.575 | 0.139 |
| purchase from 2 chains | 0.106 | 0.231 | 0.231 | 0.092 |
| purchase from 3 chains | 0.020 | 0.102 | 0.102 | 0.066 |
| purchase from 4+ chains | 0.006 | 0.093 | 0.093 | 0.703 |
| expd. share, chain of highest expd. | 0.954 | 0.827 | 0.827 | 0.387 |
| observations | 551,214 | 355,718 | 355,718 | 23,976 |

**Notes:** Number of distinct chains the HH purchases from, conditional on making a purchase in a given time window.

# D Additional results of the effect of distance

**Flexible functional form.** We present a flexible specification on the effect of distance. We estimate Equation (5) with the same set of control variables, but with a series of distance bins to capture the effect of distance in a flexible way. We find that the shape of the effect is concave, with the marginal effect of distance decreasing the further a consumer is away from the store. In addition, we find that the $\log\left(D_{ijt}+1\right)$ specification is almost exactly correct when the consumer is within 15km of the store, which is a range with the majority of offline purchases (See Figure 4). Beyond this range, the marginal effect of distance further declines.




Appendix Figure 2: Heterogeneous marginal effect of $\log\left(D_{ijt}+1\right)$

**Note:** This figure visualizes the estimates of a more flexible distance-sensitivity regression. The x-axis presents distance bins that are rescaled to the $\log(D+1)$ specification, and the y-axis is the marginal effect of distance for each bin.

**The effect of distance by-format.** We present estimation results of Equation (5) separately by retail format. While we find significant heterogeneity in the sensitivity to distance across retail formats, we consistently find no evidence that offline and online channels are net substitutes or complements.

Appendix Table 4: Sensitivity to distance (Y is purchase $\times$ 100): full table

Panel A: sensitivity to own distance: offline demand

| | offline: branded | dept. | disc. | gen. merch. | specialty |
|---|---|---|---|---|---|
| log(dist + 1) | -0.257*** | 0.069 | -0.713*** | -1.317*** | -0.183*** |
| | (0.013) | (0.141) | (0.041) | (0.122) | (0.020) |
| consumer-chain FE | Yes | Yes | Yes | Yes | Yes |
| chain-year FE | Yes | Yes | Yes | Yes | Yes |
| month FE | Yes | Yes | Yes | Yes | Yes |
| demographics | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.15 | 0.15 | 0.14 | 0.15 | 0.08 |
| observations | 28583331 | 1632373 | 6531595 | 4085050 | 12759106 |

Panel B: sensitivity to own distance: online demand

| | online: branded | dept. | disc. | gen. merch. | specialty |
|---|---|---|---|---|---|
| log(dist + 1) | 0.004 | 0.034 | 0.000 | 0.001 | -0.005 |
| | (0.002) | (0.021) | (0.002) | (0.009) | (0.004) |
| consumer-chain FE | Yes | Yes | Yes | Yes | Yes |
| chain-year FE | Yes | Yes | Yes | Yes | Yes |
| month FE | Yes | Yes | Yes | Yes | Yes |
| demographics | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.13 | 0.09 | 0.06 | 0.09 | 0.06 |
| observations | 28583331 | 1632373 | 6531595 | 4085050 | 12759106 |

**Notes:** Panel A presents the sensitivity of householdconsumer's shopping trip to offline storeschain choice to the distance to the chain's nearest store. The results are presented by retail formats: branded chains, department stores, discounters, general merchandizers, and specialty chains. Panel B presents the sensitivity of online shopping trips to the distance to the closest store. conditions on existing users, i.e. household has purchased from the focal chain before.

**Robustness.** One might be concerned that consumers or chains move into growing markets (or chains enter markets where it has growing demand). While these concerns are beyond the set of controls in Equation (5), we test for robustness of our empirical results by including chain-zipcode-year fixed effects or consumer zipcode-year fixed effects.

# E Additional figures and tables

Appendix Table 5: Sensitivity to distance (Y is purchase $\times$ 100): robustness check for all consumers

| | baseline (offl.) | (onl.) | + store zip-year FEs (offl.) | (onl.) | + HHzip-year FEs (offl.) | (onl.) |
|---|---|---|---|---|---|---|
| log(distance + 1) | -0.338*** | 0.003 | -0.326*** | 0.003 | -0.340*** | 0.003 |
| | (0.012) | (0.002) | (0.014) | (0.002) | (0.013) | (0.002) |
| consumer-chain FE | Yes | Yes | Yes | Yes | Yes | Yes |
| chain-year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| hhzip-year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| month FE | Yes | Yes | Yes | Yes | Yes | Yes |
| demographics | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.14 | 0.10 | 0.14 | 0.10 | 0.14 | 0.10 |
| observations | 54181026 | 54181026 | 54181019 | 54181019 | 54181026 | 54181026 |

**Notes:** This table contrasts estimates of Equation (5) with chain-zip4-year fixed effects (the main specification), chain fixed effects, or consumer zipcode-year fixed effects. See notes under Table 3.

Appendix Table 6: Sample selection

| | fraction of sample |
|---|---|
| chain identity not missing | 0.644 |
| individual location not missing | 0.999 |
| individual never moved | 0.831 |
| all of the above | 0.531 |
| observations | 2,267,772 |

**Notes:** This table reports our sample selection criteria.

Appendix Table 7: Store-exit effect on offline shopping incidence

| | distance change 5-20km | 1-5km | 0.1-1km | 0-0.1km | no store entry (placebo) |
|---|---|---|---|---|---|
| post entry | -0.143** | -0.029 | -0.131 | 0.001 | 0.021** |
| | (0.047) | (0.050) | (0.086) | (0.099) | (0.006) |
| pre trend | -0.008*** | -0.009*** | -0.005 | -0.007 | -0.008*** |
| | (0.002) | (0.002) | (0.004) | (0.004) | (0.000) |
| post trend | -0.003** | -0.002 | -0.004 | -0.016*** | -0.007*** |
| | (0.001) | (0.002) | (0.003) | (0.004) | (0.000) |
| R-squared | 0.09 | 0.09 | 0.15 | 0.15 | 0.16 |
| observations | 277,559 | 239,369 | 178,476 | 130,080 | 27,299,660 |

**Notes:** This table presents estimates of Equation (4), focusing on consumer-chain pairs with store exits and dividing the sample into groups where the exit has different impact on the travel distance of the customer. See notes below Table 2.

Appendix Table 8: Additional parameter estimates for the structural model
Panel C: chain-channel intercepts, trends, and $\Delta u$ for new shoppers

| | offl. | s.e. | onl. | s.e. | trend (year) | s.e. | std dev of rc | s.e. |
|---|---|---|---|---|---|---|---|---|
| branded chain 1 | 3.304 | 0.452 | -2.663 | 0.481 | -0.065 | 0.006 | 0.423 | 0.027 |
| branded chain 2 | 2.420 | 0.418 | -1.560 | 0.442 | -0.144 | 0.010 | 0.526 | 0.043 |
| branded chain 3 | 3.666 | 0.542 | -1.178 | 0.566 | -0.146 | 0.010 | 0.236 | 0.057 |
| branded chain 4 | 3.979 | 0.598 | 0.831 | 0.614 | -0.230 | 0.016 | 0.197 | 0.061 |
| branded chain 5 | 2.645 | 0.500 | -2.938 | 0.552 | -0.117 | 0.013 | 0.226 | 0.050 |
| branded chain 6 | 3.135 | 0.528 | -2.110 | 0.602 | -0.153 | 0.019 | 0.151 | 0.064 |
| branded chain 7 | 3.390 | 0.546 | -2.708 | 0.631 | -0.122 | 0.016 | 0.124 | 0.112 |
| other branded chains | 4.649 | 0.569 | -0.238 | 0.586 | -0.119 | 0.006 | 0.816 | 0.030 |
| department store 1 | 3.448 | 0.454 | -1.746 | 0.481 | -0.120 | 0.010 | 0.458 | 0.034 |
| department store 2 | 4.133 | 0.626 | -0.460 | 0.652 | -0.115 | 0.016 | 0.882 | 0.060 |
| discounter 1 | -0.536 | 0.192 | -Inf | 0.000 | -0.046 | 0.008 | 0.162 | 0.045 |
| discounter 2 | 1.011 | 0.380 | -Inf | 0.000 | -0.010 | 0.024 | 0.188 | 0.067 |
| other discounters | 0.054 | 0.228 | -7.120 | 0.327 | -0.059 | 0.007 | 1.036 | 0.024 |
| general merch. 1 | 0.694 | 0.273 | -5.296 | 0.321 | -0.050 | 0.007 | 0.223 | 0.033 |
| other general merch. | -0.677 | 0.240 | -6.857 | 0.326 | 0.036 | 0.008 | 1.130 | 0.035 |
| online retailer 1 | -Inf | 0.000 | 1.843 | 0.578 | -0.237 | 0.017 | 0.439 | 0.060 |
| online retailer 2 | -Inf | 0.000 | 2.339 | 0.691 | -0.187 | 0.027 | 0.123 | 0.093 |
| other online retailers | -Inf | 0.000 | 1.338 | 0.523 | -0.244 | 0.013 | 1.388 | 0.044 |
| specialty chain 1 | 2.021 | 0.377 | -3.311 | 0.420 | -0.119 | 0.010 | 0.395 | 0.044 |
| other specialty chains | 5.245 | 0.616 | -0.030 | 0.633 | -0.083 | 0.006 | 0.416 | 0.031 |

**Notes:** Panel C complements Panels A and B in Table 6, and presents chain-level estimates for intercepts (online and offline separately), time trends, and differences in utility for consumers never shopped at the chain. See notes in the main table.

Appendix Table 9: Average price elasticity matrix

| | (A) | (B) | (C) | (D) | (E) | (F) | (G) | (H) | (I) | (J) | (K) | (L) | (M) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (A) branded chain 1 | -1.62 | 0.11 | 0.12 | 0.07 | 0.12 | 0.11 | 0.12 | 0.12 | 0.10 | 0.12 | 0.11 | 0.12 | 0.12 |
| (B) branded chain 2 | 0.05 | -1.68 | 0.05 | 0.04 | 0.05 | 0.05 | 0.05 | 0.06 | 0.05 | 0.05 | 0.04 | 0.05 | 0.05 |
| (C) branded chain 3 | 0.05 | 0.05 | -1.67 | 0.03 | 0.05 | 0.05 | 0.05 | 0.05 | 0.04 | 0.05 | 0.04 | 0.05 | 0.05 |
| (D) branded chain 4 | 0.02 | 0.02 | 0.02 | -1.71 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.01 | 0.02 | 0.02 |
| (E) branded chain 5 | 0.03 | 0.03 | 0.03 | 0.02 | -1.69 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |
| (F) branded chain 6 | 0.02 | 0.02 | 0.02 | 0.01 | 0.02 | -1.70 | 0.02 | 0.02 | 0.01 | 0.02 | 0.01 | 0.02 | 0.02 |
| (G) branded chain 7 | 0.02 | 0.02 | 0.02 | 0.01 | 0.02 | 0.02 | -1.71 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| (H) department store 1 | 0.09 | 0.08 | 0.09 | 0.06 | 0.09 | 0.09 | 0.09 | -1.67 | 0.08 | 0.09 | 0.08 | 0.09 | 0.09 |
| (I) department store 2 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | -1.72 | 0.02 | 0.02 | 0.02 | 0.02 |
| (J) discounter 1 | 0.06 | 0.05 | 0.06 | 0.04 | 0.06 | 0.06 | 0.06 | 0.06 | 0.05 | -1.67 | 0.07 | 0.06 | 0.06 |
| (K) discounter 2 | 0.04 | 0.03 | 0.04 | 0.02 | 0.04 | 0.04 | 0.04 | 0.03 | 0.04 | 0.05 | -1.71 | 0.04 | 0.04 |
| (L) general merch. 1 | 0.10 | 0.08 | 0.09 | 0.06 | 0.10 | 0.09 | 0.10 | 0.10 | 0.08 | 0.10 | 0.10 | -1.64 | 0.09 |
| (M) specialty chain 1 | 0.05 | 0.04 | 0.05 | 0.03 | 0.05 | 0.05 | 0.05 | 0.05 | 0.04 | 0.05 | 0.04 | 0.05 | -1.69 |

**Notes:** Average (across consumers and time) price elasticity between pairs of chains. Each cell in the table represents the percent-change in purchase incidence to the *column* chain, in response to a percent-change in the price of the *row* chain.

Appendix Table 10: Average price, marginal costs, and profit margins

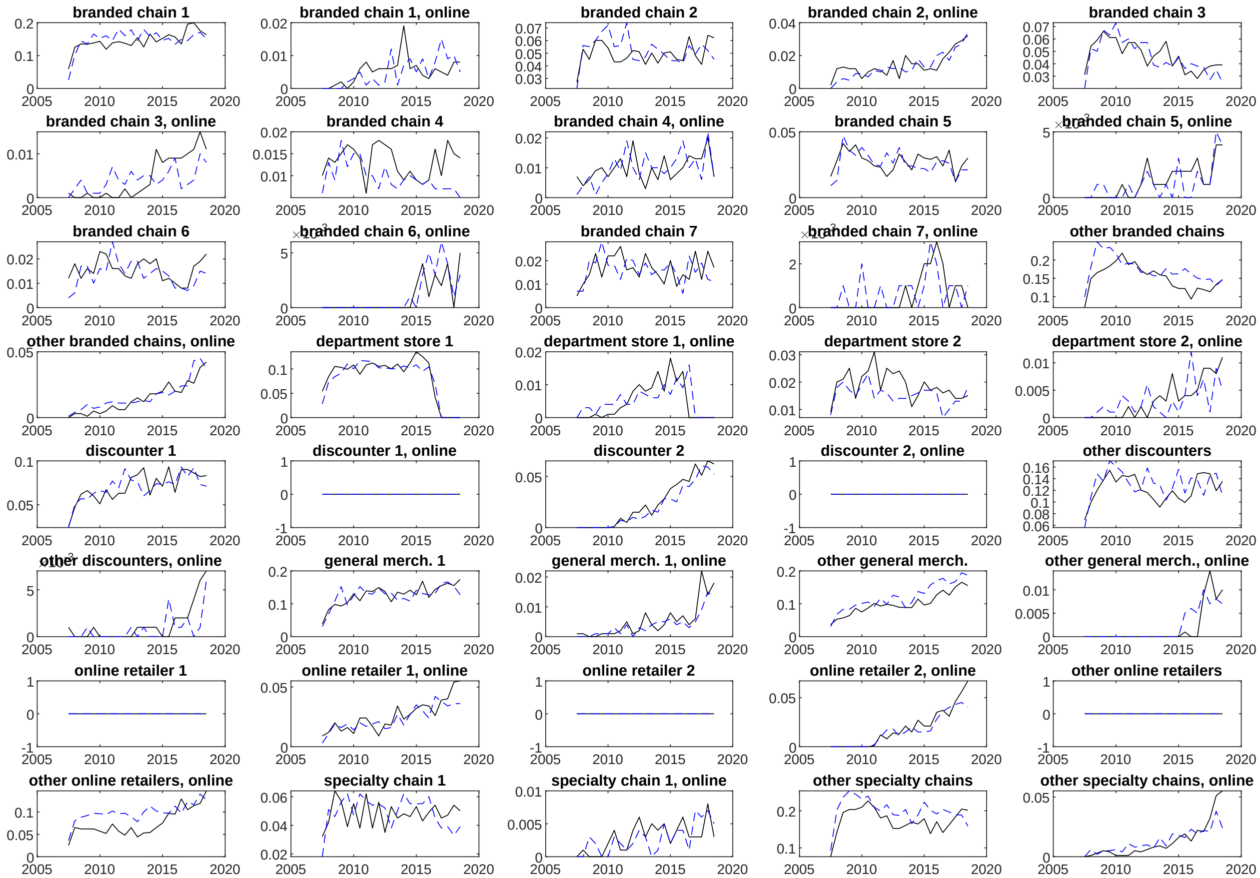
| | average price | average cost | average %gross margin |
|---|---|---|---|
| branded chain 1 | 17.08 | 5.25 | 0.60 |
| branded chain 2 | 12.38 | 4.96 | 0.51 |
| branded chain 3 | 28.35 | 9.26 | 0.58 |
| branded chain 4 | 35.33 | 13.35 | 0.53 |
| branded chain 5 | 21.18 | 7.39 | 0.56 |
| branded chain 6 | 23.42 | 7.85 | 0.57 |
| branded chain 7 | 26.94 | 9.23 | 0.57 |
| department store 1 | 15.12 | 5.18 | 0.57 |
| department store 2 | 46.62 | 15.34 | 0.58 |
| discounter 1 | 2.98 | 1.07 | 0.55 |
| discounter 2 | 6.31 | 2.19 | 0.56 |
| general merch. 1 | 5.01 | 1.75 | 0.56 |
| online retailer 1 | 30.52 | 9.67 | 0.59 |
| online retailer 2 | 46.22 | 14.59 | 0.59 |
| specialty chain 1 | 9.26 | 3.43 | 0.54 |

**Notes:** This table presents estimates for the average prices, estimated marginal costs, and implied percent gross profit margins. These estimates are derived from the 2016-2018 sample.

Appendix Table 11: Changes in equilibrium prices from taking away e-commerce

| | mean (across chains) | (std err) | std | (std err) |
|---|---|---|---|---|
| price change: 2007-2009 | 0.025 | 0.004 | 0.021 | 0.004 |
| ... 2016-2018 | 0.061 | 0.013 | 0.056 | 0.013 |
| ... fully learned | 0.116 | 0.025 | 0.102 | 0.025 |

**Notes:** Counterfactual changes in equilibrium prices when e-commerce is taken away.




Appendix Figure 3: Observed and model-predicted quantity

**Notes**: This figure reports observed (in solid) and simulated (in dash) choice paths for each chain-channel over time.