The Role of Data in the Retail Industry: An Economic Assessment

University of Tokyo Economic Consulting Inc.*

December 8, 2021

Contents

* This report was commissioned by Amazon Japan G.K. We thank Antara Dutta, Thibault Larger, and Ritsuko Okada for their helpful comments.

Executive summary

- Data collection and usage enable businesses to make better decisions. In many cases, this involves using data to generate a predictive model, and employing the model to make business-relevant predictions. For example, data are used in the retail industry to predict consumer demand, which in turn is used to make decisions regarding pricing, product assortment, inventory, and product development. Data are also used for service customization; examples include product recommendations, customized promotions, and comparison shopping.
- While individual retailers may benefit to some extent by accumulating a large amount of data, it is often the case that the relevant data are available from third-party sources. Some types of retail service, including checkout coupons and comparison shopping services, are carried out by specialized firms who obtain data from individual retailers.
- Data usage by firms has the potential to promote competition by reducing the search costs faced by consumers, which in turn leads to lower prices and greater consumer welfare. On the other hand, some policy makers have recently expressed concerned about the competitive effect of data accumulation by digital platforms. It has been pointed out by some scholars and commentators that data accumulation can create entry barriers, partly by enabling or enhancing scale economies and network effects. It has also been pointed out that data accumulation by firms could increase consumers' switching costs. We examine these possibilities in the context of the retail industry.
- The relevant data for starting a retail business are generally available from third-party providers, so that the accumulation of first-party data is not a prerequisite for market entry. It is also unlikely that data accumulation plays a significant role in the realization of scale economies; there are other factors that are far more important, such as capital requirements for stores and distribution centers. Moreover, service differentiation between retailers, as well as the prevalence of multi-homing by consumers, imply that numerous retailers can coexist in the market notwithstanding any network effects that may exist. Taken together, these observations imply that data accumulation by firms is unlikely to give rise to entry barriers in the retail industry.
- Following the classification of switching costs by Klemperer (1995), the only types of switching costs that are encountered in retail settings and that relate even tangentially to data are (i) transaction costs that accrue from switching suppliers, and (ii) the loss of benefits from discount coupons and similar devices (e.g., loyalty programs). The first type of switching cost is likely to

be small in the retail industry, compared to other industries such as banking where providers require information on customers to provide proper service. The second type of switching cost is probably more relevant in the retail sector, but its dependence on data accumulation is likely to be weak. Customized coupons can be profitably deployed by retailers even when only small amounts of information are directly available to them, since the predictive tasks are often carried out by specialized service providers. Where loyalty programs give rise to switching costs, they are primarily due to accumulated points rather than accumulated data.

- Based on the perception that asymmetric access to data has impeded competition, some governments have implemented, or are considering implementing, policies that enhance data portability or mandate data access. The idea is that these measures will enhance competition by reducing switching costs and by facilitating multi-homing. In the retail industry, where consumers and business users face relatively low switching costs (and any switching costs that do exist are caused primarily by non-data factors), and face little barriers to multi-homing, these policies are likely to be limited in their ability to promote competition. Meanwhile, such policies are likely to discourage data collection and utilization by firms, which could harm both consumers and business users.

1 Introduction

The development of information technology has made it possible to collect and analyze large amounts of data for business purposes. Such practices are now common in a multitude of industries including manufacturing, transportation, logistics, telecommunications, construction, finance, retail, and healthcare. While such trends are observed across the board, the economic role of data varies across industries, due to differences in the way they are collected and used.

For this reason, instead of trying to cover a wide range of industries and topics, we focus primarily on the retail industry and examine the role of data in determining the competitiveness of firms. We also address some concerns that have been raised in policy circles regarding the effect that data has on competition. In particular, we examine whether the accumulation of data by firms in the retail industry has the potential to affect market entry and switching costs.

The remainder of the report is structured as follows. In Section 2, we discuss the economic value of data, not just in the retail sector but more generally. In Section 3, we begin our focus on the retail industry, describing the production and use of data in that sector. Section 4 discusses the effect of data accumulation on competition in the retail industry. In particular, we examine the relationship between data and market entry, and that between data and switching costs. Based on our findings, Section 5 evaluates recent policy initiatives that aim to promote competition by mandating data access and requiring data portability. Section 6 concludes.

2 Economic value of data

2.1 Relationship between data, information, knowledge, and action

It is well known that data have the potential to create value for economic decision-makers – including firms, government agencies, and consumers – by improving their decisions. On the other hand, the detailed pathway by which data influences decision-making is not necessarily common knowledge. To fill this gap, we use this section to present a succinct description of this pathway.[1]

First of all, it is important to note that raw unorganized data are not useful until they are used to create "information" by processing and organizing it in some way.[2] For example, a dataset containing individual purchases made by consumers is worthless to manufacturers or retailers until it is converted into information on consumers' behavior. Such a conversion involves finding patterns or rules that exist in the data – e.g., the relationship between product features, price, consumer characteristics, and purchase incidence. Traditionally, analysts have uncovered such relationships by manually constructing a statistical model with a specific functional form and a selection of variables. In recent years, machine learning methods are increasingly being used to find patterns or rules in the data (Mullainathan and Spiess, 2017; Hastie, Tibshirani, and Friedman, 2017).

The next step for decision-makers is to use the information derived from data to create more information or alternatively, to create "knowledge". For example, a retailer who obtains a statistical model of consumer behavior can use it to make predictions about sales volume under different pricing scenarios. This is an instance where information is used to generate additional information. The retailer may also gain insight into whether its price is too high or too low, using the statistical model of consumer behavior along with other information. This is an example of information being used to create knowledge.

Finally, the knowledge gained by decision-makers is used to guide their actions. For example, a retailer who learns that its price is too high would act upon that knowledge by lowering its price. Likewise, a manufacturer launching a new product uses its knowledge of consumer demand to set what it thinks is the "optimal" price for the product. Consumers often use data-derived information and knowledge to guide their actions as well. For example, many consumers use comparison shopping services to find products and sellers. They use various pieces of information, including prices and customer reviews, to decide whether or not to buy a product and whom to buy it from. All of these decisions generate significant economic value for the decision-maker, either in the form

---

[1] The following conceptualization of the use of data by decision-makers is due to Varian (2019).

[2] Unless otherwise noted, in this report the term "data" refers to raw data that has not yet been processed, and "information" refers to intermediate output (including secondary data) that has been created by processing raw data.

of larger profits or increased utility.

At this juncture, it is worthwhile noting the different classes of personal data. The oft-cited World Economic Forum's classification divides personal data into the following, according to the way in which they are collected or generated (World Economic Forum, 2011):

(i) Volunteered data: Data explicitly entered by individuals, e.g., personal profiles.

(ii) Observed data: Data that record the actions of individuals, e.g., purchase histories.

(iii) Inferred data: Data on individuals derived from the analysis of volunteered or observed data, e.g., credit scores.

Note that what the World Economic Forum calls "inferred data" would actually be classified as "information" under Varian's (2019) typology which we adopt in this report. Some types of inferred data, including credit scores, are recorded at the individual level, so that at first glance they may appear indistinguishable from volunteered or observed data. However, since inferred data have already gone through a certain level of processing, which typically results in some value being added, we find it more appropriate to classify them as information. Once we reclassify inferred data as "information" rather than "data", it becomes clear that the economic value of data is realized only after it has been processed into decision-relevant information.

It should also be noted that the economic value of data depends, to a large extent, on the quality of the algorithm used for extracting information. Thus, the same dataset may yield vastly different economic value depending on the quality of the algorithm. This is exemplified in the results of the ImageNet Challenge, which is a well-known visual recognition contest. As documented by Agrawal, Gans, and Goldfarb (2018), as the quality of algorithms improved over time, the error rate of the winning algorithm decreased from 28 percent in 2010 to less than 5 percent in 2017.

2.2 Prediction and its various applications

Data are often used to make predictions, and predictions often have great economic value. The prediction of product sales under alternative pricing scenarios, described in the previous section, is but one example of their use in business. Other examples include search engines, targeted marketing and recommendations, language translation, autonomous vehicles, fraud detection, and various artificial intelligence applications such as virtual assistants and game-playing programs (Agrawal, Gans, and Goldfarb, 2018).

The value of prediction can be easily seen through the example of credit card transactions, where machine learning algorithms are used by credit card companies to predict whether a transaction is legitimate or fraudulent. These algorithms typically use "labeled" data on past transactions (i.e., transactions that have been confirmed to be legitimate or fraudulent) to generate a predictive model, and this model is used to predict whether newly generated transactions are fraudulent. These predictions constitute information which issuing banks (i.e., the clients of credit card companies) use in deciding their action: whether to approve or decline a transaction. Credit card companies have been successful at improving their predictions over time, with the result being a higher probability of fraudulent transactions being declined, and a lower probability of legitimate transactions being declined.[3] This improvement has been realized purely through better use of data, namely the application of better algorithms.

Another application of machine learning in finance is credit scoring, where algorithms are used to estimate the probability that borrowers default on their loans. Such predictions are used by potential lenders to decide whether or not to provide credit. While financial institutions have been the main users of credit scoring models, in recent years other firms have begun using them as well.[4]

Perhaps the best-known applications of machine learning algorithms are search engines such as Google. Data on past queries by users – search terms, search results, and users' clicking behavior – are fed into a machine learning algorithm to generate a predictive model. That model is then applied to a new query to predict the likelihood of the user clicking on each webpage in the search engine's database. These predictions form the basis for the search results presented to the user.[5]

2.3 Data accumulation and prediction accuracy

In general, data accumulation leads to improvement in prediction accuracy (i.e., better information), and this can lead to improved decision-making. However, this improvement in accuracy tends to be a decreasing function of data; as more data are accumulated, it becomes harder to improve prediction accuracy simply through additional data (Varian, 2019). In addition, the value of data tends to deteriorates over time. For example, Chiou and Tucker's (2017) analysis of the search engine market shows that the retention of historical data does not lead to significant competitive advantages for search engine service providers.

That said, a relatively small improvement in the accuracy of predictions can sometimes generate substantial incremental value. As an example, for credit card companies that use prediction to detect fraudulent transactions, a relatively small improvement in prediction accuracy can generate a large benefit. This is because a small reduction in the frequency of false positives (i.e., wrongly labeling a legitimate transaction as fraudulent) can dramatically increase customer satisfaction, while a small reduction in the frequency of false negatives (i.e., failing to detect a fraudulent transaction) can

---

[3] See, for example, the description of Mastercard's fraud detection service: https://newsroom.mastercard.com/press-releases/mastercard-rolls-out-artificial-intelligence-across-its-global-network/

[4] As an example, University of Tokyo Economic Consulting has developed a machine learning-based credit-scoring model that covers a large number of firms in Thailand. The model has been used by Diamond Generating Asia, an electrical power supplier, to assess the creditworthiness of potential clients. See: https://prtimes.jp/main/html/rd/p/000000002.000065767.html

[5] See: https://www.google.com/intl/ja/search/howsearchworks/

significantly increase revenues. Moreover, such improvements make it less necessary to involve human judgment in the decision-making process, which can lead to significant cost reductions (Agrawal, Gans, and Goldfarb, 2018). Therefore, the incremental value that firms gain from accumulating data and improving prediction accuracy varies across industries, and depends on the type of data used as well as the nature of the predictive tasks.

3 Production and use of data in the retail industry

The types of data used, the methods for collecting them, and the way in which they are used by firms vary across industries. This implies that the economic role of data varies across industries as well. For this reason, in the remainder of this report we focus on a specific sector – the retail industry – rather than attempt to cover a wide range of industries.

The retail industry provides an interesting case study for several reasons. To begin with, while there are various benefits to retailers from collecting and using data (as we will discuss shortly), it is not immediately clear that the sophisticated use of data is a prerequisite for success in this industry. This contrasts with other industries, such as those of credit cards and search engines, where the competitiveness of firms is closely linked to how well they collect and use data. Secondly, the retail industry is characterized by the coexistence of offline (brick-and-mortar) and online firms, as well as firms that utilize a mix of offline and online channels, all of which make some use of data. The methods for collecting and using data vary to some extent across channels, but there are important similarities as well. By observing these similarities, and taking into account that most of the methods were initially developed by offline businesses, it should become clear that data collection and usage are deeply engrained in this industry. Finally, there has been some interest in policy circles on how the collection and usage of data by retail businesses, especially online shopping malls, affects their competitiveness as well as the nature of market competition.[6]

To set the stage for a discussion on the competitive role of data in Section 4, in this section we examine how data are actually being collected and used by retailers. We also discuss what competitive advantages, if any, retailers can gain from the accumulation of data.

3.1 Types of data collected

The retail industry connects final consumers with the suppliers of goods and services. This puts retailers in a position where they can observe the purchasing behavior of individual consumers. In particular, a retailer can observe how a customer, faced with a set of alternatives, makes a choice among the available offerings. This is not to say that the retailer has a perfect grasp of consumers'

[6] See, for example, Japan Fair Trade Commission (2019).

purchase behavior. Observations are limited to those consumers who visit the retailer and the products or services that are offered there. If a customer decides not to purchase anything at the retailer, instead making a purchase elsewhere, the eventual purchase is not observed by the retailer.

While retailers vary in terms of their access to and use of data, the types of data collected by retailers include the following:

(i) Customer attributes: Data on individual consumers' age, sex, location, etc., are sometimes provided by consumers as volunteered data. In other cases, the information is obtained as observed or inferred data. For example, major convenience store chains in Japan record the sex and age category of each customer[7] Online retailers as well as brick-and-mortar retailers that run loyalty programs often record customers' purchase histories.

(ii) Product assortment/price/inventory and other conditions faced by customers: This information is recorded for each outlet and each time interval. Conditions faced by customers include special offers, product recommendations, and the layout of stores/websites.

(iii) Customers' browsing and purchasing actions: At the very minimum, retailers obtain product-level sales during each time interval. For online retailers as well as brick-and-mortar stores with loyalty programs, individual purchases are linked to data on customer attributes. In some cases, pre-purchase browsing behavior is also observed.[8]

(iv) Customers' post-purchase behavior: Information such as repeat purchases, product returns, and product reviews are observed at the customer level by online retailers as well as brick-and-mortar retailers that run loyalty programs.

3.2 Uses of data

As mentioned in Section 2, data often generate economic value through their use in prediction. In that section, we sketched how retailers use data to predict the demand for individual products, and use that information to make pricing decisions. Here, we elaborate on this aspect of data usage, and also describe other uses of data by retailers.

[7] This is done by store clerks who enter the sex and age category of each customer at the cash register. See "Why Are Those Buttons Are Gone from the Cash Registers at Convenience Stores?", *Yomiuri Shimbun*, July 18, 2018.

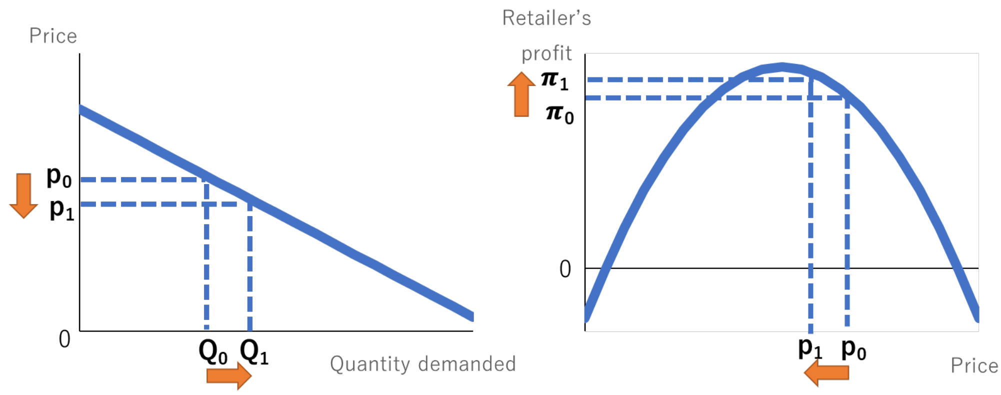
[8] The observation of customer browsing behavior is not limited to online retailers. For example, Aeon Retail, which runs a large chain of supermarkets in Japan, uses in-store cameras connected to artificial intelligence software to observe and analyze customers' behavior. See: https://www.aeonretail.jp/pdf/210513R_1.pdf

3.2.1 Demand forecasting and pricing

By using data on consumers' purchases under various conditions – including product assortment and prices as well as the consumers' own attributes – a retailer can generate a predictive model of consumers' purchase behavior (Grewal et al., 2011). Such a model predicts whether a given consumer, when faced with a given set of circumstances (product assortment, prices, etc.), purchases a particular product. In economics, this model is called a "demand function".

One way of using this model is to forecast the number of purchases under various pricing scenarios. This equates to drawing a demand curve: a graph depicting the relationship between a product's price and its volume of demand, as shown in the left panel of Figure 1. By using the demand curve in combination with other information, the retailer can predict what its profit would be under different pricing scenarios. For instance, the right panel of Figure 1 shows that the retailer's profit is higher under price p_1 than under price p_0. These predictions allow the retailer to set an optimal price, such as one that maximizes profit.

Figure 1: Use of demand curve to guide pricing



The use of demand estimation for pricing has become widespread in recent years, not only by retailers but also manufacturers and service providers. Several research papers in the marketing field document how some retailers employ price optimization software that are based on demand estimation techniques (Grewal et al., 2011; Ferreira, Lee, and Simchi-Levi, 2016; Simchi-Levi, 2017). One example in a non-retailer setting involves the use of demand estimates by sports teams to set ticket prices.[9] Another example is that of manufacturers using estimated demand functions to set the launch price of new products.[10]

[9] For example, the Softbank Hawks uses demand estimates to set seat-specific dynamic prices for baseball games. See: https://www.softbankhawks.co.jp/ticket/ticketprice.html

[10] For example, Buffalo, a large Japanese manufacturer of computer peripherals, collaborated with

3.2.2 Other uses of demand forecasting

Applications of demand function estimation are not limited to pricing. One notable area is the use of estimated demand functions to plan product assortment and inventory (Kök, Fisher, and Vaidyanathan, 2008). For example, the Dutch supermarket chain Albert Heijn reportedly estimates the demand for each product, and places orders on the basis of those estimates to maximize expected revenues. As another example, Mitsubishi Shokuhin, the largest wholesaler of processed foods in Japan, uses data from Lawson, a major convenience store chain, to estimate product-level demand. It then uses those demand estimates to guide the ordering of products from manufacturers and optimize its inventory levels.[11]

Another area where demand estimates are used is the design of new products. For example, Iris Ohyama, a Japanese manufacturer of household appliances, uses data on the sale of its existing products to estimate consumer demand in various product categories, and uses that information to design new products. According to Shimada (2002), Iris Ohyama has access to detailed store-level data because it has a wholesaling arm that deals directly with most of its retailers. This allows it to observe the quantity of sales and inventory at different retail outlets, and relate it to prices and other factors.

Retailers sometimes use higher-level demand estimates in their decision-making, such as those pertaining to entire stores. In particular, retail chains often use predictions regarding the number of customer visits to a particular location when deciding whether or not to open a new outlet. For example, Lawson has applied artificial intelligence techniques to data on existing stores to predict revenues at hypothetical stores in specific locations.[12]

3.2.3 Data-driven service customization

The use of data by retailers is not limited to demand estimation. In particular, consumers' purchase histories and data volunteered by consumers can be used to customize the retail services provided to them. Here, we discuss three important service types: product recommendations, customized promotions, and comparison shopping services.

University of Tokyo Economic Consulting to estimate the potential demand for a network recording and storage device that it launched in April 2021. It used this information to set the product's launch price. See "Sold Out on the Same Day, Eight Times in a Row: Nasne Makes an Unlikely Comeback After 2 Years", *Nikkei X-Trend*, June 15, 2021.

[11] See "Mitsubhishi Shokuhin Uses AI to Make Predictions About Optimal Inventory", *Nikkei Shimbun*, September 1, 2021.

[12] See "Lawson Uses Area Data and AI to Predict Store Profitability" *SankeiBiz*, February 20, 2018.

Product recommendations

Many retailers use data to identify the products that a given customer is likely to purchase. For example, many online retailers use data on customers' purchase histories to generate predictive models that link current purchasing behavior with past purchases. These models are then used to predict a customer's purchase probabilities for a range of products, based on purchasing decisions by other customers in similar circumstances. The products yielding the highest purchase probabilities are then recommended to the customer (Bodapati, 2008). Empirical research shows that such recommendation systems are effective at presenting new alternatives to consumers and increasing retailers' revenues (Senecal and Nantel, 2004; Kawaguchi, Uetake, and Watanabe, 2019).

From an economic viewpoint, recommendation systems enable consumers to learn about the quality of products from other consumers. As discussed by Tucker and Zhang (2011), a system-generated recommendation for a particular product informs current consumers that past consumers perceived the product to be of high quality (and hence purchased it). To the extent that such perceptions are correlated with true quality, current consumers accordingly update their own beliefs about the product's quality.

An interesting question is whether recommendation systems tend to favor "mass-appeal" products chosen by a large number of consumers, or "niche" products that appeal only to a small minority of consumers. While it may seem at first glance that recommendation systems strengthen consumers' bias towards mass-appeal products, that need not be the case. Indeed, some empirical studies, such as Brynjolfsson, Hu, and Simester (2011), find that the products sold through online retail channels have "longer tails", that is, sales are more evenly spread out across different products. They attribute this to the prevalent usage of recommendation systems and other tools that allow consumers to find and evaluate niche products.

Customized promotions

Another common usage of data by retailers is to target product-specific promotions to certain customers on the basis of their purchase histories (Grewal et al., 2011). A well-known example is Catalina Marketing's Checkout Coupon program, which issues a customized coupon for a specific product to each customer at the cash register, taking into account the customer's current purchase.[13] The target product may be the same as that purchased by the customer, as in the case of "loyalty promotions", or it may be a different product – either a competing product or complementary product – depending on the settings chosen by the retailer. The choice of the specific target product is driven by consumer-level shopping data gathered by Catalina Marketing through its network of clients. Retailers can also issue their own customized coupons by investing in specialized software

[13] Catalina Marketing's current roster of clients in Japan include major retailers such as Aeon and Ito-Yokado. See: https://jp.catalina.com/about/

and hardware.[14]

In recent years, retailers have increasingly turned to the issuance of electronic coupons through smartphone apps. For example, the communication app developer LINE offers a popular service called LINE Coupon which covers approximately 50,000 retail outlets in Japan (as of September 2020). Given LINE's estimated domestic user base of 86 million, the service offers a remarkably wide customer reach to participating retailers. Since 2019, LINE and Catalina Marketing have operated a joint coupon-issuing program which combines Catalina's in-depth purchase data with LINE's broad coverage of users.[15]

Research in the fields of economics and marketing suggests that customized promotions can benefit customers as well as increase retailers' revenue and profitability. Microeconomic theory teaches us that prices under customized promotions (which economists call "behavior-based price discrimination") can be lower than those under uniform (i.e., non-discriminatory) pricing. This implies that customized promotions can make consumers better off.[16]

According to empirical research, customized promotions allow retailers to enjoy increased revenues and profits. For example, Rossi, McCulloch, and Allenby (1996) find that customized coupons that utilize even a small amount of purchase history data can increase retail revenue by 2.5 times more than what could be achieved by a blanket coupon policy. Similarly, Zhang and Wedel (2009) find that customized promotions for a given brand of butter, generated on the basis of consumers' purchases during the directly preceding shopping trip, increase retailers' profit by 121 to 311 percent. Both studies show that the amount of customer-level data required to profitably implement customized promotions is not necessarily large.

Comparison shopping services

An important benefit of online retail for consumers is that it provides easy access to information on the availability and prices of products and services, and reduces search costs faced by them. In particular, "comparison shopping services" allow consumers to search for, and compare, different products that fit their needs as well as different sellers offering the same product. Comparison shopping services perform this function by presenting a list of offerings in response to search queries entered by users; for each offering, information on product features, prices, and user reviews are

---

[14] For example, retailers that use NEC's NeoSarf/POS system are able to manage their own coupon programs. See: https://www.nec-solutioninnovators.co.jp/ss/retail/products/neosarf-pos/function/pos

[15] See: https://jp.catalina.com/solution/line-catalina-coupon/

[16] For example, Fudenberg and Tirole (2000) employ a two-period model with two symmetric firms to show that price levels are lower when firms can price discriminate in period 2 on the basis of consumers' purchases in period 1, relative to the situation where price discrimination is not possible. Note, however, that the results depend to some extent on specific assumptions of the model, and can change under a different set of assumptions.

provided. In effect, comparison shopping services act as consumers' assistants, collecting the relevant data and presenting them in a way that facilitates decision-making. In Japan, these services are provided both by specialized firms such as Kakaku.com and online retail stores such as Amazon.co.jp and Rakuten.

Economists have found that comparison shopping services boost consumer welfare. For instance, Baye and Morgan (2001) show that they improve consumer welfare by causing average prices to fall.[17] A large body of empirical research supports this finding, including Brynjolfsson and Smith (2000) who find that internet retailers offer prices for books and CDs that are 9-16 percent lower than those at brick-and-mortar outlets; they point to online comparison services as a likely cause.[18] Ellison and Ellison (2009) find that computer component suppliers listed on a popular price search engine face extremely high price elasticities of demand for low-quality products, implying that suppliers are driven to set prices near marginal cost.[19]

The benefit of comparison shopping services goes beyond reduction in search costs. The user reviews provided by comparison shopping services allow sellers to establish a reputation for high quality; this information is then used by consumers to improve their shopping decisions. Such reputation effects are important in markets for "experience goods", where consumers are not able to ascertain the quality of a product (including the reliability of the seller) until they make an actual purchase. In such markets, sellers with a favorable reputation are able to set higher prices than those without it. Such reputation premia, in turn, incentivize sellers to provide high quality products and services (Klein and Leffler, 1981; Shapiro, 1983). The online feedback mechanisms that are built into comparison shopping services function as a powerful "reputation system". As discussed by Dellarocas (2003), the strength of these reputation systems is one reason for the remarkable success of online platforms in various industries.

The data used by comparison shopping services are collected in several different ways. Specialized comparison shopping service providers generally receive data on offerings, such as that on prices, from the retailers themselves. In addition to data, some of them receive payment from retailers for sending users to their sites, usually on a pay-per-click basis. The reason why retailers willingly offer their data to comparison shopping services is because they would lose sales to rival retailers if they did not. The user review data used by comparison shopping services are volunteered by the users themselves. Users are free to choose where to leave their reviews. For example, a user

[17] On the other hand, Ronayne (2021) shows theoretically that under certain circumstances, the existence of price comparison services can actually cause a reduction in consumer welfare through higher prices.
[18] At the same time, Brynjolfsson and Smith (2000) find that the degree of price dispersion across retailers is large even in the online segment.
[19] On the other hand, these low-quality products appear to serve as "loss leaders" that attract customers who are then offered higher-quality products that are more profitable for the suppliers.

who purchased a product from an online retailer may choose to write a review on a stand-alone comparison shopping service such as Kakaku.com instead of the retailer's site. Thus, the more attractive a comparison shopping service is to consumers and sellers, the better access to data it has.

3.3 Value of data accumulation for retailers

In the retail setting, "more data" could mean multiple different things. For example, it could mean more customer-level observations for a given product, or it could mean more shopping session-level observations for a given customer. Depending on what one means by "more data", the value that retailers gain from accumulating first-party data (e.g., data on transactions carried out by the retailer or on a platform managed by the retailer) differs.

It has been documented that accumulation of data on a greater number of products has limited impact on prediction accuracy at the individual product level, and hence on decision-making (Bajari et al., 2018). On the other hand, accumulation of data on a greater number of consumers for a given product (or a group of substitutable products) may yield better estimates of demand. As discussed in Section 3.2.1, such demand estimates can be used by firms to guide pricing and promotion decisions. They can also be used to optimize product assortment and inventory levels, or to improve the functioning of product recommendation algorithms.

The above discussion pertains to the accumulation of first-party data by retailers. Meanwhile, point-of-sale data are often collected by specialized data vendors as well, and sold to retailers and manufacturers. Major vendors in Japan include Intage, Nikkei, and GfK. This implies that data accumulation at the individual retailer level is not necessary for generating decision-relevant information such as demand function estimates (Tucker, 2019). Moreover, data collected by an individual retailer can be less useful compared to market-level data collected by specialized data vendors, since the former necessarily cover only a small portion of the market and lacks critical variables, such as product assortment, prices, and sales volumes at rival retailers.

To see why this is the case, suppose that two online retailers, labeled 1 and 2, both sell two brands – A and B – of wireless headphones.[20] Suppose also that Retailer 1 implements a 10 percent discount on Brand A, while Retailer 2 implements a 10 percent discount on Brand B. The net effect is that Retailer 1's sales of Brand A increase by 30 percent, while its sales of Brand B decrease by 15 percent. Meanwhile, Retailer 2's sales of Brand A decrease by 15 percent, while those of Brand B increase by 30 percent. Now, suppose Retailer 1 wants to estimate the cross-price elasticity of demand between Brands A and B – i.e., the degree of competition between the two brands.

If Retailer 1 is restricted to using data obtained from its own outlet, it estimates the cross-price

[20] We are restricting our discussion to two brands and two retailers for the sake of simplicity. In reality, many brands of headphones are sold through many retailers, both brick-and-mortar and online.

elasticity to be 1.5, because the 10 percent price reduction for Brand A coincided with a 15 percent volume reduction for Brand B. However, the volume reduction in Brand B sold at Retailer 1 may have been caused more by the Retailer 2's price reduction for Brand B, and less by Retailer 1's price reduction for Brand A[21]. If this were the case, Retailer 1's estimate of 1.5 for the cross-price elasticity between the two brands would be an overstatement. As this example demonstrates, an individual retailer's first-party data are often insufficient for observing the overall response of consumers to price changes, because they cover only part of the market. By comparison, aggregated point-of-sale data offered by specialized data vendors tend to cover the entire market; such data are therefore more suitable when one wants to estimate the cross-price elasticity between two brands.

It should be noted that many customized services in the retail industry do not rely on data accumulated at individual retailers. In particular, the comparison shopping services described in Section 3.2.3 use data that are voluntarily provided by numerous retailers (as in the case of price information) and consumers (as in the case of user reviews). Thus, in many cases, specialized service providers have an advantage over individual retailers in collecting the necessary data.

4 The competitive role of data in the retail industry

Our discussion in the preceding two sections highlights the potential for data to affect the competitiveness of firms, including those in the retail industry. In this section, we take a somewhat broader perspective and examine the effect that data collection and usage by individual firms have on competition between firms. As in the previous section, we maintain our focus on the retail industry.

The competitive role of data in the retail industry is manifold. Some aspects of data usage have the potential to reduce the search costs faced by consumers. In particular, comparison shopping services and recommendation systems allow consumers to more easily search for products that fit their needs. In addition, comparison shopping services allow consumers to compare prices and service qualities across suppliers. Such reductions in search costs generally lead to more intense competition and lower prices (Diamond, 1971; Anderson and Renault, 1999). This implies that the benefits of data usage tend to accrue largely to consumers, rather than be held as profit by retailers.

Meanwhile, some scholars and commentators have warned that the accumulation of data by digital platforms has the potential to raise entry barriers or create lasting asymmetries between firms in terms of competitiveness, and possibly give rise to consumer switching costs (e.g., Furman et al., 2019; Crémer, de Montjoye, and Schweitzer, 2019). In the remainder of this section, we examine these possibilities in the context of the retail industry.

[21] In other words, it is likely that the "intra-brand competition" for Brand A between the two retailers is stronger than the "inter-brand competition" between Brands A and B sold by Retailer 1.

4.1 Relationship between data and market entry

The possibility of market entry is an important determinant of competition. Incumbent firms that are protected by entry barriers, and hence face a low probability of rival entry, may be able to profitably raise prices or lower quality beyond competitive levels. They may also be less driven to innovate. Conversely, firms that face competitive pressure from actual or potential entrants are driven to keep prices low, maintain a high level of quality, and continue to invest in innovation. For this reason, one of the most important questions faced by competition authorities and fact-finders in competition cases is whether or not entry barriers exist in the specific market under consideration. Here, we focus not on a specific market but a specific industry, and ask whether differential access to data across firms has a tendency to affect entry in the retail industry. To do so, we begin by defining the concept of entry barriers. We then consider whether differential access to data fits the definition of an entry barrier in the context of the retail industry.

4.1.1 What constitutes an entry barrier?

Economists have employed various definitions for the concept of entry barriers over the years. One standard definition is a cost that must be incurred by an entrant that an incumbent does not have to (or has not had to) incur (Carlton and Perloff, 2005). A broader definition is “any cost that delays entry and thereby reduces social welfare” (Mcafee, Mialon, and Williams, 2004). Typical examples of entry barriers include the cost of accessing essential natural resources or technologies that are protected by the exclusive rights of incumbent firms, and entry restrictions through government regulation.

The concept of entry barriers is sometimes expanded to include economies of scale, such as those arising from large capital requirements. This is because scale economies affect the number of firms that can profitably operate in the market (Cabral, 2017). It is important to note that while most industries have economies of scale which require firms to be larger than some threshold size, in many practical situations they seem to have no detrimental effect on competition. To illustrate, a supermarket must have a certain floor area if it is to carry enough merchandise to satisfy shoppers. Nevertheless, competition between supermarkets is generally quite intense. Keeping these caveats in mind, in the next subsection we examine how differential access to data affects entry through its effect on scale economies.

In recent years, network effects have also come to be regarded as a source of entry barriers. Network effects are observed in markets where firms offer services that connect users to other users, or where firms facilitate exchange between two or more different classes of users. The former are called “direct network effects”, while the latter are called “indirect network effects”. When either type of network effect is present, the services of a firm with a large user base are more highly valued than those of a smaller firm, all else equal. For this reason, network effects are sometimes referred to as "demand-side economies of scale" (Baker, 2019). As with conventional scale economies, network effects have the potential to affect the number of firms that can profitably operate in the market. However, it is important to note that in markets where firms offer differentiated services, or where users find it easy to "multi-home" between two or more firms, the existence of network effects is less likely to constitute a significant entry barrier (Crémer, de Montjoye, and Schweitzer, 2019).

4.1.2 Does data accumulation give rise to entry barriers in the retail industry?

We now consider whether differential access to data constitutes an entry barrier in the retail industry. For purposes of clarity, the data in question are limited to those collected by retailers about their users (both consumers and business users), including observed data on transactions. In line with the definition of entry barriers, we pose the following questions:

(i) Do data constitute an essential input that a new entrant must acquire, or can they be substituted by other inputs?

(ii) Does profitable new entry require the realization of scale economies or network effects, and does access to data play a role?

With regard to the first question, a new entrant into the retail market is likely to require some amount of information on consumers' tastes as well as the range of products available from suppliers. Without such information, it would be difficult for the retailer to build a product assortment that appeals to consumers. However, such information need not take the form of users' volunteered data or observed data on user transactions. In fact, in many circumstances the relevant data are available from third party vendors. For example, suppliers' product lines and consumers' tastes in various segments of the consumer electronics industry can be directly observed or estimated from point-of-sale data offered by GfK, BCN, and others.

Regarding the second question, while some scale economies may exist in the retail industry so that new entrants are required to fulfill some minimum efficient scale, they are unlikely to be large enough to discourage entry. As evidence of this, the Japanese online retail sector continues to experience significant new entry. One example is the online flea market operator Mercari, which announced in October 2021 that it will launch a new online marketplace.[22] Data may play some role in the realization of scale economies in the retail industry. For example, Baker (2019) notes how "customer data may allow sellers to reduce their quality-adjusted promotional costs, thereby achieving scale economies" (p.129). However, the role of data as an engine of scale economies is likely to be small in comparison to other factors, such as bargaining power over suppliers and capital requirements for physical infrastructure.

The prevalence of service differentiation among retailers imply that the existence of network

[22] See "Mercari Enters Online Shopping Mall Market", *Nikkei Shimbun*, October 7, 2021.

effects alone need not substantially discourage entry in the retail industry, and the continued emergence of new players attests to this. In the Japanese online retail sector, highly differentiated firms that specialize in specific segments, such as the fashion retailer ZOZO, have been able to capture significant market share. One driving force is sellers' demand for retail services that are tailored to specific product segments. For example, Cake.jp, a retail platform that specializes in sweets, offers same-day product returns to consumers who are unsatisfied with minor blemishes. This has apparently expanded the demand faced by its sellers. TANP, another retail platform that specializes in gift products, generates extra sales by combining the products of different sellers and offering them as bundled packages.[23]

Another factor that weakens the ability of scale or network effects to deter new entry is the prevalence of multi-homing by consumers and sellers, because markets have less of a tendency to "tip" when consumers are able to transact with a large number of sellers on multiple platforms. A survey of online commerce users in Japan carried out by the marketing research firm Do House in December 2020 found that 74.9 percent of respondents use multiple online retailers on a regular basis.[24] Asked about their reasons for doing so, many respondents replied that different retailers have different strengths – e.g., some retailers offer free shipping while others offer generous loyalty points. This suggests that service differentiation and multi-homing in online retail are mutually reinforcing.

4.2 Relationship between data and switching costs

Consumers who have previously purchased a good or service from one firm may face a cost when switching to another firm. These costs are called switching costs. Economists have understood that switching costs give firms a degree of market power over existing customers, and has the potential to affect pricing and service quality. At the same time, it has been recognized that switching costs come in different forms, and that the different forms affect competition differently. We therefore begin by reviewing the typology of switching costs. Klemperer (1995) classifies the switching costs faced by consumers into the following six categories:

(i) Need for compatibility with existing equipment
(ii) Transaction costs that accrue from switching suppliers
(iii) Costs of learning to use new brands
(iv) Uncertainty about the quality of untested brands
(v) Loss of benefits from discount coupons and similar devices

---

[23] See "Internet Retailers Find an Opportunity in Specialization", *Nikkei Shimbun*, September 14, 2021.
[24] See "30 Percent of Consumers Consciously Use Multiple E-commerce Sites", *SankeiBiz*, January 27, 2021.

(vi) Psychological costs of switching, or non-economic "brand loyalty"

Types (i) and (iii) are unlikely to be encountered in retail markets, given that most consumers can use new retail services without specialized equipment or training. While types (iv) and (vi) may exist to some extent in retail markets, they are basically unrelated to data. We therefore focus our attention on the remaining two types: (ii) transaction costs that accrue from switching suppliers, and (v) loss of benefits from discount coupons and similar devices. In particular, we examine whether data accumulation by retailers has the potential to increase such switching costs.

As a general matter, the transaction costs borne by a consumer who switches from one retailer to another is likely to be small relative to, say, a consumer who switches between different banks or different hairstylists. This is because unlike a lending bank that needs to know the creditworthiness of its customers, or a hairstylist who must know his clients' tastes, a retailer can provide sufficient service without having detailed knowledge of the customer. Thus, consumers can transact with a new retailer without having to provide much information about themselves, except for their shipping address and payment details such as credit card information.

That said, there are some types of retail service, most notably product recommendations, that require some knowledge of customers' tastes or demographics. Information on customers' tastes is typically gleaned from their purchase histories. It is theoretically possible, therefore, that the accumulation and use of data by retailers generate some transaction-costs-type switching costs. It is important to recall, however, that a majority of consumers multi-home among different retailers. This implies that consumers' purchase histories are often maintained by multiple retailers, making the existence of switching costs less relevant for competition. Perhaps more importantly, consumers who multi-home can take the recommendation provided by one retailer, and use it to make a purchase from another retailer. In other words, retailers can free-ride on rivals' recommendation systems. As discussed by Akman and Sokol (2017), such free-riding between online retailers is quite prevalent. This implies that in practice, the ability of recommendation services to generate switching costs for consumers is likely to be limited.

Discount coupons and similar devices, such as loyalty programs that accumulate points, can generate switching cost if consumers expect the benefits accruing from them to disappear once they switch from one retailer to another. While customized coupons are sometimes issued on the basis of purchase histories, they can be profitably deployed by retailers even when only small amounts of information are directly available to them. This is because the predictive tasks are often carried out by specialized service providers (recall our discussion in Section 3.2.3). In the case of loyalty programs, it is not the collected data *per se*, but rather the accumulated points, that is the primary source of switching costs for consumers.[25] Therefore, it appears that switching costs arising from discount

[25] See, for example, Hartmann and Viard (2008), who show that customers in reward programs face substantial switching costs only when they are close to receiving a reward. This suggests that

coupons and similar devices are not particularly dependent on data accumulation.

5 Implications for data portability and mandated data access

Based on the perception that asymmetric access to data has impeded competition in some industries, mandated data sharing between platforms has been considered as a policy option in Europe and elsewhere (e.g., Crémer, de Montjoye, and Schweitzer, 2019). Article 6 of the Digital Markets Act, which is currently under consideration in the European Union, contains a provision that obligates “gatekeeper” platforms to provide, to business users and third parties authorized by them, real-time access to data generated in the context of those business users’ activities on the platform.

Meanwhile, certain governments have begun to recognize the rights of individual consumers to access and utilize personal data that relate to them and that are held by digital platforms. For example, Article 20 of the European Union’s General Data Protection Regulation states that individuals have the right to receive their personal data from a digital platform “in a structured, commonly used and machine-readable format”, as well as the right to transmit such data to another digital platform.

While these initiatives have the potential to promote competition between digital platforms in some industries, they may not be as effective in others. In this section, we draw from our discussions in the preceding sections to consider the effects of implementing such initiatives in the retail sector. Given that our analysis of the competitive effects of data in Section 4 has been limited to those that work through market entry and switching costs, our discussions here are restricted to those channels as well.

The economic objective of data portability requirements is generally understood to be the promotion of competition in markets where data accumulation by incumbent firms allow them to have durable market power (OECD, 2021). By allowing consumers and other individuals who use a firm’s services to transfer their personal data to another firm, data portability requirements are envisioned to reduce the switching costs faced by those individuals and allow them to more easily engage in multi-homing. The presumption is that these changes would facilitate entry into the relevant markets. The underlying premise for this chain of reasoning is that customers face high switching costs and do not multi-home.

The objectives of mandated data access, such as that described in Article 6 of the European Union’s Digital Markets Act, are similar. It is envisioned that, by requiring digital platform operators to provide data access to business users and third-party firms designated by them (including rivals of the digital platform operators), the switching costs faced by business users are reduced and their ability to multi-home are enhanced. These changes, in turn, are expected to improve the entry

switching costs originate primarily from accumulated points.

prospects of rival platforms. Again, the premise is that, as a result of data accumulation by incumbent firms, business users face significant switching costs and have insufficient ability to multi-home.

Our analysis in the preceding sections suggest that the underlying assumptions of these policy interventions do not necessarily hold in the retail industry. As discussed in Section 4.2, consumers' switching costs in the retail industry are likely to be smaller than in other industries, such as banking, where service providers require users' data to provide adequate service. Moreover, both consumers and business users have a strong tendency to multi-home in the retail industry.

We have already mentioned the tendency of consumers in Japan to use multiple online retailers (see Section 4.1.2). The benefits of using multiple retail channels are widely recognized by sellers as well. For example, a 2019 survey of Amazon sellers by Feedvisor, a firm that provides data analytics and pricing services, found that only 13 percent of the surveyed sellers used Amazon as the sole retail outlet.[26] This proportion appears to be decreasing over time, which suggests that multi-homing by sellers is becoming more prevalent.[27]

To give a more specific example, Working Unit Japan, which acts as a distributor of European luggage brands, uses the three major retail stores in Japan – Amazon.co.jp, Rakuten, and Yahoo Shopping – as well as its own e-commerce site hosted on Shopify. Part of the allure of Shopify for Working Unit Japan is that it allows the distributor to manage inventory and sales not only on the Shopify-hosted site but also on Rakuten.[28] As this example suggests, retail stores compete for sellers not only with each other, but also with emerging e-commerce service providers such as Shopify, BASE, and STORES.

Given the prevalence of multi-homing as well as the limited nature of switching costs in the retail industry, the competition enhancement effects arising from mandated data access and/or data portability requirements are likely to be limited. On the other hand, it is quite likely that such interventions would reduce retailers' incentive to invest in the collection and utilization of data. Indeed, after the General Data Protection Regulation came into force in 2018, technology venture investment in the EU (not limited to the retail sector) experienced a decline relative to similar investments in the U.S. and other parts of the world (Jia, Jin, and Wagman, 2021). Given the importance of data as an input for technology start-ups, this is highly suggestive of the disincentivizing effect of the GDPR. To the extent that a reduction in investment causes a decline in

[26] See Feedvisor, *The State of the Amazon Marketplace 2019* (https://feedvisor.com/resources/amazon-trends/the-state-of-the-amazon-marketplace-2019/).

[27] See Feedvisor, *The State of the Amazon Marketplace 2018* (https://feedvisor.com/resources/industry-news/the-state-of-the-amazon-marketplace-2018-findings-released/), which reports that the proportion of sellers that used Amazon as the sole retail outlet was 20 percent in 2018.

[28] See: "Which One to Use? Own E-commerce Site or Online Shopping Mall: Testimony from 2 Companies that Use Both", *Nikkei X-Trend*, September 15, 2021.

the quality of data-based services, both consumers and business users may be negatively affected.

6 Conclusion

The collection and use of data by businesses has created value for firms and consumers by improving the predictions used by them, and by extension, the quality of their decisions. In the retail industry, this has taken the form of improved demand forecasting which has aided firms in their various activities. Data collection and usage have also reduced the search costs faced by consumers, by enabling new types of retail service such as product recommendation and comparison shopping. This has had the effect of promoting competition among retailers and lowering prices in both online and offline channels, implying that the benefits of data usage have largely accrued to consumers.

While some concerns have been raised in policy circles about the potential for digital companies' accumulation of data to raise entry barriers and switching costs, the characteristics of the retail industry render such concerns far less serious than in some other industries. In particular, the prevalence of service differentiation among retailers, as well as consumers' and suppliers' tendency to multi-home, weakens the ability of network effects to deter new entry. Thus, the Japanese online retail industry has continued to see the emergence of new players, including segment-specialized retailers such as ZOZO and e-commerce service providers such as Shopify and BASE. In addition, switching costs play a relatively small role in the retail industry, given that retailers do not require detailed information on their customers to provide proper service. To the extent that switching costs do exist, they are primarily caused by non-data factors such as accumulated points within loyalty programs. Taken together, these observations imply that the data-focused policies currently under consideration, namely mandated data access and data portability requirements, are unlikely to have a significant competition enhancement effect in the retail industry. Given their potentially negative effect on firms' investment incentives, the need for such policies should be carefully assessed on an industry-by-industry basis.

References

AGRAWAL, A., J. S. GANS, AND A. GOLDFARB. (2018): *Prediction Machines: The Simple Economics of Artificial Intelligence*, Harvard Business Press.

AKMAN, P., AND D. SOKOL. (2017): "Online RPM and MFN Under Antitrust Law and Economics," *Review of Industrial Organization*, 50, 133–51.

ANDERSON, S. P., AND R. RENAULT. (1999): "Pricing, Product Diversity, and Search Costs: A Bertrand-Chamberlin-Diamond Model," *The RAND Journal of Economics*, 30, 719–35.

BAJARI, P., V. CHERNOZHUKOV, A. HORTACSU, AND J. SUZUKI. (2018): "The Impact of Big Data on Firm Performance: An Empirical Investigation," NBER Working Paper 24334.

BAKER, J. (2019): *The Antitrust Paradigm*, Harvard University Press.

BAYE, M. R., AND J. MORGAN. (2001): "Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets," *American Economic Review*, 91, 454–74.

BODAPATI, A. V. (2008): "Recommendation Systems with Purchase Data," *Journal of Marketing Research*, 45, 77–93.

BRYNJOLFSSON, E., Y. HU, AND D. SIMESTER. (2011): "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales," *Management Science*, 57, 1373–86.

BRYNJOLFSSON, E., AND M. D. SMITH. (2000): "Frictionless Commerce? A Comparison of Internet and Conventional Retailers," *Management Science*, 46, 563–85.

CABRAL, L. (2017): *Introduction to Industrial Organization*, 2nd ed., MIT Press.

CARLTON, D., AND J. PERLOFF. (2005): *Modern Industrial Organization*, 4th ed., Pearson.

CHIOU, L., AND C. TUCKER. (2017): "Search Engines and Data Retention: Implications for Privacy and Antitrust," NBER Working Paper 23815.

CRÉMER, J., Y.-A. DE MONTJOYE, AND H. SCHWEITZER. (2019): *Competition Policy for the Digital Era.*, LU: Publications Office.

DELLAROCAS, C. (2003): "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science*, 49, 1407–24.

DIAMOND, P. A. (1971): "A model of price adjustment," *Journal of Economic Theory*, 3, 156–68.

ELLISON, G., AND S. F. ELLISON. (2009): "Search, Obfuscation, and Price Elasticities on the Internet," *Econometrica*, 77, 427–52.

FERREIRA, K. J., B. H. A. LEE, AND D. SIMCHI-LEVI. (2016): "Analytics for an Online Retailer: Demand Forecasting and Price Optimization," *Manufacturing & Service Operations Management*, 18, 69–88.

FUDENBERG, D., AND J. TIROLE. (2000): "Customer Poaching and Brand Switching," *The RAND Journal of Economics*, 31, 634-657.

FURMAN, J., D. COYLE, A. FLETCHER, D. MCAULEY, AND P. MARSDEN. (2019): *Unlocking Digital Competition: Report of the Digital Competition Expert Panel*, Digital Competition Expert Panel.

GREWAL, D., K. L. AILAWADI, D. GAURI, K. HALL, P. KOPALLE, AND J. R. ROBERTSON. (2011): "Innovations in Retail Pricing and Promotions," *Journal of Retailing*, 87, S43–52.

HARTMANN, W. R., AND V. B. VIARD. (2008): "Do frequency reward programs create switching costs? A dynamic structural analysis of demand in a reward program," *Quantitative Marketing and Economics*, 6, 109–37.

HASTIE, T., R. TIBSHIRANI, AND J. FRIEDMAN. (2017): *The Elements of Statistical Learning: Data Mining, Infernce, and Prediction*, 2nd ed., Springer.

JAPAN FAIR TRADE COMMISSION. (2019): *Report on the Fact-Finding Survey on the Trading Practices of Digital Platforms: Business-to-Business Transactions in the Online Shopping Mall and App Store Sectors*, Japan Fair Trade Commission.

JIA, J., G. Z. JIN, AND L. WAGMAN. (2021): "The Short-Run Effects of the General Data Protection Regulation on Technology Venture Investment," *Marketing Science*, 40, 661–84.

KAWAGUCHI, K., K. UETAKE, AND Y. WATANABE. (2019): "Effectiveness of Product Recommendations Under Time and Crowd Pressures," *Marketing Science*, 38, 253–73.

KLEIN, B., AND K. B. LEFFLER. (1981): "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy*, 89, 615–41.

KLEMPERER, P. (1995): "Competition when Consumers have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade," *Review* *of Economic Studies*, 62, 515–39.

KÖK, A. G., M. L. FISHER, AND R. VAIDYANATHAN. (2008): "Assortment Planning: Review of Literature and Industry Practice," in *Retail Supply Chain Management*, ed. by Agrawal, N. and S. A. Smith, Boston, MA: Springer US, 99–153.

MCAFEE, R. P., H. M. MIALON, AND M. A. WILLIAMS. (2004): "What Is a Barrier to Entry?," *American Economic Review Papers and Proceedings*, 94, 461–65.

MULLAINATHAN, S., AND J. SPIESS. (2017): "Machine Learning: An Applied Econometric Approach," *Journal of Economic Perspectives*, 31, 87–106.

OECD. (2021): *Data Portability, Interoperability and Digital Platform Competition*, OECD Competition Committee Discussion Paper.

RONAYNE, D. (2021): "Price Comparison Websites," *International Economic Review*, 62, 1081–1110.

ROSSI, P. E., R. E. MCCULLOCH, AND G. M. ALLENBY. (1996): "The Value of Purchase History Data in Target Marketing," *Marketing Science*, 15, 321–40.

SENECAL, S., AND J. NANTEL. (2004): "The influence of online product recommendations on consumers' online choices," *Journal of Retailing*, 80, 159–69.

SHAPIRO, C. (1983): "Premiums for High Quality Products as Returns to Reputations," *Quarterly Journal of Economics*, 98, 659-680.

SHIMADA Y. (2002): "New product development using demand prediction system," *Nikkei Computer*, July 15, 60–65.

SIMCHI-LEVI, D. (2017): "The New Frontier of Price Optimization," *MIT Sloan Management Review*, Fall 2017 Issue.

TUCKER, C. (2019): "Digital Data, Platforms and the Usual [Antitrust] Suspects: Network Effects, Switching Costs, Essential Facility," *Review of Industrial Organization*, 54, 683–94.

TUCKER, C., AND J. ZHANG. (2011): "How Does Popularity Information Affect Choices? A Field Experiment," *Management Science*, 57, 828–42.

VARIAN, H. (2019): "Artificial Intelligence, Economics, and Industrial Organization," in *The Economics of Artificial Intelligence: An Agenda*, ed. by Agrawal, A., J. Gans, and A. Goldfarb. University of Chicago Press, 399–422.

WORLD ECONOMIC FORUM. (2011): *Personal Data: The Emergence of a New Asset Class*, World Economic Forum.

ZHANG, J., AND M. WEDEL. (2009): "The Effectiveness of Customized Promotions in Online and Offline Stores," *Journal of Marketing Research*, 46, 190–206.