RAND Journal of Economics
Vol. 48, No. 3, Fall 2017
pp. 611–646

Advertising, consumer awareness, and choice: evidence from the U.S. banking industry

Elisabeth Honka*

Ali Hortaçsu**

and

Maria Ana Vitorino***

How does advertising influence consumer decisions and market outcomes? We utilize detailed data on consumer shopping behavior and choices over bank accounts to investigate the effects of advertising on the different stages of the shopping process: awareness, consideration, and choice. We formulate a structural model with costly search and endogenous consideration sets, and show that advertising in the U.S. banking industry is primarily a shifter of awareness as opposed to consideration or choice. Advertising makes consumers aware of more options, search more, and find better alternatives. This increases the market share of smaller banks and makes the industry more competitive.

* University of California Los Angeles; elisabeth.honka@anderson.ucla.edu.
** University of Chicago and NBER; hortacsu@uchicago.edu.
*** University of Minnesota; vitorino@umn.edu.
We are grateful to Frederico Belo, Mark Bergen, Thomas Holmes, Mitch Lovett, John Lynch, Amil Petrin, Enno Siemsen, Aaron Sojourner, and Joel Waldfogel for their numerous comments and suggestions. We also thank the participants of the 2013 Marketing Science conference (Istanbul), the 2014 UT Dallas FORMS conference, the 2014 IIOC conference (Chicago), the 2014 Yale Customer Insights conference (New Haven), the 2014 CIRPÉE Conference on Industrial Organization (Montreal, Canada), the 2014 Summer NBER Industrial Organization Meeting (Cambridge, MA), the 2014 Summer Institute in Competitive Strategy (Berkeley), the 2014 Marketing Dynamics conference (Las Vegas), the 2014 Annual Federal Trade Commission Microeconomics Conference (Washington DC), the 2016 Workshop on the Economics of Advertising and Marketing (Vilnius, Lithuania), and seminar participants at the Carlson School of Management (University of Minnesota), Brown University, Anderson School of Management (UCLA), Booth School of Business (University of Chicago), Simon Business School (University of Rochester), Johnson School of Management (Cornell University), Kellogg School of Management (Northwestern University), Universitat Autònoma de Barcelona, and Columbia University for their comments. We specifically thank the discussants Tat Chan, Judith Chevalier, Gautam Gowrisankaran, Sanjog Misra, Sridhar Narayanan, Marc Rysman, Anna Tuchman, and Lawrence White for the detailed comments on our work. We are grateful to RateWatch and to an anonymous market research company for providing us with the data. Mark Egan provided excellent research assistance. This article was generously supported by the NSF (grant no. SES-1426823) and by the Dean's Small Grants Program at the University of Minnesota Carlson School of Management. All errors are our own.

1. Introduction

■ How does advertising work? This is a central question in the economics literature (Bagwell, 2007). In his classic book, Chamberlin (1933) argued that advertising affects demand because (i) it conveys information to consumers with regard to the existence of sellers and the prices and qualities of products in the marketplace, and (ii) it alters consumers' wants or tastes. Measuring the separate contributions of these roles is, however, challenging. Incorporating the different channels through which advertising affects demand into a model, especially one that is suitable for estimation, is a complex task. More importantly, it is rare to have access to data that describes the entire consumer purchase process, from awareness to consideration and finally choice, which is necessary to identify the separate roles of advertising.

This article uses detailed survey data to empirically disentangle the roles of advertising in the different stages of the consumer purchase process and to measure the mechanisms through which advertising ultimately affects consumer choices. More specifically, we measure how much advertising influences consumer behavior directly as a utility shifter versus as a way of increasing consumer awareness for a product or brand. We do so in the context of the banking industry and conduct our measurement through a fully specified structural model that contains the awareness-consideration-choice stages and, in particular, endogenizes the "choice" of consideration set by each consumer using a costly search framework.

Understanding the drivers of demand for retail banking products and services, a very large and growing sector of the economy, is an important question. With its $14 trillion of assets, 7,000 banks, and more than 80,000 bank branches, the U.S. banking industry comprises a very important portion of the "retail" economy with significant attention from regulators and policy makers. Despite the importance of the banking sector, structural demand analyses to date (e.g., Dick, 2008; Mólnar, Violi, and Zhou, 2013; Wang, 2016) have been based on aggregated market share data on deposits. There has been very little research using detailed consumer-level data to characterize consumers' heterogeneous response to drivers of demand. Moreover, although the banking and financial industry spends more than $8 billion per year on advertising,[1] and the industry's market value relies heavily on advertising (as shown, e.g., in Belo, Lin and Vitorino 2014 and Vitorino 2014), there is little academic research that investigates the precise way in which advertising affects consumer demand in this industry. Some recent exceptions in the literature are Gurun, Matvos, and Seru (2016) on the marketing of mortgages and Hastings, Hortaçsu, and Syverson (forthcoming) on retirement savings products. However, neither of these studies can differentiate between the awareness and the utility-shifting functions of advertising.

The primary data source for our study consists of individual-level survey data from a major market research company. The data contain information about which banks consumers are aware of, the set of banks consumers consider, and the identity of the banks consumers decide to open one or more new bank accounts with. In addition, we observe a nearly complete customer profile containing information on demographics such as age, gender, marital status, and education. We complement these data with three additional sets of data on the retail banking industry. Data provided by RateWatch contain information on interest rates and fees for the most common account types for all banks over the same time period as the survey data. Advertising data come from Kantar Media's *Ad$pender* database. Kantar tracks the number of advertisements and advertising expenditures in national media as well as in local media at the Designated Market Area (DMA) level. Last, we collect information on the location of bank branches from the Federal Deposit Insurance Corporation (FDIC). Together, these data give us a detailed picture of consumers' shopping and purchase processes and of the main variables affecting them.

The data show that consumers are, on average, aware of 6.8 banks and consider 2.5 banks when shopping for a new bank account. There is substantial heterogeneity in consumers' awareness and consideration sets as reflected in the large variation in the sizes of these sets across

[1] kantarmediana.com/intelligence/press/us-advertising-expenditures-increased-second-quarter-2013.

consumers. The relationship between the size of consumers' awareness and consideration sets is weak, indicating a distinct difference between the two stages. This difference is further reflected in the large variation across consumers in what concerns which banks enter consumers' awareness and consideration sets. The data also show that there are large differences in the conversion rates from awareness to consideration and from consideration to choice/purchase across banks. In terms of final choices, the most common account types consumers open are checking accounts (85% of consumers), savings accounts (58%), and credit cards (26%). Finally, our data also confirm the crucial importance of local bank presence—that is, bank branches—in the consumer's decision process: given that the consumer decides to consider or to purchase from a bank, we find that the probability that a bank has a local branch within five miles of the consumer's home is on average 81% and 84%, respectively.

To quantify the effects of advertising (and other variables) on economic outcomes, we develop a structural model of the three stages of the consumer purchase process: awareness, consideration, and choice. We model the decision process of a consumer who is considering to add (or not to add) one or more bank accounts to his existing portfolio. The model incorporates informational heterogeneity across consumers (reflected in the size and content of the consumers' awareness and consideration sets) and allows for costly search (about interest rates). We show that incorporating both informational heterogeneity and costly search is important in our context. In addition, we address the potential endogeneity of the banks' advertising intensity variable using the control-function approach.

Awareness is a function of bank advertising, local bank presence, and demographic factors. A consumer searches among the banks he is aware of. Searching for information is costly for the consumer, because it takes time and effort to contact financial institutions and is not viewed as pleasant by most consumers. Thus, a consumer investigates only a few banks that together represent his consideration set and makes the final decision to open one or more new accounts with a bank from among the ones in the considered set. Our utility-maximizing modeling approach connects and models all three outcome variables: the set of banks the consumer is aware of, the consumer's decision of which banks to include in his consideration set given his awareness set, and the decision of which bank to open one or more accounts with, given his consideration set.

The estimates from the awareness stage highlight the importance of both advertising and branch presence in driving how aware consumers are of a bank. The results from the consideration and choice stages indicate that the average consumer search cost that rationalizes the amount of search conducted by consumers within their awareness sets is about four basis points (0.04%) per bank searched, which is in line with consumer search costs estimated in other financial products settings, for example, Hortaçsu and Syverson (2004). The results also show that convenience is a major driver in the consumer shopping process. Convenience is captured by the fact that consumers are more likely to open bank accounts with banks located in close proximity to where the consumers live. The positive effect of local bank presence at the choice stage suggests that, in spite of the widespread availability and convenience of online banking, consumers still value having the possibility of talking to a bank employee in person.

To evaluate the economic implications of our findings, we perform several analyses. First, to answer the question of how advertising works, we compute elasticities of awareness and choice probabilities with respect to changes in advertising. In addition, we perform a counterfactual analysis in which we shut down advertising and investigate its effects on selected economic variables. Overall, our results show that advertising has a large indirect effect on consumer choices via awareness and that it affects consumers' choices directly only marginally. Advertising makes consumers aware of more alternatives; thus, consumers search more and find better alternatives than they would otherwise. In turn, this increases the market share of smaller banks, making the U.S. banking industry more competitive. This suggests that, in the retail banking industry, the primary role of advertising is to inform consumers about the existence and availability of retail banks and their offerings. This finding stands in contrast to other recent research that has also investigated consumers' demand for financial products. For example, Gurun, Matvos, and

Seru (2016) and Hastings, Hortaçsu, and Syverson (forthcoming) suggest a negative (persuasive) effect of advertising for mortgages and retirement savings products, respectively.

Second, we evaluate the extent to which banks can substitute local branch presence with advertising. This analysis is motivated by policy considerations. Banking consolidation is an ongoing feature of the U.S. banking industry, but merger guidelines mostly focus on market concentration measures (calculated as a function of branch presence). Banks, however, can use advertising strategically to compensate for branch closures imposed by regulators. We show that the increase in advertising expenditures that is required for a bank to compensate for a potential branch closure is not prohibitively large, and hence it is a feasible response for banks. For example, a bank with only one branch within five miles of consumers can substitute that branch for about \$600,000 in annual advertising spending. This value is significantly smaller for banks with more than one branch within five miles of consumers.

Third, we evaluate the impact of alternative assumptions about choice sets on demand estimation results. In most studies, consumers' consideration sets are not known to the econometrician and several different assumptions have been proposed. We show that the definition of consumers' choice sets strongly influences empirical results. For example, when we estimate a version of the model assuming full information (i.e., in which consumers consider all banks), we get estimates that are quantitatively and, often, qualitatively different from the ones we get from the structural model in which choice sets are endogenously determined. Indeed, to get meaningful estimates of the effects of key variables such as interest rates and advertising on choice, the researcher has to carefully define choice sets to be as close to the true consideration sets as possible. In most applications, this is of course impossible to test. Unfortunately, we show that seemingly reasonable assumptions about choice sets can lead to large biases in parameter estimates and to wrong inferences about the economic impact of variables.

The remainder of the article is organized as follows. In the next section, we discuss the related literature. In Section 3, we describe our data and in Section 4 we show evidence of consumers' limited information. Then we introduce our model and discuss identification in the following two sections. We present our estimation approach in Section 7 and show our results in Section 8. In Section 9, we discuss several policy implications and conduct counterfactual analyses. Next, we present robustness checks. Finally, we conclude by summarizing our findings in the last section.

2. Related literature

■ This article is related to four streams of literature, namely, literature on the roles of advertising, on multistage models of consumer demand, on consumer search, and on consumer purchase behavior for financial services.

Since Chamberlin's (1933) seminal work in which he describes the informative and persuasive effects of advertising, several empirical researchers have tried to distinguish between these two effects of advertising in a variety of industries. For example, Ackerberg (2001) and Ackerberg (2003) investigate the roles of advertising in the yogurt market. Narayanan, Manchanda, and Chintagunta (2005), Chan, Narasimhan, and Xie (2013), and Ching and Ishihara (2012) study the pharmaceutical market and Lovett and Staelin (2016) investigate entertainment (TV) choices. Consistent with this literature, advertising in our data can be interpreted as having an informative role if it primarily affects awareness, or as having a persuasive effect if it primarily affects choice conditional on awareness. Clark, Doraszelski, and Draganska (2009) use data on over 300 brands and find advertising to have a positive effect on awareness but no significant effect on perceived quality. Our focus is on financial products and, more specifically, retail banking. There is little academic research that investigates the precise way through which advertising affects consumer demand for financial products. Gurun, Matvos, and Seru (2016) and Hastings, Hortaçsu, and Syverson (forthcoming) explore the effects of advertising in the mortgage and social security markets, but neither of these studies can differentiate between the awareness and the utility-shifting functions of advertising. Because we observe consumers' awareness, consideration, and choice of individual banks, we can distinguish between advertising affecting consumers' awareness and advertising shifting consumers' utility.

Although it is well known that consumers go through several stages (awareness, consideration, and choice) in their shopping process before making a purchase decision (as discussed, e.g., in Winer and Dhar, 2011), most demand side models maintain the full information assumption that consumers are aware of and consider all available alternatives. This assumption is mostly driven by data restrictions, as information other than the purchase decision is rarely available to researchers. Among the set of articles that explicitly acknowledge and model the different stages of the consumer shopping process, a crucial distinction relates to the data and identification strategy used. A first group of articles models at least two stages, usually consideration and choice, and uses purchase data for estimation purposes (e.g., Allenby and Ginter, 1995; Siddarth, Bucklin, and Morrison, 1995; Chiang, Chib, and Narasimhan, 1998; Zhang, 2006; Goeree, 2008; Van Nierop et al., 2010; Terui, Ban, and Allenby, 2011). A second, smaller group of articles, also models at least two stages, but makes use of available data on each of the shopping stages by incorporating it directly in the estimation (e.g., Franses and Vriens, 2004; Lambrecht, Seim, and Tucker, 2011; Abhishek, Fader, and Hosanagar, 2012; Chintagunta and Lee, 2012; De los Santos, Hortaçsu, and Wildenbeest, 2012; Honka, 2014; Moraga-González, Sándor, and Wildenbeest, 2016).

Further distinction should be made between work that has estimated consumers' consideration sets and work that has also modeled *how* consumers form their consideration sets. Examples of the former include Allenby and Ginter (1995), Siddarth, Bucklin, and Morrison (1995), Chiang, Chib, and Narasimhan (1998), Zhang (2006), Van Nierop et al. (2010), whereas examples of the latter include Mehta, Rajiv, and Srinivasan (2003), Kim, Albuquerque, and Bronnenberg (2010), Honka (2014), Moraga-González, Sándor, and Wildenbeest (2016), Honka and Chintagunta (2017). The latter set of articles is also part of a growing body of literature on consumer search. Although earlier literature developed search models without actually observing search in the data (e.g., Mehta, Rajiv, and Srinivasan, 2003; Hong and Shum, 2006), in the most recent search literature, search is observed in the data either directly through data on the consumers' consideration sets (e.g., De los Santos, Hortaçsu, and Wildenbeest, 2012; Honka, 2014) or indirectly through other variables (e.g., Kim, Albuquerque, and Bronnenberg, 2010). In this article, we develop a structural model of all three stages of the consumer purchase process where consumers form their consideration sets through costly search and we estimate the model using data on awareness, consideration, and choice.

Similar to this article, Goeree (2008) also studies the effects of advertising and the adequacy of the full information assumption in the personal computer (PC) industry. She shows evidence of consumers' limited information in this industry and that consumers' "choice sets" are driven by advertising. Our article differs from hers in several respects. First, we differentiate between consumer awareness and consideration sets, whereas Goeree (2008) remains agnostic on this issue and uses the term "choice sets." Second, she probabilistically models consumers' choice sets because she does not observe them in her data, whereas we observe consumers' awareness and consideration sets. Third, in Goeree's (2008) model, advertising can affect choice sets but cannot affect choice directly, whereas advertising can affect all three stages, awareness, consideration, and choice, in our model. Last, Goeree (2008) does not model how choice sets are formed; our model endogenizes consumers' choices of consideration sets by positing that consumers undertake costly search.

Finally, our article is also related to the literature examining consumer purchase behavior for financial services and products. Hortaçsu and Syverson (2004) study consumer purchase behavior for S&P 500 index funds and Allen, Clark, and Houde (2014) look at consumer behavior when buying mortgages. There is also a stream of literature on consumer adoption and usage of payment cards (e.g., Rysman, 2007; Cohen and Rysman, 2013; Koulayev et al., 2016; see also Rysman and Wright, 2014, for an overview). Lambrecht, Seim, and Tucker (2011) study the adoption of Internet-based customer self-service applications, such as online payments in the

German retail banking section, using a multistage model. Somewhat surprisingly, and despite its size and importance for both consumers and the economy, the literature on consumer demand for retail banks and their products is very sparse. Using survey data, Kiser (2002) finds that both advertising and prices and other bank-specific characteristics such as customer service are important drivers of consumers' bank choices. Dick (2008) and Wang (2016) develop aggregate-level, structural models of consumer demand for retail banks. Dick (2007), Hirtle (2007), and Ishii (2008) investigate branching structures and Dick (2007) and Mólnar, Violi, and Zhou (2013) study competition in retail banking. Similar to Ishii (2008), Dick (2008), and Wang (2016), we estimate demand for retail banks, but, in contrast to the aforementioned articles, our model describes consumer shopping and purchase behavior using consumer-level data.

3. Data

■ To conduct our analysis, we combine several data sets. We describe these data sets below, before turning to the presentation of our model and to the empirical results.

□ **Consumer-level data.** We benefit from access to survey data collected by a major market research company during March and April 2010, for a sample of 4280 respondents. Respondents were asked to refer to their bank shopping experiences during the previous 12 months. Given that we do not know the specific dates when respondents were shopping for banks, the studied period refers to bank activities (across all respondents) from March 2009 to April 2010 (herein referred to as "reference period").

In these data, we observe a consumer's previous and current primary bank[2]; the account types the consumer has with his primary and other banks; the banks the consumer is aware of (aided awareness)[3]; the banks the consumer considered during his search process[4]; the accounts the consumer moved from his previous to his current primary bank or opened with another (nonprimary) bank, and the identities of these banks.[5] In addition, we observe a nearly complete customer profile containing information on demographics such as age, gender, marital status and education.

For tractability reasons, we focus on the 18 largest financial institutions in the United States which had a combined national market share of 56% (measured in total deposits) in 2010. The leftmost column in Table 1 shows the list of included banks. We drop all respondents who have at least one institution in their consideration sets that is not among the 18 institutions listed. Further, we also remove all respondents with invalid zip codes. This results in a final sample of 2214 consumers. To ensure that dropping consumers does not introduce a selection problem, we compare the demographics of the initial and final set of respondents in Table A.1 in web Appendix A. The descriptives show that the final data set contains consumers with similar demographics to those in the initial data.

Table 2 shows descriptive statistics for all respondents in our final sample, as well as for the two subgroups of respondents: "shoppers" (1940 consumers) and "nonshoppers" (274 consumers). Shoppers are consumers who shopped and opened one or more new accounts, and nonshoppers are consumers who neither shopped nor opened new accounts during the reference

[2] There are many ways to define "primary bank," for example, by number of accounts, type of accounts, frequency of transactions, or some combination of these. In the survey, a definition was not provided to respondents, but most respondents indicated that this was the bank they had their primary checking account with.

[3] Aided awareness was obtained through the answer to the following survey question: "Please review the list below and select ALL the financial institutions that you recognize."

[4] The set of considered banks comes from the answer to the survey question: "Which of the financial institution(s) below did you investigate/research but decide not to get an account with?" (note that the answer to this question does not contain a consumer's final choice) and each consumer's final bank choice.

[5] Choices were coded based on the responses to the questions: "In the past 12 months, have you opened a new banking account?" and "Which financial institution did you open your new account(s) with?"

TABLE 1 Share of Respondents That Were Aware/Considered/Chose Each Bank

Institution	Aware	Considered	Chose
Bank of America	94.31	44.04	12.60
BB&T	17.62	5.92	3.39
Capital One	31.26	6.55	3.30
Chase/WaMu	72.85	31.48	12.47
Citibank	62.24	17.48	7.09
Citizens Bank	24.39	7.59	4.47
Comerica Bank	10.70	1.94	0.95
Fifth Third Bank	24.98	7.59	3.48
HSBC	20.69	7.32	2.98
Keybank	23.53	5.74	2.80
M&T	8.49	4.20	2.39
PNC/National City Bank	33.60	10.75	4.47
Regions Bank	21.45	6.10	3.12
Sovereign Bank	16.71	4.52	2.12
Suntrust Bank	31.07	11.70	7.54
TD Bank	20.42	8.58	3.84
U.S. Bank	31.93	13.60	8.85
Wells Fargo/Wachovia	87.35	35.00	14.14

This table presents the percentage of respondents in the sample that were aware, considered, or chose each of the institutions listed. The "Aware" column shows the breakdown of responses of 2214 respondents to the question "Please review the list below and select ALL the financial institutions that you recognize." The "Considered" column shows the breakdown of responses of 2214 respondents regarding the banks in their consideration sets. Consumers' consideration sets are the result of responses to the question "Which of the financial institution(s) below did you investigate/research but decide not to get an account with?" (note that the answer to this question does not contain the consumer's final choice) and to the set of questions describing a respondent's final choice (listed next). The "Chose" column shows the breakdown of the combined responses of 2214 respondents to the three questions: (1) "In the past 12 months, have you opened a new banking account?" (2) "Which one of these banks do you consider to be your primary financial institution for conducting your personal banking business?" and (3) "Which financial institution did you open your new account(s) with?" (responses to question (2) were used for the "Chose" column calculations whenever the response to question (1) was "No"). Note that all respondents that responded "No" to question (1) (274 respondents) are, by definition, "nonshoppers."

period.[6] We see that 62% of all respondents are female; 65% are between 30 and 59 years old; 79% are white; 33% are single/divorced; and 65% are married/with partner. With respect to income, households are almost equally distributed among the three categories "Under $49,999," "$50,000–$99,999," and "$100,000 and over," with the last category having a slightly smaller percentage of respondents than the other two. Finally, regarding education, 8% of all respondents have a high school degree or less, whereas the remaining 92% of respondents are evenly split among the "Some College," "College Graduate," and "Postgraduate" categories. Looking at shoppers and nonshoppers separately, we find nonshoppers to be older and to have lower income and less education.

We also observe the number and type(s) of bank account(s) respondents opened during the reference period.[7] On average, shoppers opened 2.25 different types of accounts with a minimum of 1 and a maximum of 10 account types within two months of opening a new account. Table 3 contains the percentages of shoppers that opened different types of accounts. The most common

[6] We use the terms "shop" and "search" interchangeably. Note that all the respondents in the data who were looking to open an account ended up doing so, that is, all shoppers/searchers ended up purchasing. More details on the different subgroups of respondents are provided in web Appendix B.

[7] The types of accounts considered in the survey fall into three groups. "Deposit accounts" include checking, savings, CD, and money market accounts. "Borrowing accounts" include credit cards, mortgages, home equity loans, or home equity lines of credit and personal loans (including auto loans and student loans). Last, "Investment accounts" include mutual funds/annuities and stocks/bonds.

TABLE 2 **Demographics By Respondent Type**

	Respondent Type		
	Shopper ($n = 1940$) %	Nonshopper ($n = 274$) %	All ($n = 2214$) %
Gender			
Female	61.5	60.9	61.5
Male	38.5	39.1	38.5
Age			
19–29	19.7	5.1	17.9
30–44	33.8	17.5	31.8
45–59	31.0	44.5	32.7
60+	15.4	32.8	17.6
Household Income			
Under $49,999	34.9	47.4	36.4
$50,000–$99,999	38.2	32.5	37.5
$100,000 and over	26.9	20.1	26.1
Race			
White	77.5	85.8	78.5
Black	5.7	4.0	5.5
Asian	9.7	4.4	9.1
Hispanic	5.3	1.8	4.9
Other	1.8	4.0	2.1
Education			
High school or less	6.8	12.8	7.5
Some college	30.8	35.8	31.4
College graduate	31.8	26.6	31.1
Postgraduate	30.6	24.8	29.9
Marital Status			
Single/Divorced	33.5	30.3	33.1
Married/Partner	64.4	66.1	64.6
Widowed	2.1	3.6	2.3
Region			
New England	6.5	3.6	6.1
Mid-Atlantic	28.3	18.6	27.1
Midwest	5.5	16.1	6.8
North Central	8.1	12.0	8.6
Southeast	8.8	7.7	8.6
South Central	3.0	5.1	3.3
Texas	4.6	4.7	4.7
Florida	11.0	7.7	10.6
Southwest	5.8	7.3	6.0
Northwest	4.6	5.1	4.7
California	12.8	9.9	12.5
Other	0.8	2.2	0.9

This table reports descriptive statistics for all respondents in our final sample as well as for the two subgroups of respondents: “shoppers” (1940 consumers) and “nonshoppers” (274 consumers). Shoppers are consumers who opened one or more new accounts, and nonshoppers are consumers who did not open new accounts during the reference period.

account types consumers shop for are checking accounts (85% of consumers), followed by savings accounts (58%), and credit cards (26%).[8]

[8] Note that we do not model consumers’ choices of bank account types in our structural model in Section 5. Instead, we model consumers’ choices of banks. The above information on bank account types is provided for descriptive purposes.

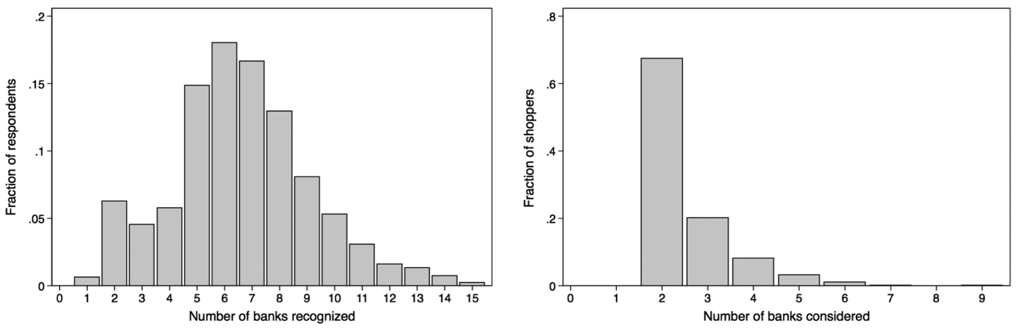
TABLE 3 Account Types Opened (Shoppers Only)

Account Type	% Respondents Opening	Account Type	% Respondents Opening
Deposit Accounts		*Borrowing Accounts*	
Checking	84.74	Credit card	25.67
Savings	57.78	Mortgage	9.23
Certificate of deposit	11.60	Home equity loan	6.08
Money market account	12.11	Personal loan	8.25
Investment Accounts			
Mutual funds	4.48		
Stocks/Bonds	4.12		

This table shows the types of new accounts opened during the reference period for the subsample of shoppers.

FIGURE 1

SIZE OF AWARENESS SETS AND CONSIDERATION SETS



(a) **Size of awareness sets** (b) **Size of shoppers' consideration sets**

This figure shows the distribution of awareness set sizes across all consumers (shoppers and nonshoppers) and the distribution of consideration set sizes for shoppers in our data.

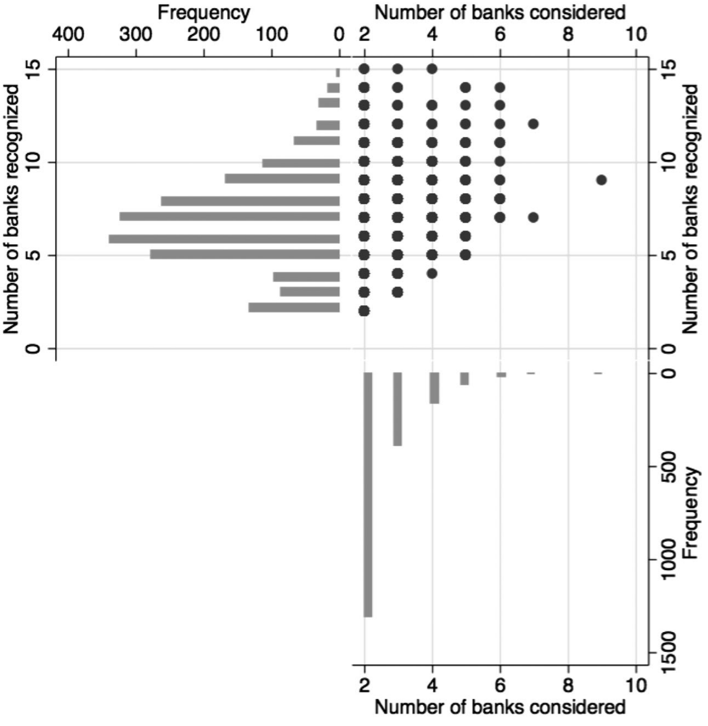
Table 1 displays the percentage of respondents who are aware of, consider, or choose a bank. The percentage of consumers who are aware of a given bank ranges from around 90% for the largest banks such as Bank of America and Wells Fargo/Wachovia to around 10% for the smaller banks in our data, such as M&T and Comerica Bank. Similarly, the percentage of consumers considering a given bank varies from around 40% for the larger banks to around 1%–2% for the smaller banks. Finally, the rightmost column in Table 1 shows the percentage of consumers who choose to open an account with each of the banks listed in the table. The purchase shares range from less than 1% to more than 13%.

Figure 1 shows histograms of the awareness and consideration set sizes. Consumers are, on average, aware of 6.8 banks and consider 2.5 banks. There is large variation in the sizes of consumers' awareness and consideration sets, which range from 2 to 15 and 2 to 9, respectively. Further, the relationship between the sizes of consumers' awareness and consideration sets is weak (see Figures 2 and 3). This suggests that there are distinct differences between these two types of sets and how they are formed and that looking at one of the stages may not be enough to understand consumers' choices. Indeed, the lack of a clear monotonic relationship between the two processes makes it difficult to justify estimating a "reduced-form" model that combines the two processes into one.

The differences between the awareness and consideration stages are further reflected in the large variation across consumers in what concerns which specific banks enter consumers' awareness and consideration sets (not tabulated). There are also large differences in the conversion

FIGURE 2

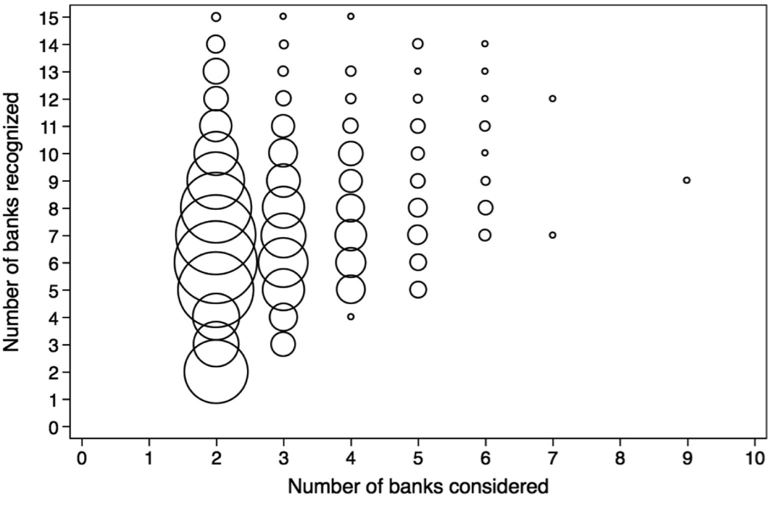
AWARENESS VERSUS CONSIDERATION (SHOPPERS ONLY)



This figure shows the distribution of awareness and consideration set sizes for shoppers in our data and the relationship between the sizes of the two types of sets.

FIGURE 3

AWARENESS VERSUS CONSIDERATION (SHOPPERS ONLY)



This figure shows the relationship between the size of the awareness sets and the size of the consideration sets for shoppers in our data. The area of each circle is proportional to the number of respondents for each combination of awareness and consideration set sizes.

rates from awareness to consideration and from consideration to choice/purchase across banks (see Table 1). For example, Bank of America, Chase/WaMu, M&T, and TD Bank get about 40% to 50% of consumers who are aware of these banks to consider them, whereas Capital One, Keybank, and Sovereign Bank get only 20% to 30% of consumers who are aware of these banks to consider them. Similarly, U.S. Bank, Suntrust Bank, and Citizens Bank have conversion rates between 60% and 75% from consideration to choice/purchase, whereas the conversion rates for Bank of America, Chase/WaMu, HSBC, and WellsFargo/Wachovia lie between 30% to 40%. Interestingly, it is not true that banks with the largest conversion rates from awareness to consideration also have the largest conversion rates from consideration to choice/purchase. For example, Bank of America has a very high conversion rate from awareness to consideration and a very low conversion rate from consideration to purchase. We see that the opposite is true for Comerica Bank and Keybank, for example. This holds true even when we compare banks with similar market shares. The market shares of HSBC, Keybank, M&T, and Sovereign Bank all lie between 2% and 3%. However, the awareness probabilities for this set of banks range from 8% to 23% indicating that predicting awareness from choice (and vice versa) is difficult.

□ **Sample representativeness.** The survey conducted by the market research company that provided us with the data focuses mostly on shoppers. We correct for the oversampling of shoppers by using weights in the model estimation so that the results are representative and accurately reflect the search and switching behavior of the overall U.S. population of retail banking consumers.

We calculate these sampling weights using information from another survey conducted by the same market research company, which they shared with us. This "screener" survey does not contain the same level of detail as the data described in Section 3, but has a much larger scale (around 100,000 respondents) and a sampling design that ensures population representativeness.[9] The sampling weights are used in all estimation specifications.

□ **Price data.** Previous work (e.g., Dick, 2008; Ishii, 2008) has imputed price data from deposit revenues (in the case of checking accounts) and from deposit expenses (in the case of savings deposits) because data on actual interest rates is typically available only from small-sample surveys. We benefit from access to a comprehensive database with branch-level deposit product prices. These price data, provided by RateWatch, include the interest rates and fees offered on various deposit products at the branch-level. The data are in panel format, that is, there are multiple measurements for the same branch and account type over time. We focus on the data for the reference period.

We combine the price data with the consumer-level data to obtain measures of the fees and interest rates that each consumer faced while shopping for a bank account. From the survey data, we know which respondents have checking and savings accounts with each bank and which banks were part of the respondents' consideration sets. Because we do not observe what specific types of checking or savings accounts respondents have, we focus on the account types that are most common within each bank; all of these correspond to some specific type of 2.5K savings account.[10,11] More specifically, we calculate the median (over time) interest rate for each bank in each respondent's zip code and use that information both to estimate the distribution of prices

[9] We refer the reader to web Appendix B for more details on the construction of the sampling weights.

[10] RateWatch has indicated that the amount of data points (over time and geographies) collected on deposit products at the branch-level is proportional to the market popularity of those products (i.e., data are collected more frequently and more consistently for products with higher share of deposits).

[11] A 2.5K savings account is a type of account that requires an average monthly balance of $2500 to avoid fees. We reestimated our model using data on all savings accounts weighted by the number of observations for each account type across geographies and time, and the results remain robust.

expected by the consumer prior to searching, and as a proxy for the rates that each respondent obtained upon searching among the banks in his consideration set.[12,13]

Specifying the interest rates for this type of account as the relevant price variable in the model estimation is reasonable because the correlation among the different savings accounts' rates and the checking accounts' fees across banks is very high. The Spearman rank correlation coefficient between the mean rates (across the reference period and geographies) for the most common 2.5K savings accounts and the most common 1K, 5K, and 10K savings accounts lies between 0.78 and 0.96 (the Cronbach's alpha is 0.83). Further, the Spearman rank correlation coefficient between the mean rates for the most common 2.5K savings accounts and the mean out-of-network ATM fees for the most common checking accounts is -0.72 (p value$=$ 0.02).[14,15]

Across all banks, the mean interest rate for the most common 2.5K savings accounts during the reference period is 0.135% and the standard deviation is 0.06%. Although the interest rates during the reference period were at a historically low level, the existing variation is suggestive that there were consumer gains to be had from search. In fact, in the data, we see that *ex ante* (i.e., before choosing a bank out of their consideration sets) consumers could get on average an 18 basis points higher interest rate by choosing the highest interest rate bank in their consideration set. *Ex post*, 52% of consumers end up choosing the bank with the highest interest rate, whereas 31% of consumers choose the second highest interest rate bank. This is consistent with consumers making decisions based on interest rates (after searching for such information).

□ **Advertising data.** Advertising data come from Kantar Media's *Ad$pender* database. Kantar tracks advertising expenditures and the number of advertisements (also called "units" or "placements"[16]) placed in national media (e.g., network TV and national newspapers) as well as in local media (e.g., spot TV and local newspapers) at the Designated Market Area (DMA) level. A DMA is a geographic region where the population can receive the same (or similar) television and radio station offerings.

We calculate total advertising expenditure and total number of placements by institution and DMA over the period from March 2009 until April 2010 (the reference period). Respondents' locations are identified by zip code and not DMA, so we match each respondent's zip code to a specific DMA to find how much each bank advertised in each respondent's DMA. Table A.2 in web Appendix A reports average advertising expenditures and the average number of placements at the DMA level for each bank during the reference period. In the estimation, we focus on placements as a measure of advertising intensity. This is so that we have a measure of advertising that is independent of the cost of advertising and that thus can be more easily compared across DMAs and banks.

Figure 4 shows the geographic distribution of DMA-level advertising placements for each of the "Big Four" banks (Bank of America, Chase/WaMu, Citibank, and Wells Fargo/Wachovia). The maps clearly show that there is significant variation in advertising intensity across banks and DMAs. This regional variation will be useful for identifying the effects of advertising on bank awareness and choice.

---

[12] Whenever zip code data for a specific bank in a respondent's consideration set were not available, we used data from branches located in adjacent zip codes.

[13] Henceforth, we will use the terms "price" and "interest rate" interchangeably.

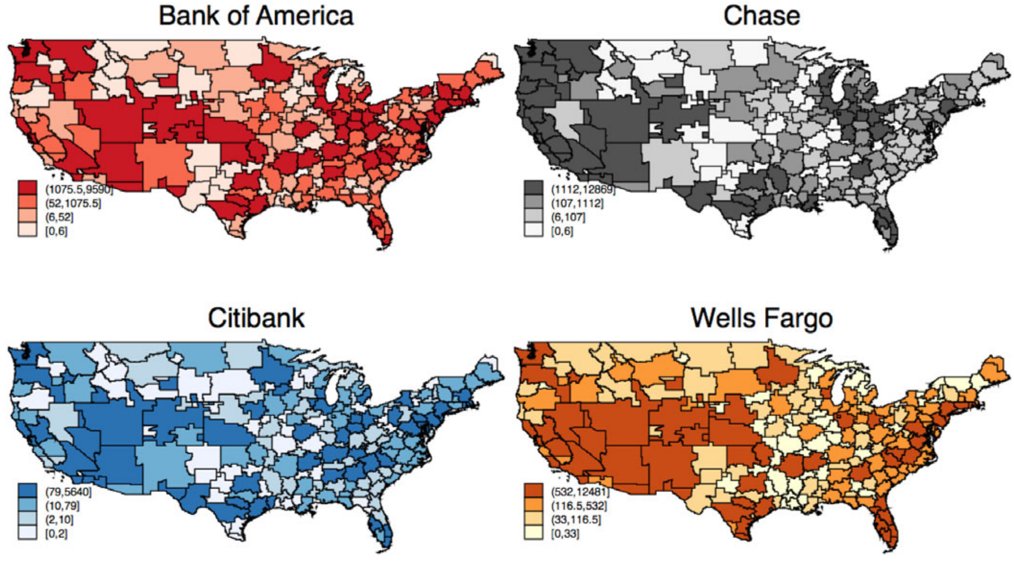
[14] According to the survey "The State of Checking Account Consumers in 2013" (thefinancialbrand.com/33346/bank-checking-account-customers-research/), the fees that irritate consumers the most are those triggered by out-of-network ATMs, so we choose to focus on those when calculating the correlations reported here.

[15] Because there are several banks that have exactly the same ATM fee ($2), we calculate the average of the interest rates within banks with the same fee before calculating the Spearman rank correlation coefficient.

[16] According to Kantar, "units" is simply the number of advertisements placed. These data are reported by Kantar without any weighting (based on spot length, size, etc.).

FIGURE 4

GEOGRAPHIC DISTRIBUTION OF THE "BIG FOUR" BANKS' DMA-LEVEL ADVERTISING (PLACEMENTS) [Color figure can be viewed at wileyonlinelibrary.com]



These maps display the spatial distribution of the total DMA-level number of advertising placements in the 206 DMAs in the continental United States over the reference period for the "Big Four" banks.

□ **Bank branch data.** In Mintel's "Retail Banking—U.S." October 2012 report, half of consumers stated that they chose their bank because there was a branch near their home—a significantly higher proportion than for any other reason listed. Nine in 10 consumers said that it is important to them to have a bank branch nearby. We use the respondents' five-digit zip code information to find their zip code centroid and to calculate the distance to the different institutions in their neighborhood using branch location data obtained from the FDIC.

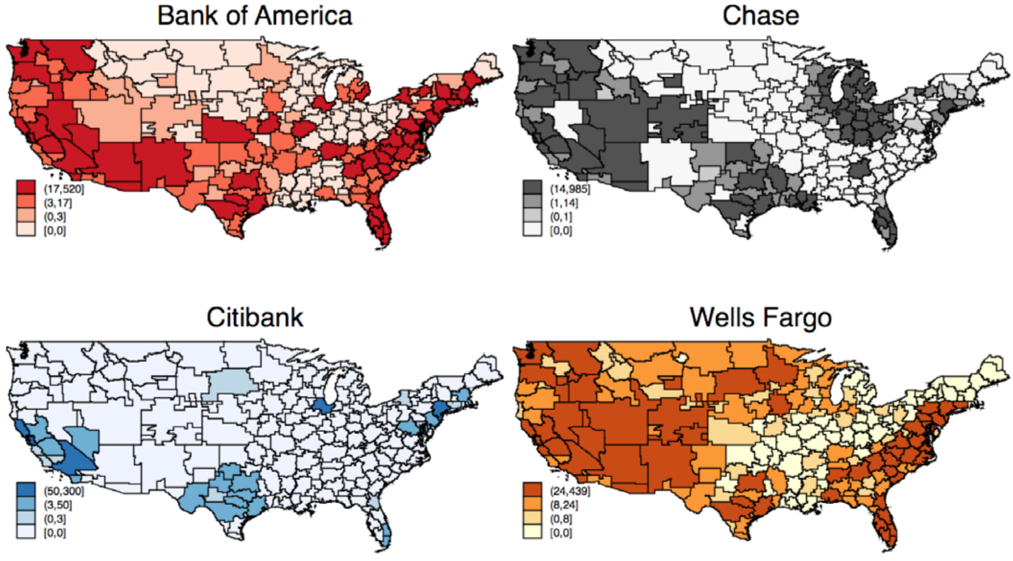
Our data confirm the crucial importance of local bank presence—that is, bank branches—in the consumer decision process: conditional on the consumer deciding to consider or to purchase from a bank, we find that the probability that the bank has a local branch within five miles of the consumer's home is on average 81% and 84%, respectively (Table A.3 in web Appendix A).

Figure 5 shows the geographic variation of branch presence (at the time of the survey) for each of the "Big Four" banks. Again, the large variation across banks and regions will be important for parameter identification purposes.

□ **Data limitations.** Although our data are well suited to study the consumer's shopping and purchase process for retail bank accounts because we observe awareness, consideration, and choice, the data have a few limitations. First, our data are cross-sectional. As a consequence, our ability to control for consumer-level unobserved heterogeneity, beyond the factors that are observable and that we use in the estimation, is limited. Second, our data do not contain information on credit unions, which have a significant share of the retail banking sector in the United States. Third, we do not observe the exact terms and conditions of the accounts opened by the survey respondents. To overcome this limitation, and as discussed in Section 3, we use interest rates for savings accounts as prices for all consumers. Hence, our interest rate/price data are to be interpreted as a proxy for the actual interest rates/prices observed by consumers.

FIGURE 5

GEOGRAPHIC DISTRIBUTION OF THE “BIG FOUR” BANKS’ DMA-LEVEL NUMBER OF BRANCHES
[Color figure can be viewed at wileyonlinelibrary.com]



These maps display the spatial distribution of the total DMA-level number of branches in the 206 DMAs in the continental United States in the reference period for the “Big Four” banks.

4. Evidence of consumers’ limited information

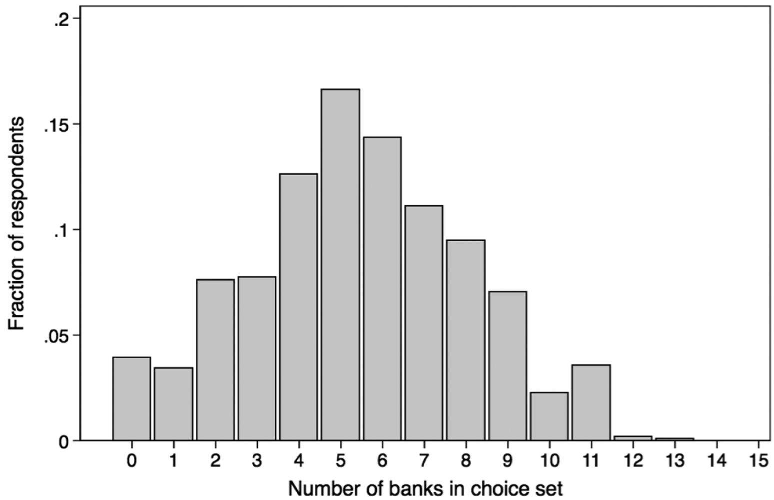
■ Access to data on consumers’ awareness and consideration sets is extremely rare. Previous empirical research studying demand for financial services has not had access to information on consumers’ consideration sets and therefore had to make assumptions about consumers’ “choice sets.”[17] For example, Dick (2008) assumes that consumers’ choice sets include all institutions with a physical presence in geographic proximity of the consumer. Here, we explore the consequences of four different assumptions on consumers’ choice sets when estimating demand for financial services. The four choice set definitions that we study are as follows: (i) choice set containing all 18 banks; (ii) choice set containing all banks that have at least one branch within five miles of the consumer’s zip code centroid[18] (“five-mile set” in what follows)—à la Dick (2008); (iii) choice set containing all banks the consumer is aware of; and (iv) choice set containing all banks the consumer considers (for shoppers) and all banks the consumer is aware of (for nonshoppers). The last choice set definition is the one that is closest to the sets truly faced by the consumers in our data (this is so because, although we observe the consideration sets faced by shoppers, we do not observe nonshoppers’ consideration sets because they do not shop and thus do not form consideration sets). Empirically, we evaluate the consequences of imposing each of these four choice set assumptions by estimating four reduced-form logit regressions. Note that we can do this only because awareness and consideration sets are observed in our data. Thus, we are in a privileged position to study the impact of making different assumptions regarding consumers’ choice sets on the estimated parameters.

[17] Note that, for the length of this section, we define “choice set” as the set of banks for which a consumer has full information about all characteristics (whether those are observed or unobserved by the researcher) and from which he chooses the bank to open one or more accounts with.

[18] To ensure that our results are not driven by our choice of a 5-mile radius, we also did the same analysis using a 10-mile radius and found similar results.

FIGURE 6

SIZE OF CHOICE SETS (FIVE-MILE RADIUS)



This figure shows the distribution of choice set sizes under the assumption that consumers' choice sets include all banks with at least one branch within five miles of consumers' zip code centroids.

Before discussing our empirical results, we provide descriptive evidence that five-mile sets are different from consideration sets and therefore, five-mile sets are imperfect measures of consumers' consideration sets. Figure 6 shows a histogram of the sizes of the five-mile sets. The distribution of the five-mile sets sizes is clearly different from the distribution of the consideration sets sizes (see Figure 1b). The average size of consumers' five-mile sets is 5.4, which is about three banks larger than the average consideration set size of 2.5 banks. There is also significant variation in the size differences between consumers' consideration sets and five-mile sets: some consumers have consideration sets that contain up to 4 banks more than their five-mile sets, whereas other consumers have consideration sets that contain up to 11 banks less than their five-mile sets (not tabulated). In addition, the shares of banks in the five-mile sets (not tabulated) are different from the shares of banks in consumers' consideration sets reported under the column "Considered" in Table 1. For example, Bank of America's share in consumers' consideration sets is 44.0%, whereas its share is 13.9% in five-mile sets. Similarly, Suntrust Bank has a share of 11.7% in consumers' consideration sets, whereas its share is 4.2% in five-mile sets.[19] Furthermore, 29.5% of consumers consider at least one bank outside of their five-mile set and 15.9% of consumers choose a bank outside of their five-mile set.[20]

The five-mile sets are also different from consumers' awareness sets and, hence, cannot be used as a proxy for these sets, either. Although the shape of the distribution of five-mile set sizes across consumers is similar to the shape of the distribution of awareness set sizes across consumers (see Figure 1a), awareness sets and five-mile sets of individual consumers tend to be quite different. More specifically, some consumers have awareness sets that contain up to 12 banks more than their five-mile sets, whereas other consumers have awareness sets that contain 10 banks less than their five-mile sets (not tabulated). Further, five-mile sets are identical (in terms of size and the identity of the banks in them) to awareness sets for only 5.7% of consumers

[19] Shares of the other banks in the five-mile sets are available from the authors upon request.

[20] If we extend the radius to 10 instead of 5 miles, 22.8% of consumers consider at least one bank outside of their 10-mile sets and 10.8% of consumers choose a bank outside of their 10-mile sets.

TABLE 4 Full Information Logit Choice Models

	(1) 18 banks	(2) Five miles	(3) Aw. Set	(4) Mix
Bank branches (Y/N)	2.107**		1.163**	1.210**
	(0.130)		(0.041)	(0.057)
Interest rate	−0.488	−0.858	0.306	0.504
	(0.534)	(0.581)	(0.404)	(0.452)
Advertising	0.050**	0.074**	0.019	0.014
	(0.015)	(0.024)	(0.017)	(0.022)
Primary bank	3.093**	2.685**	2.496**	2.229**
	(0.059)	(0.037)	(0.038)	(0.045)
Bank fixed effects	Yes	Yes	Yes	Yes
Ad elasticity	0.040	0.200	0.050	0.030
IR elasticity	−0.070	−0.080	0.040	0.040
N	2,214	1,831	2,214	2,214
Avg. choice set size	18	5.4	6.8	4.5

This table reports the results from four reduced-form logit regressions (with control-function correction) with different choice set definitions. In regression (1), consumers choose a bank from among all 18 banks in our data. In regression (2), consumers choose a bank from among all the banks that have at least one branch within five miles of the consumer's zip code centroid. In regression (3), consumers choose from among all the banks that they are aware of. In regression (4), shoppers choose from among the banks in their consideration sets whereas nonshoppers choose from among the banks in their awareness sets. Local bank presence is operationalized as a dummy variable that captures whether there is at least one branch of a given bank within five miles of each respondent's zip code centroid. Advertising corresponds to the number of placements at the DMA-level (in thousands). "Ad Elast" is the advertising elasticity and "IR Elast" is the interest rate elasticity. N is the number of respondents used in the estimation.
$^{*}p < 0.10$; $^{**}p < 0.05$.

(not tabulated).[21] Thus, we conclude that five-mile sets are different from both awareness and consideration sets.

We now demonstrate the consequences of different choice set assumptions on parameter estimates. To that end, we estimate four reduced-form logit regressions using the four alternative choice set definitions discussed at the beginning of this section. In all four regressions, we include the same set of independent variables: interest rates, advertising, a dummy variable for the primary bank, and a dummy variable indicating whether there is at least one branch within five miles of the consumer's zip code centroid. The only exception is the second regression, where we use five-mile sets and do not include a dummy variable indicating whether there is at least one branch within five miles of the consumer's zip code centroid, as the effect of such a dummy variable cannot be identified. The results are shown in Table 4. We start by discussing the results for the interest rate coefficients across the four regressions and then examine the results for the advertising coefficients.

Given that the interest rates we observe in our data are for deposit products, we expect consumers to prefer banks that offer higher interest rates to banks that offer lower interest rates, which should translate into a positive coefficient for the interest rate variable. Looking at the results from regression (1), in which consumers choose among all 18 banks, we find a negative coefficient for interest rates of −0.488. Similarly, we also find a negative coefficient on interest rates in regression (2) using five-mile sets. In contrast, once we use, arguably, the more relevant awareness and awareness/consideration sets as choice sets in regressions (3) and (4), respectively, we get positive coefficient estimates for interest rates of 0.306 and 0.504, respectively. Although the coefficients for the interest rate variable are not statistically significant, they move in the expected direction as consumers' choice set definitions become closer to the sets consumers truly use.

[21] If we extend the radius to 10 instead of 5 miles, 10-mile sets are the same as awareness sets for only 4.3% of consumers.

The intuition behind the change in the interest rate coefficient as we move across specifications is as follows: the higher the number of (irrelevant/incorrect) options that are in consumers' choice sets (which is the case in regressions (1) and (2) when compared to (3) and (4)), the more likely it is that consumers do not choose the option with the highest interest rate from within those sets. In the estimation of the choice model, this consumer behavior may then translate into an interest rate coefficient that suggests that consumers are insensitive to interest rates or, as in this specific case, even prefer lower to higher interest rates (holding everything else constant). As we make choice sets closer to what they truly are, the model becomes able to distinguish between consumers not choosing a bank with a high interest rate because they do not know about it (due to not being aware of that bank or not considering it) and consumers being insensitive to interest rates.

Next, we turn to the effects of advertising. In regressions (1) and (2), we estimate significant advertising coefficients of 0.050 and 0.074, respectively. Similar to the results for the interest rate variable, the picture changes once we look at regressions (3) and (4). Although the advertising coefficients remain positive, they become much smaller in magnitude and not statistically significant. Recall that consumers' awareness sets are used as choice sets in regression (3) and consumers' awareness/consideration sets are used as choice sets in regression (4). Thus, the results for the advertising variable indicate that, conditional on a consumer being aware of a bank, advertising does not have a statistically significant effect on consumer choice and that the prime mechanism through which advertising works is by increasing consumer awareness for a bank.

To summarize, we find that the definition of consumers' choice sets strongly influences empirical results. To get meaningful estimates of the effects of key variables such as interest rates and advertising on choice, the researcher has to carefully define choice sets. Consideration sets (where available) best describe the relevant set of banks among which a consumer makes a decision. Further, we find evidence suggestive of advertising primarily affecting consumer awareness and not choice. A model describing consumers' bank choice behavior needs to take these empirical observations into account.

5. Model

■ Our model describes the three stages of the purchase process: awareness, consideration, and choice. We view awareness as a passive occurrence, that is, the consumer does not exert any costly effort to become aware of a bank. A consumer can become aware of a bank by, for example, seeing an ad or driving by a bank branch. Consideration is an active occurrence, that is, the consumer exerts effort and incurs costs to learn about the interest rates offered by a bank. The consumer's consideration set is thus modeled as the outcome of a simultaneous search process, given the consumer's awareness set. Finally, purchase is an active, but effortless, occurrence in which the consumer chooses the bank which gives him the highest utility. The consumer's purchase decision is modeled as a choice model, given the consumer's consideration set. Consideration and choice are modeled in a consistent manner by specifying the same utility function for both stages (thus providing a structural interpretation for our estimates). This assumption is supported by Bronnenberg, Kim, and Mela (2016), who find that consumers behave similarly during the search and purchase stages.

□ **Awareness stage.** There are N consumers indexed by $i = 1, \ldots, N$. Consumer i lives in market m ($m = 1, \ldots, M$), and his awareness of bank j ($j = 1, \ldots, J$) is a function of bank fixed effects ς_{0j}, advertising adv_{jm}, demographic variables D_i, local bank presence b_{ij}, and an error term ξ_{ij}, and can be written as

$$a_{ijm} = \varsigma_{0j} + \varsigma_{1j} adv_{jm} + D_i \varsigma_{2j} + f(b_{ij})\varsigma_{3j} + \xi_{ij}, \quad \forall j \neq j_{PB}, \tag{1}$$

in which a_{ijm} is the latent awareness score for consumer i who lives in market m. We assume that consumer i is aware of bank j if his awareness score for bank j is larger than 0 and unaware otherwise. The term adv_{jm} denotes bank j's advertising intensity in market m (where consumer i resides). A market is defined as a DMA. The vector D_i includes observed demographic variables (age, gender, etc.), and $f(b_{ij})$ captures the effects of bank branches within five miles of consumer i's zip code centroid in a flexible way (we describe the specific functional form we use in the results section). The set of parameters to be estimated is given by $\theta_1 = (\varsigma_{0j}, \varsigma_{1j}, \varsigma_{2j}, \varsigma_{3j})$.

Note that we exclude the consumer's primary bank j_{PB} from the model, because we assume that consumers are aware of their primary bank. By this logic, we should also exclude any other banks the consumer has accounts with, given that the consumer should be aware of those banks as well. Unfortunately, although the survey data contain information on whether a consumer has other accounts other than those with his primary bank, it does not have information on the identities of the banks the consumer has one or more accounts with.

Last, note that we do not consider interest rates when modeling consumers' awareness sets, because a consumer logically cannot have interest rate beliefs for banks he is not aware of.

□ **Utility function.** Let u_{ijm} be the utility that consumer i who lives in market m obtains from bank j. Utility is specified as

$$u_{ijm} = \alpha_j + \beta_1 p_{ij} + \beta_2 I_{ij_{PB}} + \beta_3 adv_{jm} + f(b_{ij})\beta_4 + \epsilon_{ij}, \tag{2}$$

in which ϵ_{ij} is observed by the consumer but not by the researcher. The α_j terms are bank-specific brand intercepts and p_{ij} denotes prices. One of the challenges of modeling the consumers' shopping process for retail bank accounts stems from the definition of "price." In most retail settings, price is the posted amount the consumer has to pay to acquire a product. When it comes to retail banking, the definition of price is not as straightforward, as price can have multiple components such as fees and interest rates, and consumers can have multiple account types. As described in Section 3, we define "price" as the interest rate on 2.5K savings accounts. The term $I_{ij_{PB}}$ is a dummy variable indicating whether bank j is consumer $i's$ primary bank and $f(b_{ij})$ captures the effects of bank branches within five miles of consumer $i's$ zip code centroid in a flexible manner (to be described in the results section). The set of parameters to be estimated is given by $\theta_2 = (\alpha_j, \beta_1, \beta_2, \beta_3, \beta_4)$.

□ **Consideration stage.** The consumer makes the decisions of which and how many banks to search at the same time. For expository purposes, we first discuss the consumer's decision of which banks to search, followed by the consumer's decision of how many banks to search. Both decisions are jointly estimated. Consumers search for interest rates. We assume interest rates follow an Extreme Value Type I distribution with location parameter η and scale parameter τ. Consumers know the distribution of interest rates in the market, but search to learn the specific interest rate a bank will offer them.[22]

We use the approach developed by Honka and Chintagunta (2017) to identify the search method (simultaneous versus sequential) consumers use.[23] We find the proportion of above-expectation actual interest rates in consumers' consideration sets to be around 50% and constant across different consideration set sizes. This pattern indicates that consumers search in a simultaneous fashion.[24]

[22] Although there are websites that aggregate information on interest rates, they are not used by most consumers. Moreover, banks offer several products and consumers still need to search over exact rates. Further, 55% of shoppers in our data indicated that interest rates/fees were a shopping trigger for them.

[23] Honka and Chintagunta (2017) prove analytically that the proportion of below-price expectation actual prices among consumers searching once equals (is larger than) the probability of getting a below-price expectation actual price draw under simultaneous (sequential) search.

[24] Simultaneous search, also called fixed-sample or nonsequential search, means that consumers precommit to searching a specific set of companies or products and do not stop searching until they have collected the information from all companies or products in that set. The actual information collection can happen in a sequence.

Given the distributional assumption for interest rates, utility (from the consumer's perspective) u_{ijm} is an Extreme Value Type I distributed random variable with location parameter $a_{ij} = \alpha_j + \beta_1 \eta + \beta_2 I_{ij_{PB}} + \beta_3 adv_{jm} + f(b_{ij})\beta_4 + \epsilon_{ij}$ and scale parameter $g = \frac{\tau}{\beta_1}$. A consumer's search decision under simultaneous search depends on the Expected Indirect Utilities (EIUs; Chade and Smith, 2005). Consumer $i's$ EIU for bank j in market m, where the expectation is taken with respect to price, is given by

$$E[u_{ijm}] = \alpha_j + \beta_1 E[p] + \beta_2 I_{ij_{PB}} + \beta_3 adv_{jm} + f(b_{ij})\beta_4 + \epsilon_{ij}, \quad \forall j \in A_i. \tag{3}$$

Consumer i observes the EIUs for every bank he is aware of, including ϵ_{ij}. We also considered specifications where the consumer does not observe ϵ_{ij}. Note that such model specifications are computationally much simpler. Although the reality is most likely in between the perfectly observed (by the customer) ϵ_{ij} and the unobserved ϵ_{ij}, we believe that a model where consumers know attributes about banks that the econometrician does not is closer to reality. However, our main results as related to advertising are qualitatively robust to the alternative model specification where consumers do not observe ϵ_{ij}.

To decide which banks to search over, consumer i ranks all banks according to their EIUs (Chade and Smith, 2005) and then picks the top k banks to search from. The theory developed by Chade and Smith (2005) on the optimality of the ranking according to EIUs holds only under the assumption of first-order stochastic dominance among the interest rate distributions. Because we assume that interest rates follow a market-wide distribution, this assumption is automatically fulfilled. Further, we also need to impose a second restriction on the simultaneous search model to be able to use Chade and Smith (2005): search costs *cannot* be bank-specific.

To decide on the number of banks k, with $k \geq 0$, for which to obtain interest rate information, the consumer calculates the net benefit of all possible search sets *given the ranking of the EIUs*. A consumer's benefit of a searched set S_i is then given by the expected maximum utility among the searched banks. The term R_{ik} denotes the set of top k banks consumer i ranked the highest according to their EIUs. For example, R_{i1} contains the bank with the highest expected utility for consumer i, R_{i2} contains the banks with the two highest expected utilities for consumer i, etc. The consumer picks the size of his searched set S_i which maximizes his net benefit of searching denoted by Γ_{ik}, that is, the expected maximum utility among the searched banks minus the cost of search

$$\Gamma_{ik} = E\left[\max_{j \in R_{ik}} u_{ijm}\right] - kc, \tag{4}$$

in which c denotes the consumer's search cost (per bank searched).[25] The consumer picks the number of searches k which maximizes his net benefit from searching.

□ **Choice stage.** After the consumer has formed his consideration set and learned the interest rates of the considered banks, all uncertainty is resolved. At this stage, both the consumer and the researcher observe the actual interest rates. The consumer then chooses the bank with the highest utility among the searched banks, that is,

$$j = \arg\max_{j \in S_i} u_{ijm}, \tag{5}$$

in which u_{ijm} now contains the actual interest rate of bank j faced by consumer i in market m and S_i is the set of searched banks.

[25] Note that the effects of variables such as advertising on consumers' search costs and utility function are not separately identified without further assumptions. See, for example, Honka (2014) and Yao, Wang, and Chen (forthcoming) for a discussion. Accordingly, we assume that advertising affects consumers' utility.

6. Identification

■ The identification of the utility parameters comes from both consumers' consideration and choice/purchase decisions. Observing consumers' account opening decisions allows us to identify the parameters capturing differences in the bank intercepts and the effects of advertising, price, and bank branches that vary across banks. Identification is standard, as in a conditional choice model. Observing consumers' consideration decisions allows us to better identify the aforementioned parameters and to additionally also identify the base bank intercept.

The size of a consumer's consideration set helps to pin down search costs. We can identify only a range of search costs because it is utility-maximizing for all consumers with search costs in that range to search a specific number of times. Beyond the fact that a consumer's search cost lies within a range that rationalizes searching a specific number of times, the variation in our data does not identify a point estimate for search costs. The search cost point estimate is identified by the functional form of the utility function and the distributional assumption on the unobserved part of the utility.

The composition of a consumer's consideration set allows us to better identify the parameters in the utility function when compared to using information from the choice decisions alone (i.e., the choice of one option out of the consideration set). This is so because, when deciding over which consideration set to search, consumers compare the expected maximum utility from different possible sets, which provides an additional source of identification.

The base bank intercept is identified from the consumer's decision to search or not to search, that is, having nonshoppers in our data is essential to identify this parameter. Intuitively, the option not to search and not to open at least one account is the outside option and allows us to identify the base bank intercept. So, the search cost estimate is pinned down by the average number of searches, whereas the base bank intercept is identified by the consumer's decision to search or not.

7. Estimation

■ The unconditional purchase/choice probability is given by

$$P_{ij} = P_{iA_i} \cdot P_{iS_i|A_i} \cdot P_{ij|S_i}, \tag{6}$$

in which P_{iA_i} is the probability that consumer i is aware of the set of banks denoted by A_i, $P_{iS_i|A_i}$ denotes the probability that consumer i searches set S_i given his awareness set A_i, and $P_{ij|S_i}$ is the probability that consumer i chooses bank j conditional on his consideration set S_i. In the following three subsections, we discuss how each of these probabilities are estimated. Note that, for computational reasons, we assume that there is no correlation between the unobservables in the awareness and the utility functions, that is, the awareness probability P_{iA_i} does not have any error terms (or parameters) in common with the conditional consideration and conditional purchase probabilities. Thus, the awareness stage can be estimated separately from the other stages.

□ **Awareness stage.** We assume that the error term ξ_{ij} follows an Extreme Value Type I distribution. This allows us to estimate equation (1) as a binary logit regression for each bank j separately. The probability that consumer i is aware of bank j is then given by

$$P(a_{ij} > 0) = \frac{\exp(\varsigma_{0j} + \varsigma_{1j} adv_{jm} + D_i \varsigma_{2j} + f(b_{ij})\varsigma_{3j})}{1 + \exp(\varsigma_{0j} + \varsigma_{1j} adv_{jm} + D_i \varsigma_{2j} + f(b_{ij})\varsigma_{3j})}, \tag{7}$$

and the probability that consumer i is aware of the set A_i of banks is denoted by

$$P_{iA_i} = \prod_{j=1}^{J} P(a_{ij} > 0)^{\phi_{ij}} (1 - P(a_{ij} > 0))^{1-\phi_{ij}}, \tag{8}$$

in which ϕ_{ij} equals 1 if consumer i is aware of bank j and equals zero otherwise.

□ **Consideration stage.** We start by pointing out the crucial differences between what the consumer observes in our model and what the researcher observes:

1. The consumer knows the distribution of prices in the market, but the researcher has to infer this distribution from available data;
2. The consumer knows each company's position in the EIU ranking, but the researcher only partially observes the ranking by observing which companies are being searched and which ones are not being searched;
3. In contrast to the consumer, the researcher does not observe ϵ_{ij}.

The first point implies that the price distribution needs to be inferred from the data. The typical assumption of rational expectations (e.g., Mehta, Rajiv, and Srinivasan, 2003; Hong and Shum, 2006; Moraga-González and Wildenbeest, 2008; Honka, 2014; Honka and Chintagunta, 2017) is that this distribution can be estimated from the prices observed in the data. Because the parameters of the price distribution are estimated, we need to account for sampling error when estimating the other parameters of the model. We do so by integrating over the empirical distributions of the estimated parameters (McFadden, 1986).

To address the second point, we note that partially observing the EIU ranking of companies provides information that allows us to estimate the composition of consideration sets. Honka (2014) has shown that the following condition has to hold for any searched set:

$$\min_{j \in S_i}(E[u_{ijm}]) \geq \max_{j' \notin S_i}(E[u_{ij'm}]) \cap \Gamma_{ik} \geq \Gamma_{ik'} \quad \forall k \neq k', \tag{9}$$

that is, the minimum EIU among the searched banks is larger than the maximum EIU among the nonsearched banks *and* the net benefit of the chosen searched set of size k is larger than the net benefit of any other search set of size k'.

To address the last point, we assume that ϵ_{ij} has an Extreme Value Type I distribution with location parameter 0 and scale parameter 1, and integrate over its distribution to obtain the corresponding probabilities with which we can compute the likelihood function. Then, the probability that a consumer picks a consideration set $S_i = \Upsilon$ is

$$P_{iS_i|A_i,\epsilon} = \Pr\left(\min_{j \in S_i}(E[u_{ijm}]) \geq \max_{j' \notin S_i}(E[u_{ij'm}]) \cap \Gamma_{ik} \geq \Gamma_{ik'} \quad \forall k \neq k'\right). \tag{10}$$

□ **Choice stage.** We now turn to the choice decision stage given consideration. The consumer's choice probability conditional on his consideration set is

$$P_{ij|S_i,\epsilon} = \left(u_{ijm} \geq u_{ij'm} \quad \forall j \neq j', \quad j, j' \in S_i\right), \tag{11}$$

in which we now include the actual prices in the utility function. Note that there is a selection issue: given a consumer's search decision, the ϵ_{ij} do not follow an Extreme Value Type I distribution and the conditional choice probabilities do not have a logit form. We solve this selection issue by using Simulated Maximum Likelihood Estimation (SMLE) when estimating the conditional choice probabilities.

In summary, the researcher estimates the price distributions, only partially observes the utility rankings, and does neither observe ξ_{ij} in the consumer's awareness nor ϵ_{ij} in the consumer's utility function. Given this, our model has awareness probability given by equation (8), conditional consideration set probability given by equation (10), and conditional choice probability given by equation (11). We maximize the joint likelihood of the consumer's awareness set, consideration set, and choice. The likelihood of our model is given by

$$L = \prod_{i=1}^{N} w_i \cdot P_{iA_i} \cdot \left[\int_{-\infty}^{+\infty} \prod_{l=1}^{L} \prod_{j=1}^{J} P_{iS_i|A_i,\epsilon}^{\vartheta_{il}} \cdot P_{ij|S_i,\epsilon}^{\delta_{ij}} f(\epsilon)\, \mathrm{d}\epsilon\right], \tag{12}$$

in which w_i are the consumer-specific representativeness weights, ϑ_{il} indicates the chosen consideration set, and δ_{ij} the bank with which the consumer chooses to open one or more accounts with. The set of parameters to be estimated is given by $\theta = \{\theta_1, \theta_2, c\}$.

Neither the consideration set probability as shown is equation (10) nor the choice probability as shown in equation (11) have a closed-form solution. We use SMLE to estimate the consideration set and choice probabilities described in equations (10) and (11), respectively, and provide details about our estimation approach in web Appendix C.

□ **Advertising endogeneity.** One estimation concern is the potential endogeneity of the advertising intensity variable, which may arise both in the awareness and utility equations (equations (1) and (2), respectively).[26] The correlation between advertising intensity and the unobserved portions of latent awareness and latent utility is caused by omitted variables: the econometrician does not observe all the factors that affect consumers' awareness or utility and that may be correlated with advertising intensity. For example, banks may set advertising levels according to their regional performance measures, for instance, as a function of levels of customer satisfaction. Because customer satisfaction is not observed by the researcher, but may be observed by the bank management, this can give rise to endogeneity concerns. Ideally, this advertising endogeneity issue could be addressed by including bank-specific regional fixed effects in the awareness and utility equations. Unfortunately, the number of respondents that we have for each bank-region combination does not allow us to follow this approach.

We allow for endogeneity in advertising intensity adv_{jm} and address the problem that ξ_{ij} in equation (1), and ϵ_{ij} in equation (2) may not be independent of adv_{jm} by using the control-function approach (Hausman, 1978; Heckman, 1978; Blundell and Powell, 2004; Petrin and Train, 2010). The idea behind the control-function correction is to derive a proxy variable that conditions on the part of adv_{jm} that depends on ξ_{ij} and on ϵ_{ij} so that the remaining variation in the endogenous variable becomes independent of the errors.

More formally, let the endogenous explanatory variable adv_{jm} be expressed as a linear function of all relevant exogenous variables, denoted as X, which enter the latent awareness and the utility function specifications, of the variables Z that do not enter latent awareness or utility directly but affect adv_{jm} (also called instruments), and of an unobserved term μ_{jm}[27]

$$adv_{jm} = \alpha_j + X\beta_j + Z_{jm}\gamma_j + \mu_{jm}. \tag{13}$$

Substituting this expression for adv_{jm} into equations (1) and (2), the endogeneity issue becomes clear. Although ξ_{ij}, ϵ_{ij}, and μ_{jm} are independent of X and Z_{jm}, μ_{jm} is correlated with ξ_{ij} and with ϵ_{ij}. This correlation implies that adv_{jm} is correlated with ξ_{ij} and with ϵ_{ij}, and this is the source of the endogeneity concern. More specifically, there may be bank-DMA-specific variables that affect all consumers living in a given DMA and that are not observed by the researcher, such as the regional branch performance mentioned above. Note that we assume that μ_{jm} and ξ_{ij} and ϵ_{ij} are independent for all $k \neq j$.

Further, we specify ξ_{ij} in the awareness equation (1) to consist of a part that is correlated with adv_{jm} and that can be explained by a general function of μ_{jm} (a first-order approximation can be $\lambda \cdot \mu_{jm}$) and a part that is i.i.d. Extreme Value Type I. In particular, let

$$\xi_{ij} = CF(\mu_{jm}; \lambda_a) + \tilde{\xi}_{ij}, \tag{14}$$

and, similarly, for the utility equation (2) we can write

$$\epsilon_{ij} = CF(\mu_{jm}; \lambda_u) + \tilde{\epsilon}_{ij}, \tag{15}$$

[26] We do not explicitly account for endogeneity in interest rates because there is less geographic variation in interest rates across branches within a bank and we include bank-level fixed effects in both the awareness and consideration/choice models.

[27] Note that the point of the advertising equation is not to model advertising-setting behavior, but rather to have a robust way of uncovering the true parameters of the awareness and utility models (Rivers and Vuong, 1988; Villas-Boas, 2007; Petrin and Train, 2010).

in which CF(·) denotes the control functions with parameters λ_a and λ_u, and $\tilde{\xi}_{ij}$ and $\tilde{\epsilon}_{ij}$ are i.i.d. Extreme Value Type I.

To use the control-function approach, the model is estimated in two steps. First, the advertising equation (13) is estimated and its residuals $\hat{\mu}_{jm}$ are retained. To estimate equation (13), we use X variables that capture the bank presence across DMAs and that are similar to the bank presence variables included in the awareness and utility equations. Namely, we use an indicator variable for whether the bank has a branch present in the DMA and the number of branches it has in that DMA. As instruments (Z variables), we use the number of TV homes for each DMA (which captures the size of each market) and the cost of advertisements at the DMA- and bank-level. Market size and advertising costs act as exogenous shifters of advertising placement decisions because they are likely to be correlated with advertising intensity but uncorrelated with latent awareness or utility. Average advertising costs are calculated for each type of media by using total advertising expenditures and units across all industries for each market and media type. Because different banks have different allocations of advertising units across media, we calculate an average advertising cost per bank and DMA (weighted by the share allocated to each media; we assume that this allocation share is not determined at the DMA level, but is set at the bank-level). This means that advertising costs are not just market- but also bank-specific.

Second, both the awareness model and the consideration/choice model are estimated with these residuals entering as explanatory variables. These residuals are the control functions that are included to account for omitted attributes. Because the second step uses an estimate of μ_{jm} from the first step as opposed to the true μ_{jm}, the asymptotic sampling variance of the second-step estimator needs to take this extra source of variation into account. We implement a bootstrap procedure to address this issue.

8. Results

□ **Awareness stage.** We start by discussing our results on consumer awareness for retail banks. Table 5 shows the estimates from 18 (one for each bank) binary logit regressions (we refer to these regressions as model (A)—which stands for "Awareness"). Each regression includes a bank intercept and, as regressors, advertising intensity, local bank branch presence, and demographics. We control for local bank presence in a flexible manner by including two dummy variables: the first one indicating whether there is exactly one branch within five miles of the consumer's zip code centroid; and the second one indicating whether there are two or more branches within five miles of the consumer's zip code centroid.[28] Further, we control for a set of demographic variables, namely, age, gender, marital status, race, income, and education.

All bank intercepts other than those for Citibank, Chase/WaMu, Wells Fargo/Wachovia, and Capital One are negative and significant (not tabulated). The coefficients associated with the two dummy variables describing the effects of local branch presence are both positive and significant. Further, the effects of advertising (measured in 1000 units/placements) are positive for all banks with the exception of Wells Fargo/Wachovia, which has a negative but not significant coefficient. In addition, the advertising coefficients for three of the "Big Four" banks (Bank of America, Chase/WaMu, Wells Fargo/Wachovia) are not significant. This could be because of the very high baseline awareness levels for these banks, possibly built based on past (as discussed in Vitorino 2014) or current national advertising which would be captured (together with other factors that are constant across regions) by the bank fixed effects.

To interpret the impact of advertising on awareness in economic terms, we convert the advertising coefficients in the Awareness Model into advertising elasticities for each bank (reported under the column labelled "Awareness" in Table 6—we examine the full set of results in this

[28] We also estimated awareness models with separate dummy variables for a larger number of branches. The results obtained there were similar, and so we chose the most parsimonious specification.

TABLE 5 Results from Awareness Stage

	Model (A): Awareness		
		Local Bank Presence	
	Advertising	One Branch	>1 Branch
Bank of America	0.029	0.250	0.616
	(0.153)	(0.943)	(0.559)
BB&T	4.945**	2.628**	2.691**
	(1.896)	(0.508)	(0.579)
Citibank	0.539**	1.057	1.315**
	(0.149)	(0.647)	(0.402)
Citizens Bank	2.287**	0.330	0.967*
	(0.486)	(0.736)	(0.554)
Comerica	6.199**	1.093	2.369**
	(1.688)	(0.876)	(0.807)
Fifth Third	1.514*	1.919*	2.075**
	(0.864)	(0.992)	(0.639)
HSBC	0.864**	0.358	0.331
	(0.163)	(0.572)	(0.464)
Chase/WaMu	0.052	1.193**	0.888**
	(0.083)	(0.607)	(0.384)
Keybank	0.682*	1.388**	3.546**
	(0.402)	(0.664)	(0.632)
M&T	2.938**	1.847**	2.608**
	(1.296)	(0.805)	(0.792)
PNC/N. City Bank	0.247**	0.942**	1.870**
	(0.083)	(0.462)	(0.308)
Regions	2.677**	2.797**	2.964**
	(0.641)	(0.653)	(0.584)
Sovereign	1.066**	3.701**	3.371**
	(0.478)	(1.416)	(0.649)
SunTrust	1.258**	2.610	3.006**
	(0.609)	(2.891)	(0.610)
TD	0.863**	3.183**	1.402**
	(0.168)	(0.831)	(0.662)
U.S. Bank	1.976**	2.994**	2.766**
	(0.828)	(0.799)	(0.507)
Wells Fargo/Wachovia	−0.034	2.440**	0.212
	(0.111)	(0.540)	(0.502)
Capital One	0.056**	4.834**	2.503**
	(0.025)	(0.750)	(0.706)

This table reports the results from the Awareness stage model (denoted as model (A)). A standard logit model with control-function correction was estimated for each bank listed (for a comparison with results without the control-function correction, please refer to Table D.1 in web Appendix D). The dependent variable for each model is a self-reported dummy variable indicating whether the consumer is aware of a given bank. The second column reports the parameters associated with the advertising variable for each estimated model, and the third and fourth columns report the parameters associated with the local bank presence variables. Advertising corresponds to the number of placements at the DMA-level (in thousands). Local bank presence is operationalized as two dummy variables that capture whether there is one branch or more than one branch, respectively, of a given bank within five miles of each respondent's zip code centroid. All models include the following variables, which are omitted from this table: intercept, respondent demographics (namely, age, gender, marital status, race, income, and education) and the advertising residual that resulted from the control-function approach first stage. Standard errors were calculated via bootstrapping and are reported in parentheses under the coefficient estimates.
$*p < 0.10$; $**p < 0.05$.

table in the third subsection of Section 8). The average elasticity across all banks is 0.10, which means that when advertising (measured in 1000 units/placements) is increased by 1%, the average probability of awareness (across all banks and consumers) increases by 0.10%. Evaluated at the mean levels (across all banks and respondents) of advertising and awareness (1145 units and 38%,

TABLE 6 Advertising Elasticities

Bank	Awareness (A)	Choice (CC-LI)
Bank of America	0.01	0.03
BB&T	0.10	0.00
Citibank	0.10	0.02
Citizens Bank	0.21	0.01
Comerica	0.15	0.01
Fifth Third	0.11	0.02
HSBC	0.16	0.02
Chase/WaMu	0.05	0.03
Keybank	0.07	0.03
M&T	0.09	0.01
PNC/N. City Bank	0.15	0.02
Regions	0.06	0.01
Sovereign	0.13	0.01
SunTrust	0.09	0.02
TD	0.25	0.03
U.S. Bank	0.12	0.00
Wells Fargo/Wachovia	−0.01	0.03
Capital One	0.05	0.06
Average	0.10	0.02

This table reports the advertising elasticities that correspond to the model estimates reported in Tables 5 and 7. Elasticities for each bank are calculated by first calculating individual-level elasticities for each respondent and bank and then averaging across respondents within banks, taking the representativeness weights into account. The reported "Average" elasticities at the bottom of the table are calculated as simple averages of the bank-specific elasticities.

respectively) this means that, if the number of advertising units is increased by 11.45 units,[29] we expect the probability of awareness to go up from 38% to 38.038% (i.e., an additional 4 in 10,000 people become aware).

Finally, we also find several parameters associated with demographic variables to be significant (not tabulated). For example, older consumers are more aware of banks such as BB&T, M&T, and Wells Fargo/Wachovia, and consumers with some college education are more aware of banks such as Chase/WaMu, Keybank, U.S. Bank, and Wells Fargo/Wachovia.

□ **Consideration and choice stages.** Table 7 shows the estimates for the consideration and choice (CC) parts for both the limited information (LI) model (CC-LI), and, for comparison purposes, for the full information (FI) model (CC-FI), which we describe in detail below. We operationalize the effects of local bank presence by including four dummy variables: the first one indicating whether there is exactly one branch; the second one indicating whether there are two or three branches; the third one indicating whether there are four to seven branches; and the fourth one indicating whether there are eight or more branches within five miles of the consumer's zip code centroid. We use this flexible functional form for local bank presence to allow us to capture the decreasing marginal utility consumers get as the number of bank branches in their vicinity increases.

For the (CC-LI) model, the results in Table 7 show that all coefficients, with the exception of the advertising residual from the control-function approach, are significant (but note that the advertising coefficient becomes significant and increases in magnitude when we use the control-function approach—see Table D.1 in web Appendix D—hence, including the control function is indeed important here). As expected, local bank presence increases consumers' utility for a

[29] If we use a rough estimate for advertising costs of $20 per unit advertised (calculated across all types of media, DMAs, and banks), 11.45 units of advertising translate into about $230 per 13-month period (i.e., for a period of the same length as the reference period).

TABLE 7 Results from Consideration and Choice Stages

	(CC-FI)	(CC-LI)
Bank Intercepts		
Bank of America	−0.123	−4.016**
	(0.170)	(0.083)
BB&T	−0.486*	−3.648**
	(0.251)	(0.161)
Citibank	0.086	−4.037**
	(0.271)	(0.105)
Citizens Bank	−0.032	−3.681**
	(0.208)	(0.162)
Comerica	−1.003**	−4.505**
	(0.419)	(0.261)
Fifth Third	−0.281	−4.051**
	(0.215)	(0.142)
HSBC	0.285	−3.285**
	(0.254)	(0.138)
Chase/WaMu		−3.829**
		(0.078)
Keybank	−0.527	−4.124**
	(0.343)	(0.161)
M&T	−0.410**	−3.437**
	(0.205)	(0.189)
PNC/N. City Bank	−0.642**	−4.043**
	(0.245)	(0.105)
Regions	−0.612**	−4.130**
	(0.249)	(0.149)
Sovereign	−0.264	−3.841**
	(0.235)	(0.188)
SunTrust	0.128	−3.594**
	(0.160)	(0.112)
TD	−0.768**	−3.536**
	(0.266)	(0.143)
U.S. Bank	0.152	−3.929**
	(0.164)	(0.115)
Wells Fargo/Wachovia	0.073	−4.088**
	(0.150)	(0.086)
Capital One	−0.141	−4.390**
	(0.600)	(0.231)
Other Parameters		
Primary bank	3.020**	0.382**
	(0.061)	(0.030)
Interest rates	−0.575	2.104**
	(0.526)	(0.553)
Advertising	0.038**	0.015*
	(0.015)	(0.009)
Advertising residual	−0.043	−0.015
	(0.039)	(0.022)
Bank branches, N=1 (Y/N)	1.972**	0.504**
	(0.154)	(0.091)
Bank branches, $3 \geq N \geq 2$ (Y/N)	1.949**	0.655**
	(0.175)	(0.081)
Bank branches, $7 \geq N \geq 4$ (Y/N)	2.225**	0.796**
	(0.163)	(0.083)

(*Continued*)

TABLE 7 Continued

	(CC-FI)	(CC-LI)
Bank branches, $N > 7$ (Y/N)	2.747**	1.034**
	(0.205)	(0.094)
Search cost constant		0.001**
		(0.000)

This table reports the results from two different model specifications for the Consideration and Choice (CC) stages. Specification (FI) corresponds to a Full Information Model, equivalent to a traditional multinomial logit model in which consumers are assumed to be aware and consider all banks, and know banks' actual interest rates without engaging in search. Specification (LI) corresponds to a model that accounts for consumers' limited information. In both models, local bank presence is operationalized as a set of four dummy variables, capturing whether there is one branch, whether there are two to three branches, four to seven branches, or more than seven branches of a given bank within five miles of each respondent's zip code centroid. Advertising corresponds to the number of placements at the DMA-level (in thousands). All reported results reflect a control-function correction for potential endogeneity of the advertising variable; for a comparison with results without the control-function correction, please refer to Table D.2 in web Appendix D. Standard errors were calculated via bootstrapping and are reported in parentheses under the coefficient estimates.
$*p < 0.10$; $**p < 0.05$.

bank and the size of the effect varies with the number of branches. Having one branch increases consumer utility by 0.504, but the average effects of additional branches are significantly lower. The estimated coefficient for the "Primary Bank" dummy variable suggests that switching costs with regard to a consumer changing his primary bank are also an important factor of demand in this market. To summarize, being a consumer's primary bank, having high interest rates on savings accounts, and local bank presence increase a consumer's utility for a bank.

The estimated consumer search costs for retail banks (measured in interest rate percentage points) are 0.04 percentage points per bank searched. This translates to about \$2 per account searched for a \$5000 deposit in a 2.5K-type account (given that the interest rates in the data are for 2.5K accounts). Interestingly, at first sight, search costs seem to be on the smaller side compared to other search cost estimates in the financial products industry. For example, Hortaçsu and Syverson (2004) find median search costs to be between 5 and 21 basis points for S&P 500 index funds (typically purchased by more financially sophisticated and higher-income individuals). However, if we compare the obtained search costs with the possible returns from the low interest rates for deposits which were in place during the reference period, we see that, relatively speaking, the search costs can actually be considered high. For example, using the average interest rate of 0.135% mentioned in Section 3, we calculate the annual return/interest earned for a \$5000 deposit to be of \$6.75. Search costs per account searched are about 30% of this amount.

We now compare our estimates from the limited information model to those obtained from a model under full information (CC-FI) that has the exact same specification as our limited information search model (CC-LI).[30] The full information model is misspecified in the sense that it assumes that consumers are aware of and consider all banks when deciding which bank they would like to open one or more accounts with. Further, consumers are allowed to know the actual interest rates that any bank in the data will offer them (because this is a model without search). Under these assumptions, the full information model can be estimated as a multinomial logit model. The estimation results are shown in Table 7. Because we cannot directly compare the coefficients from the two models (CC-FI) and (CC-LI), we translate the coefficients into elasticities and marginal effects when needed.

The advertising elasticities implied by the two models are very similar, with the advertising elasticity being slightly lower in the full information model (0.027 in model (CC-FI) and 0.032 in model (CC-LI)—not tabulated). The coefficient estimate for interest rates is negative and not

[30] The only difference between the model (CC-FI) in Table 7 and the model "18 Banks" in Table 4 is a more flexible local presence specification using four dummy variables instead of two dummy variables.

significant in model (CC-FI). The marginal effects associated with the primary bank variable are more than 10 times larger in model (CC-FI) when compared to model (CC-LI) (2.85 for model (CC-FI) versus 0.23 for model (CC-LI)—not tabulated).[31] Last, the marginal effects of bank branches in model (CC-FI) are four to six times larger than in model (CC-LI), depending on the branch dummy considered (e.g., the marginal effect for the branch dummy $N = 1$ is 0.30 in model (CC-LI) and 1.86 in model (CC-FI)—not tabulated). To summarize, we find the coefficient estimates for utility shifters to be quantitatively different under the assumptions of full and limited information, which is consistent with the analysis of the different limited information models in Section 4.

□ **Does advertising influence more consumer awareness or choice?** To compare the magnitudes of the effects of advertising across the different stages in the consumer shopping process, we look at the advertising elasticities from the awareness and consideration-choice stages. Table 6 shows the results from this analysis. The average advertising elasticities for awareness and choice are 0.10 and 0.02, respectively. These elasticities indicate that advertising affects consumer awareness more than it affects choice conditional on awareness. This suggests that the role of advertising in the U.S. retail banking industry is, in accordance with the terminology used by previous research that describes the informative and persuasive roles of advertising (see Related literature section), primarily informative.

Our qualitative results are similar to those found by previous literature, albeit for different products. For example, Ackerberg (2001) and Ackerberg (2003) find that advertising has a primarily informative role in the yogurt market and Clark, Doraszelski, and Draganska (2009) also show that advertising has stronger informative effects in a study of over 300 brands. However, the results stand in contrast with other recent research that has also investigated consumers' demand for financial products. For example, Gurun, Matvos, and Seru (2016) and Hastings, Hortaçsu, and Syverson (forthcoming) suggest a negative (persuasive) effect of advertising for mortgages and retirement savings products, respectively.

Looking at the bank-specific advertising elasticities for awareness and choice, we find that the advertising elasticities for awareness vary from −0.01 to 0.25, whereas the advertising elasticities for choice conditional on awareness range from 0.00 to 0.06. For most banks, the advertising elasticity for awareness is larger than that for choice. The exceptions are Bank of America and Wells Fargo/Wachovia and, to a lesser extent, Capital One, for which the advertising elasticity for choice is larger than that for awareness. Again, as discussed before, for Bank of America and Wells Fargo/Wachovia, this could be because of the very high baseline awareness levels for these banks, possibly built based on past or current national advertising which would be captured (together with other factors that are constant across regions) by the bank fixed effects.

Although the comparison of the awareness and choice advertising elasticities is interesting per se, the effects being compared are measured in different types of outcome variables (namely, the probability of consumers being aware of a bank and the probability of consumers choosing a bank). An alternative way of answering the question of whether the effect of advertising is stronger via awareness or via choice is by analyzing how much advertising influences final choices when it is turned on and off in the different stages of the consumer shopping process. We conduct this exercise as part of the counterfactual analysis in Section 9 and find that the qualitative results from the comparison of advertising elasticities conducted in this section remain overall the same. This suggests that the effects of advertising on choice are marginal when compared to the effects of advertising on awareness.

[31] Intuitively, this can be explained as follows. In the data, consumers often choose their primary bank even when there are "better" options available. To match this data pattern, a model under the full information assumption (in which the consumer can choose from 18 banks), tends to overestimate the primary bank coefficient.

9. Counterfactual analyses

■ In this section, we perform counterfactual analyses to examine the implications of our empirical results for policy purposes and to further evaluate the economic impact of advertising on the consumer decision process and on certain economic variables of interest.

□ **Branch-advertising substitutability.** Banking consolidation is an ongoing feature of the U.S. banking industry. Merger guidelines issued by the Federal Trade Commission and the Department of Justice mostly focus on concentration metrics related to market shares, which are measured in terms of deposits and branch presence in carefully defined geographical markets. Although taking physical presence into account is definitely important, banks can use other strategic variables such as advertising to affect consumer awareness and consideration and therefore choice. The fact that advertising may affect which options consumers may be aware of is disregarded by U.S. antitrust authorities. This may lead to a mismeasurement of the real intensity of bank competition and to a misjudgment of the effectiveness of restrictions on competition. For example, if banks are present in a given geographic area but consumers are not aware of them, then the real intensity of competition is much lower than the intensity measured solely based on bank branch presence or share of deposits. At the same time, banks can substitute advertising for physical branch presence, thus mitigating the effectiveness of merger restrictions related to branch presence on bank competition.

Our empirical results indicate that bank branches have significant positive effects on both consumer awareness and choice. Here, we calculate how much banks would have to spend on advertising if they eliminated one branch and wanted to keep consumer choice unaffected by this.[32] This experiment allows us to further evaluate the economic impact of advertising and to quantify to what extent banks can use advertising to counteract restrictions on branch presence.[33]

Suppose that consumer awareness remains constant despite the closing of one branch—this might be interpreted as the short-term effect of branch closings. Our econometric specification allows the magnitude of this effect to depend on how many branches a bank has in a given area. According to the estimation results for model (CC-LI) reported in Table 7, if a bank closes the only branch within five miles of consumers, this bank would have to increase its advertising by about 31,680 units annually to keep its market share constant in a given DMA.[34] These 31,680 units annually translate into about \$633,600 if we use a rough advertising costs estimate of \$20 per unit advertised (calculated as an average across all types of media, DMAs, and banks). On the other hand, if a bank closes one of two or three existing branches, the advertising budget would have to increase by roughly \$94,261 (the average value of a marginal branch when a bank has two or three branches in an area) for the bank to maintain its market share. The marginal value of a branch for a bank that has between four and seven branches is smaller and estimated at about \$44,173. As expected, the marginal value of a branch decreases in the number of branches.

In the long run, branch closings also affect consumers' awareness of a bank. For example, if a bank closes the only branch within five miles of consumers, the bank would have to increase advertising by about 6726 units to keep awareness levels unchanged.[35] These 6726 advertising units roughly translate into \$134,526 in annual advertising spending. Note that this amount is smaller than the above \$633,600—the amount of additional annual advertising spending that

---

[32] Note that here, we investigate the effects of the elimination of a branch (i.e., divestment) and not the effects of a change in branch ownership.

[33] For example, the Department of Justice required 306 branches in four New England states to be closed in the Fleet Financial and BankBoston merger in 1999 (www.justice.gov/archive/atr/public/press_releases/1999/3027.htm).

[34] We compute the 31,680 annual figure using the estimates from model (CC-LI) in Table 7. The increase in advertising required to keep utility constant if the only branch within 5 miles of consumers closes is $0.5045 \div 0.0147 = 34,320$ advertising units (in 1000s) for a 13-month period (the length of the reference period).

[35] We calculate the 6726 figure by first dividing the branch coefficient estimates by the advertising coefficient estimates from model (A) in Table 5 for all banks with significant advertising coefficients and then averaging across all banks. Then, we convert the resulting number into an annual figure.

would be necessary to keep choice unchanged (conditional on constant awareness levels). Given that any advertising affects both awareness and choice and the amount necessary to uphold awareness levels is smaller than the amount necessary to uphold choice levels, the long-run value of the only branch of a bank within five miles of consumers would then be of $633,600 (per year).

The analysis in this section shows that, even if regulators force banks to divest from branches in a given geographic area, banks can compensate for those divestures by increasing their advertising spending in that area. This is because advertising is a strategic variable that firms can use to influence awareness and preferences for products. As the results reported here show, the increase in advertising expenditures that is required for a bank to compensate a potential branch closure is not prohibitively large, and hence, it is a feasible response for banks especially if the bank has more than one branch in that market. Although we are not advocating that it is feasible for regulators to put a limit on advertising budgets, it is important for them to be aware of the extent to which advertising can work as a substitute for branches and potentially consider reassessing the criteria for evaluating antitrust policy in the retail banking industry. Complementing the existing merger simulations analysis based on market concentration measures with an analysis of the impact of advertising on consumer awareness, along the lines proposed in this article, may lead to more accurate evaluations of the effects of merger restrictions on long-run outcomes.

□ **Advertising ban.** In this section, we perform a counterfactual analysis in which advertising is banned. This allows us to quantify how different are the effects of advertising through awareness and through consideration/choice with respect to several metrics, such as the size of consumers' awareness and consideration sets, the magnitude of incurred search costs, the prices of chosen products, and market concentration. Although this counterfactual does not reflect a particular or feasible proposed policy, it quantifies the full contribution of advertising along the different stages of the consumer shopping process and illustrates how disregarding (especially) the awareness stage may not give a complete picture of the effects of advertising.[36]

Some recent research has pointed out the pernicious effects of advertising in financial markets (e.g., Gurun, Matvos, and Seru, 2016; Hastings, Hortaçsu, and Syverson, forthcoming) and suggested that consumers are better off when advertising is banned. This recent research, however, has focused only on the choice stage. Thus, it has not considered the potential effects of advertising on consumer awareness and how this channel may ultimately affect consumer choices in a positive way (e.g., awareness of more options may allow consumers to choose a better product and it may also increase competition in the industry). In our case, given that the effects of advertising in the choice stage are marginal when compared to those in the awareness stage, the distinction between the two channels is particularly important. This distinction also helps us better understand the exact mechanism (awareness versus choice) through which advertising operates in the banking industry.

Recall that, according to our model, advertising can affect consumers' utility and final choices in two ways: indirectly through awareness (because being aware of more or fewer options affects consumers' awareness sets and consequently their consideration sets and choices) and directly through consideration and choice (because the fact that advertising is included as a utility shifter directly in the utility function makes some options more salient to the consumer and thus more likely to be part of consumers' consideration sets and to be chosen). Accordingly, in Table 8, we present the outcomes from three scenarios. Scenario 1 shows the model predictions given the observed (in the data) advertising levels. In Scenario 3, advertising is banned. Scenario 2, an intermediate scenario in which the effects of advertising are turned off for awareness but kept on for choice, allows us to measure the effects of advertising on choice directly via the utility

[36] Also note that our model is a partial equilibrium model. Thus, any counterfactuals capture only consequences on the demand side, that is, we do not model interest rates or advertising spending adjustments on the supply side. The results can be interpreted as short-run market effects.

TABLE 8 Counterfactual Analysis: Advertising Ban

	Scenario 1 Ad in Aw.: ON Ad in Ch.: ON	Scenario 2 Ad in Aw.: ON Ad in Ch.: OFF	Scenario 3 Ad in Aw.: OFF Ad in Ch.: OFF	% Change 1→ 2	% Change 2 → 3
Average awareness set size	6.33	6.33	5.34	0.00	−15.61
Average consideration set size	2.48	2.41	1.78	−2.85	−25.97
Average total search costs	0.096	0.093	0.069	−2.85	−25.97
Average consumer welfare	−0.65	−0.72	−1.00	−10.54	−38.70
Average interest rate of chosen bank	0.17	0.16	0.14	−9.29	−9.41
Market concentration—CR_4	53.27	53.93	59.30	1.23	9.96
Market concentration—HHI	1020.27	1024.60	1174.77	0.42	14.66

	Awareness Probabilities			Choice Probabilities/Market Shares				
Institution	Scenarios 1 and 2 Ad in Aw.: ON Ad in Ch.: ON/OFF	Scenario 3 Ad in Aw.: OFF Ad in Ch.: OFF	% Change 1, 2 → 3	Scenario 1 Ad in Aw.: ON Ad in Ch.: ON	Scenario 2 Ad in Aw.: ON Ad in Ch.: OFF	Scenario 3 Ad in Aw.: OFF Ad in Ch.: OFF	% Change 1→ 2	% Change 2 → 3
Bank of America	99.50	99.42	−0.08	17.09	17.27	19.53	1.06	13.06
BB&T	13.27	8.22	−38.03	1.80	1.80	1.36	0.21	−24.58
Citibank	67.39	60.82	−9.75	6.18	6.29	6.21	1.67	−1.28
Citizens Bank	18.55	7.59	−59.11	3.66	3.69	3.16	0.76	−14.31
Comerica	11.49	7.10	−38.24	0.82	0.83	0.66	1.50	−20.47
Fifth Third	25.50	15.54	−39.06	3.93	3.90	3.27	−0.70	−16.10
HSBC	12.19	5.27	−56.77	2.18	2.31	1.95	5.90	−15.27
Chase/WaMu	77.85	74.97	−3.70	11.51	11.75	13.11	2.06	11.57
Keybank	23.62	20.04	−15.13	2.34	2.34	2.31	0.01	−1.24
M&T	6.36	3.85	−39.49	2.09	2.11	2.00	1.03	−5.16
PNC/N. City Bank	32.40	23.96	−26.05	4.40	4.49	4.37	2.21	−2.72
Regions	23.81	19.56	−17.85	2.77	2.78	2.54	0.27	−8.68
Sovereign	13.07	7.21	−44.87	1.50	1.67	1.67	11.84	−0.29
SunTrust	26.19	20.09	−23.30	3.86	3.88	3.46	0.50	−10.72
TD	13.88	3.54	−74.49	1.88	2.02	1.09	7.10	−45.88
U.S. Bank	40.89	32.36	−20.87	8.22	8.28	7.92	0.68	−4.31
Wells Fargo/Wachovia	97.44	97.65	0.21	18.49	18.62	20.46	0.72	9.86
Capital One	29.65	27.06	−8.75	7.29	5.98	4.93	−18.02	−17.49

This table provides the results for the advertising ban counterfactual. Scenario 1 shows the model predictions given the observed advertising levels. In Scenario 3, advertising is banned. In Scenario 2, the effects of advertising on awareness are turned off while they are kept on for choice. Scenario 2 thus allows us to measure the effects of advertising on choice directly via the utility function (when compared to Scenario 1), and indirectly via consumers' awareness (when compared to Scenario 3). The top panel of the table shows how overall economic measures of interest change in each of the scenarios. The bottom panel shows in more detail what are the awareness and choice probabilities for the banks in each of the three scenarios. All averages are calculated across consumers taking the representativeness weights into account. Search costs represent the total search costs incurred per consumer (i.e., they are a function of the consideration set sizes) and are measured in interest rate percentage points. Consumer welfare is measured in utils. Awareness and Choice probabilities are calculated taking representativeness weights into account.

function (when compared to Scenario 1) and indirectly via consumer awareness (when compared to Scenario 3).[37]

The top panel in Table 8 (especially the last two columns) shows that, when advertising is turned off in the consideration/choice stage (Scenario 1 → Scenario 2), the economic quantities of interest barely change, especially when compared to the large effects of turning advertising off in the awareness stage (Scenario 2 → Scenario 3). Hence, in what follows, we focus most of our discussion on analyzing the results from turning advertising on/off in the awareness stage while keeping the direct effect of advertising on consideration/choice turned off (Scenario 2 → Scenario 3). This focus also provides a more clean experiment because the interpretation of the direct effect of advertising on utility is ambiguous in the literature, and different interpretations have different implications in terms of welfare analysis. On the one hand, the direct effect of advertising on utility can be literally interpreted as advertising providing utility to the consumer (e.g., through a positive informative role). In this case, shutting down advertising will lead to a loss of utility that we should take into account when computing changes in consumer welfare in the choice stage. On the other hand, an alternative interpretation is that we should not view advertising as a product characteristic that gives utility because its effect is manipulative and its only function is to distort preferences (Chamberlin, 1933). These two alternative interpretations imply that the estimated advertising coefficient in the utility function (even if positive) has ambiguous interpretations in terms of consumer welfare. As such, we perform our analysis here by shutting down this direct utility effect to provide a cleaner analysis of the advertising effects on certain economic variables of interest.

The top panel in Table 8 shows that, when we shut down the effect of advertising through awareness, the average size of consumers' awareness sets decreases from 6.33 to 5.34 (i.e., each consumer is aware of one fewer bank on average) which represents a decrease of 16%. To evaluate the effects on the awareness of each bank separately, the columns under "Awareness Probabilities" in the bottom panel of Table 8 show the proportion of consumers who are predicted to be aware of each bank in each of the three scenarios. (Turning advertising off in the choice stage has by definition no effect on awareness probabilities. Hence, the awareness probabilities under Scenarios 1 and 2 are the same.) Naturally, the proportion of consumers who are aware of a bank decreases when the effect of advertising on awareness is set to zero. More interestingly, the magnitude of the effects appears to be inversely related to bank size. For larger banks such as Bank of America, Citibank, Chase/WaMu, and Wells Fargo/Wachovia, the proportion of consumers who are aware of a bank decreases by 0% to 10%. For smaller banks, the proportion of consumers who are aware of a bank decreases by 9% to 74%.

Turning to the analysis of the effects of an advertising ban on other economic variables, the top panel in Table 8 shows that, on average, the size of consumers' consideration sets decreases from 2.41 to 1.78 banks (i.e., about two fewer banks are considered for every three consumers searching), which represents a decrease of 26%. Because consumers' consideration sets decrease in size, consumers incur 25% less search costs (measured in interest rate percentage points) and end up choosing deposit accounts with worse interest rates (although the difference in interest rates is very small, 0.16% per year in Scenario 3 versus 0.14% per year in Scenario 2). Consumer welfare also decreases significantly by 39% when we go from Scenario 2 to Scenario 3.[38,39] As consumers are aware of and consider fewer options when advertising is off, market concentration

[37] To obtain (counterfactual) predictions, we generate, for each consumer, 50 sets of draws from the error distributions for awareness and consideration/choice. For each set, we predict awareness, conditional consideration, and conditional choice and average predictions within each consumer. Last, we calculate weighted mean predictions across consumers using the representativeness weights.

[38] Note that the formula used to calculate consumer welfare (Small and Rosen, 1981) is heavily driven by the number of considered banks. Therefore, consumer welfare is likely to decrease as consumers search fewer banks.

[39] Welfare also decreases when advertising is turned off in choice (from Scenario 1 to Scenario 2). Although the direct effect of advertising on utility is hard to interpret (positive effects of advertising on utility have been interpreted as persuasive/negative), the positive effect of advertising on choice through awareness is unambiguous.

increases significantly. Overall, the Herfindahl-Hirschman Index (HHI) increases from 1025 to 1175 when the impact of advertising on awareness is zeroed out and only from 1020 to 1025 when the direct effect of advertising on preferences is turned off. The Four-Bank Concentration Ratio (CR_4) increases from 54% to 59% when the impact of advertising on awareness is zeroed out and only from 53% to 54% when the direct effect of advertising on preferences is turned off.

To better understand the changes in market concentration, we calculate how market shares change for each specific bank when advertising is turned on/off for awareness and choice. The results are shown in the columns under “Choice Probabilities/Market Shares” in the bottom panel of Table 8. The effects of turning off the direct effect of advertising on preferences are minimal for most banks when compared to the effects of turning off the effect of advertising on awareness (the exceptions are Sovereign and, to a lesser extent, Citibank and Capital One). When the effect of advertising on awareness is zeroed out, we can see that the impact appears to be asymmetric across banks with different sizes: three of the four largest banks (Bank of America, Chase/WaMu, and Wells Fargo/Wachovia) increase their market shares by over 10% on average, whereas smaller banks see their market shares decrease by 13% on average. In turn, this asymmetric impact on large and smaller banks leads to the significant increase in market concentration when advertising is banned.

To summarize, we find that the main mechanism through which advertising affects choices in this market is through awareness. As we show here, this channel is quantitatively important. Overall, advertising has a positive effect on economic outcomes and consumer welfare through increasing consumer awareness. Allowing for advertising makes consumers aware of more options and thus leads them to search more (consider more banks) and to ultimately make better choices. In turn, the increase in awareness (and consideration sets) leads to an increase in the market share of smaller banks, making the U.S. banking industry more competitive.

10. Robustness checks

■ We conduct a variety of checks to test the robustness of our results.[40] First, we change the radius for the local bank branch variable in the estimation. In the results shown, we control for local bank presence within five miles of the consumer’s zip code centroid. We also estimated our model using an alternative radius of 10 miles from the consumer’s zip code centroid for local bank presence. For both awareness and consideration/choice, we find very similar and mostly significant coefficients for the bank branch dummies. The parameter estimates are slightly larger using the 10-mile radius compared to the 5-mile radius. The estimates for the other variables remain very similar. Thus, we conclude that our results are robust to different definitions of local bank presence.

Second, we also verify the robustness of our results to an alternative definition of interest rates. The currently reported results are based on interest rate data calculated using information on the most common 2.5K savings account for each bank. We also experimented using data on all different types of 2.5K savings account for each bank and the qualitative results did not change.

Last, we check the robustness of our results with respect to a different measure of advertising. In our model, we operationalize advertising as the number of DMA-level advertisements. In this robustness check, we use the sum of national and DMA-level advertising quantity in the estimation. We find that our results are qualitatively and quantitatively robust to this alternative measure of advertising.

11. Conclusion

■ In this article, we utilize a unique data set with detailed information on the consumer’s shopping process for banking services. Using data on awareness and consideration sets and the purchase decision, we disentangle the effects of advertising on consumer awareness and choice

[40] The detailed results for all robustness checks are available from the authors upon request.

in the retail banking sector. We find that advertising primarily informs consumers about the existence of banks and only marginally shifts consumers' utility for retail banks. Our results shed light on the drivers of demand, the value of local bank presence, and the role of advertising in this very important sector of the economy. Advertising makes consumers aware of more options; thus, consumers search more and find better alternatives than they would otherwise. In turn, this increases the market share of smaller banks, making the U.S. banking industry more competitive.

Modifying some of our modeling assumptions could potentially lead to interesting extensions. For example, our model describes the consumer's shopping and account opening process, given his decision on the account types he is considering adding or moving, that is, we do not jointly model the consumer's choice of account types and the search among banks. Developing a model where consumers choose several products and search at the same time that they evaluate those products could lead to interesting predictions. Also, more work is needed to enhance our understanding of the effectiveness of price promotions versus advertising in the retail banking industry. Advertisements stating, for example, that consumers can get $200 for opening a new checking account, as advertised by Chase, are effectively price promotions and their effectiveness as compared to brand advertising is an open question for future research.

References

Abhishek, V., Fader, P., and Hosanagar, K. "Media Exposure through the Funnel: A Model of Multi-Channel Attribution." Working Paper no. 8–2012, Heinz College, Carnegie Mellon University, 2012.

Ackerberg, D.A. "Empirically Distinguishing Informative and Prestige Effects of Advertising." *RAND Journal of Economics*, Vol. 32 (2001), pp. 316–333.

———. "Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination." *International Economic Review*, Vol. 44 (2003), pp. 1007–1040.

Allen, J., Clark, R., and Houde, J.F. "Search Frictions and Market Power in Negotiated Price Markets." Working Paper no. 19883, National Bureau of Economic Research, 2014.

Allenby, G.M. and Ginter, J.L. "The Effects of In-Store Displays and Feature Advertising on Consideration Sets." *International Journal of Research in Marketing*, Vol. 12 (1995), pp. 67–80.

Bagwell, K. "The Economic Analysis of Advertising." In M. Armstrong and R.H. Porter, eds., *The Handbook of Industrial Organization*. New York: North Holland, 2007.

Belo, F., Lin, X., and Vitorino, M.A. "Brand Capital and Firm Value." *Rev. Econom. Dynam.*, Vol. 17 (2014), pp. 150–169.

Blundell, R.W. and Powell, J.L. "Endogeneity in Semiparametric Binary Response Models." *Review of Economic Studies*, Vol. 71 (2004), pp. 655–679.

Bronnenberg, B.J., Kim, J.B., and Mela, C.F. "Zooming in on Choice: How Do Consumers Search for Cameras Online?" *Marketing Science*, Vol. 35 (2016), pp. 693–712.

Chade, H. and Smith, L. "Simultaneous Search." Working Paper no. 1556, Department of Economics, Yale University, 2005.

Chamberlin, E.H. *The Theory of Monopolistic Competition: A Re-orientation of the Theory of Value*. Cambridge, Mass.: Harvard University Press, 1933.

Chan, T., Narasimhan, C., and Xie, Y. "Treatment Effectiveness and Side Effects: A Model of Physician Learning." *Management Science*, Vol. 59 (2013), pp. 1309–1325.

Chiang, J., Chib, S., and Narasimhan, C. "Markov Chain Monte Carlo and Models of Consideration Set and Parameter Heterogeneity." *Journal of Econometrics*, Vol. 89 (1998), pp. 223–248.

Ching, A.T. and Ishihara, M. "Measuring the Informative and Persuasive Roles of Detailing on Prescribing Decisions." *Management Science*, Vol. 58 (2012), pp. 1374–1387.

Chintagunta, P.K. and Lee, J. "A Pre-Diffusion Growth Model of Intentions and Purchase." *Journal of the Academy of Marketing Science*, Vol. 40 (2012), pp. 137–154.

Clark, C., Doraszelski, U., and Draganska, M. "The Effect of Advertising on Brand Awareness and Perceived Quality: An Empirical Investigation Using Panel Data." *Quantitative Marketing and Economics*, Vol. 7 (2009), pp. 207–236.

Cohen, M. and Rysman, M. "Payment Choice with Consumer Panel Data." Working Paper no. 13-6, Federal Reserve Bank of Boston, 2013.

De los Santos, B.I., Hortaçsu, A., and Wildenbeest, M.R. "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior." *American Economic Review*, Vol. 102 (2012), pp. 2955–2980.

Dick, A.A. "Market Size, Service Quality, and Competition in Banking." *Journal of Money, Credit and Banking*, Vol. 39 (2007), pp. 49–81.

———. "Demand Estimation and Consumer Welfare in the Banking Industry." *Journal of Banking & Finance*, Vol. 32 (2008), pp. 1661–1676.

FRANSES, P.H. AND VRIENS, M. "Advertising Effects on Awareness, Consideration and Brand Choice Using Tracking Data." Working Paper no. ERS-2004-028-MKT, Department of Econometrics, Erasmus University Rotterdam, 2004.

GOEREE, M.S. "Limited Information and Advertising in the U.S. Personal Computer Industry." *Econometrica*, Vol. 76 (2008), pp. 1017–1074.

GURUN, U.G., MATVOS, G., AND SERU, A. "Advertising Expensive Mortgages." *Journal of Finance*, Vol. 71 (2016), pp. 2371–2416.

HASTINGS, J.S., HORTAÇSU, A., AND SYVERSON, C. "Sales Force and Competition in Privatized Social Security: The Case of Mexico." *Econometrica* (forthcoming).

HAUSMAN, J.A. "Specification Tests in Econometrics." *Econometrica*, Vol. 46 (1978), pp. 1251–1271.

HECKMAN, J.J. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica*, Vol. 46 (1978), pp. 931–959.

HIRTLE, B. "The Impact of Network Size on Bank Branch Performance." *Journal of Banking & Finance*, Vol. 31 (2007), pp. 3782–3805.

HONG, H. AND SHUM, M. "Using Price Distributions to Estimate Search Costs." *RAND Journal of Economics*, Vol. 37 (2006), pp. 257–275.

HONKA, E. "Quantifying Search and Switching Costs in the U.S. Auto Insurance Industry." *RAND Journal of Economics*, Vol. 45 (2014), pp. 847–884.

HONKA, E. AND CHINTAGUNTA, P.K. "Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry." *Marketing Science*, Vol. 36 (2017), pp. 21–42.

HORTAÇSU, A. AND SYVERSON, C. "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds." *The Quarterly Journal of Economics*, Vol. 119 (2004), pp. 403–456.

ISHII, J. "Compatibility, Competition, and Investment in Network Industries: ATM Networks in the Banking Industry." Working Paper, Stanford Graduate School of Business, Stanford University, 2008.

KIM, J.B., ALBUQUERQUE, P., AND BRONNENBERG, B.J. "Online Demand under Limited Consumer Search." *Marketing Science*, Vol. 29 (2010), pp. 1001–1023.

KISER, E.K. "Predicting Household Switching Behavior and Switching Costs at Depository Institutions." *Review of Network Economics*, Vol. 20 (2002), pp. 349–365.

KOULAYEV, S., RYSMAN, M., SCHUH, S., AND STAVINS, J. "Explaining Adoption and Use of Payment Instruments by U.S. Consumers." *RAND Journal of Economics*, Vol. 47 (2016), pp. 293–325.

LAMBRECHT, A., SEIM, K., AND TUCKER, C. "Stuck in the Adoption Funnel: The Effect of Interruptions in the Adoption Process on Usage." *Marketing Science*, Vol. 30 (2011), pp. 355–367.

LOVETT, M.J. AND STAELIN, R. "The Role of Paid and Earned Media in Building an Entertainment Brand: Reminding, Informing, and Enhancing Enjoyment." *Marketing Science*, Vol. 35 (2016), pp. 142–157.

MCFADDEN, D. "The Choice Theory Approach to Market Research." *Marketing Science*, Vol. 5 (1986), pp. 275–297.

MEHTA, N., RAJIV, S., AND SRINIVASAN, K. "Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation." *Marketing Science*, Vol. 22 (2003), pp. 58–84.

MÓLNAR, J., VIOLI, R., AND ZHOU, X. "Multimarket Contact in Italian Retail Banking: Competition and Welfare." *International Journal of Industrial Organization*, Vol. 31 (2013), pp. 368–381.

MORAGA-GONZÁLEZ, J.L., SÁNDOR, Z., AND WILDENBEEST, M.R. "Consumer Search and Prices in the Automobile Market." Discussion Paper no. TI 2015–033/VII, Tinbergen Institute, 2016.

MORAGA-GONZÁLEZ, J.L. and WILDENBEEST, M.R. "Maximum Likelihood Estimation of Search Costs." *European Economic Review*, Vol. 52 (2008), pp. 820–848.

NARAYANAN, S., MANCHANDA, P., AND CHINTAGUNTA, P.K. "Temporal Differences in the Role of Marketing Communication in New Product Categories." *Journal of Marketing Research*, Vol. 42 (2005), pp. 278–290.

PETRIN, A. AND TRAIN, K. "A Control Function Approach to Endogeneity in Consumer Choice Models." *Journal of Marketing Research*, Vol. 47 (2010), pp. 3–13.

RIVERS, D. AND VUONG, Q.H. "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models." *Journal of Econometrics*, Vol. 39 (1988), pp. 347–366.

RYSMAN, M. "An Empirical Analysis of Payment Card Usage." *The Journal of Industrial Economics*, Vol. 55 (2007), pp. 1–36.

RYSMAN, M. AND WRIGHT, J. "The Economics of Payment Cards." *Review of Industrial Organization*, Vol. 13 (2014), pp. 303–353.

SIDDARTH, S., BUCKLIN, R.E., AND MORRISON, D.G. "Making the Cut: Modeling and Analyzing Choice Set Restriction in Scanner Panel Data." *Journal of Marketing Research*, Vol. 32 (1995), pp. 255–266.

SMALL, K.A. AND ROSEN, H.S. "Applied Welfare Economics with Discrete Choice Models." *Econometrica*, Vol. 49 (1981), pp. 105–130.

TERUI, N., BAN, M., AND ALLENBY, G.M. "The Effect of Media Advertising on Brand Consideration and Choice." *Marketing Science*, Vol. 30 (2011), pp. 74–91.

VAN NIEROP, E., BRONNENBERG, B.J., PAAP, R., WEDEL, M., AND FRANSES, P.H. "Retrieving Unobserved Consideration Sets from Household Panel Data." *Journal of Marketing Research*, Vol. 47 (2010), pp. 63–74.

Villas-Boas, J.M. "A Note on Limited versus Full Information Estimation in Non-Linear Models." Working Paper, Haas School of Business, University of California Berkeley, 2007.
Vitorino, M.A. "Understanding the Effect of Advertising on Stock Returns and Firm Value: Theory and Evidence from a Structural Model." *Management Science*, Vol. 60 (2014), pp. 227–245.
Wang, H. and Ching, A.T. "Consumer Valuation of Network Convenience: Evidence from the Banking Industry." Working Paper, Guanghua School of Management, Peking University, 2016.
Winer, R.S. and Dhar, R. *Marketing Management*. Upper Saddle River, NJ: Prentice Hall, 2011.
Yao, S., Wang, W., and Chen, Y. "TV Channel Search and Commercial Breaks." *Journal of Marketing Research*, (forthcoming).
Zhang, J. "An Integrated Choice Model Incorporating Alternative Mechanisms for Consumers' Reactions to In-Store Display and Feature Advertising." *Marketing Science*, Vol. 25 (2006), pp. 278–290.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's website:

Table A.1: Demographics for Initial and Final Samples
Table A.2: DMA-level Advertising Expenditures and Placements by Bank
Table A.3: Respondents with Bank Branches within 5 Miles of their Home (Shoppers only)
Table D.1: Results from Awareness Stage (Pre and Post Control Function)
Table D.2: Results from Consideration and Choice Stages (Pre and Post Control Function)