



**AgEcon** SEARCH  
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

# Attribute Search in Online Retailing

Timothy J. Richards, Arizona State University, trichards@asu.edu  
Stephen F. Hamilton, shamilto@calpoly.edu  
Janine Empen, janine.empen@gmail.com\*

Selected Paper prepared for presentation at the  
Agricultural and Applied Economics Association's 2015 Annual Meeting,  
San Francisco, CA, July 26-28, 2015.

May 4, 2015

## Abstract

Online shopping is common in many categories of retail goods. The recent trend towards online retailing has created an unprecedented empirical opportunity to examine consumer search behavior using click stream data. In this paper we examine consumer search intensity across a wide range of grocery products that differ in the depth of product assortment. We develop a model of attribute search in which consumers search within a chosen retailer for products that match their tastes, and that equilibrium prices reflect retailers' expectations of how intensively consumers intend to shop. The model predicts an inverse relationship between product variety and attribute search in which greater product variety reduces search intensity and leads to higher retail prices. We test these hypotheses using consumer data on online search and purchase behavior from the comScore Web Behavior Panel. Our results indicate that consumer's search less and pay higher retail prices in categories with deeper product assortments, a finding that suggests deeper product assortments can produce anti-competitive effects in retail food markets mediated through equilibrium responses in consumer search.

*Keywords:* consumer search, variety, retail prices, attribute search, market power.

*JEL Classification:* D12, D83, L13, L81.

---

\*Copyright 2015 by Timothy J. Richards, Stephen F. Hamilton, and Janine Empen. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

# 1 Introduction

Online markets represent a small but rapidly growing share of retail food trade. In 2010, online grocery sales accounted for \$13.0 billion of the retail food business (Hartman 2012), an amount that is expected to grow to over \$100.0 billion and roughly 12 percent of total grocery spending by 2019 (Cloud 2014).<sup>1</sup> Moreover, with Amazon and Walmart beginning to compete in the “order and deliver” market, many of the obstacles to growth (access, delivery, and price) are likely to be removed from the online grocery segment over the next few years.<sup>2</sup> In the U.K., where generally lower incomes, smaller baskets, and more frequent shopping trips dampen online grocery sales, online food retailing accounts for fully 6% of all retail food sales, and internet grocers are expected to double warehouse space in 2014 (Butler 2014). Given the growing market for online grocery sales, it is important to understand the implications of consumers’ online search behavior for retail assortments and prices.

In this paper we utilize clickstream data in online grocery markets to examine how consumer search for product attributes affects retail pricing among differentiated products. Attribute search represents an under-studied aspect of consumer search behavior. However, online food retailing presents a unique empirical opportunity to examine how consumers search for attributes because we can directly observe consumer search behavior in the online interface. Our data describes how consumers search within different product categories at a single retailer, which allows us to isolate the “within-retailer” effect of search on equilibrium prices (on the intra-retailer margin), while abstracting from any inter-retailer, or competitive, effects on prices. To do so, we exploit variation in online product assortments and prices across multiple grocery categories within a single retailer to examine the linkage between attribute provision, consumer search intensity, and retail prices.

While many researchers document significant variation in prices for identical products

---

<sup>1</sup>Traditional retail outlets account for roughly \$500.0 billion in sales, and non-conventional outlets, such as club stores, discounters and drugstores, the remaining \$500.0 million (Hartman 2012).

<sup>2</sup>The online grocery market consists of “order and deliver,” “order online and store pick-up,” “specialty online,” and “online order and shipment” sub-segments (Hartman 2012).

offered by online retailers (Brynjolfsson and Smith 2000; Clay, Krishnan, and Wolff 2001), it is not clear whether these differences in prices are due to search imperfections, differentiation among the retailers themselves, or in the allocation of consumer search effort over price or attribute search. For example, a consumer buying a book online may search among multiple retailers for the best price for a particular book, or may browse a given retailer to find an interesting book, so that observations of consumer search behavior may confound price and attribute search. In this regard, our study of consumer search behavior across multiple online grocery categories represents a relatively clean examination of how product proliferation affects consumer attribute search and retail pricing at the individual category level.

In general, online and offline retailing differ in at least two important ways. First, consumer search costs are substantially lower online than offline, which conventional wisdom suggests should lead to increased price competition among brands, raising demand elasticities and lowering equilibrium prices. However, an emerging literature on online search behavior provides evidence that online demand elasticities for grocery items tend to be lower than their offline counterparts (Degeratu, Wangaswamy, and Wu 2000; Andrews and Currim 2004; Chu, Chintagunta, and Cebollada 2008). While there are a number of plausible explanations for this seemingly paradoxical effect, for instance online consumers may tend to be more time constrained, have higher incomes, or have higher brand loyalty than offline consumers (Danaher, Wilson and Davis 2003), online retailing is also likely to reduce the cost of attribute search. The ability to access a greater volume of information on the attributes of goods sold online can lead to a second difference between online and offline retailing – the “long tail” effect (Anderson 2006) that arises from a flattening of the sales distribution across the product assortment in each category. Brynjolfsson, Hu, and Simester (2011) find that, after controlling for supply factors, lower search costs online imply much deeper *effective* assortments, which is the segment of the product assortment that shoppers actually sift through to find the specific product attributes they desire. The ability to shop more efficiently for desired attributes can sharpen product differentiation, reducing the elas-

ticity of demand for foods sold online, and raising retail margins (Alba, et al. 1997; Chen and Hitt 2003; Kuksov 2004; Cachon, Terwiesch, and Yu 2008). Despite the theoretical research supporting this argument, there is little empirical evidence to date that provides clear separation between the effects of changes in price and attribute search on retail prices.

In this study, we examine the relationship between price and attribute search in online food markets. We frame our analysis by decomposing search costs between attribute search, which occurs on the intra-retailer margin within particular product assortment, and price search, which result in demand effects that are internalized between products on the intra-retailer margin but that affect equilibrium prices between rivals on the inter-retailer margin. On the inter-retailer margin of the grocery market, price comparison across retailers tends to occur at the basket-level rather than at the individual product level, resulting in a non-monotonic relationship between prices and the depth of retail product assortments (Hamilton and Richards 2009). Our decomposition of consumer attribute search on the intra-retailer margin allows us to exploit variation in the depth of product assortments across online grocery categories to examine how consumer search responds to the depth of retail product assortments.

In principle, providing a deeper product assortment can have two possible effects on attribute search. First, consumers faced with a larger attribute set to sort through may respond to increased product variety by raising search intensity, which would tend to make demand more price elastic as consumers carefully consider all aspects of the products they search (Kuksov 2004; Cachon, Terweisch and Yu 2008). Second, consumers may respond to deeper product assortments by reducing their search intensity. Longer product lines tend to fill the attribute space more densely with products, making it easier for consumers to find products with desirable attribute compositions, and this can facilitate superior matches between attributes and tastes for a given amount of search.

We consider a Butters (1977)-type process for search in attribute space in which consumers select the level of effort to spend searching the attribute space for the most desired

brand. Search behavior is conditional on the depth of the product assortment offered by the retailer whose product line is searched. The model results in the clear prediction that attribute search decreases in the depth of the retail product assortment. For example, if a consumer is searching the attribute space for wheat-based breakfast cereal with low sugar and dried-fruit, attribute search intensity may decrease if the cereal category is sufficiently deep that she can find cereal options with low sugar and dried fruit without having to weigh the relative merits of exchanging sugar for dried fruit, or substituting rice-based cereal for the preferred wheat-based alternative to acquire a desired mix of sugar and fruit.

The relationship between search and prices on the intra-retailer margin is less straightforward. If consumers search only for prices, as in Tappata (2009), Hong and Shum (2006), or Wildenbeest (2011), then the orthodox result arises that lower search intensity causes demand elasticities to fall, and equilibrium prices to rise as retailers exploit the fact that customers are not actively shopping for better deals. If consumers search for attributes, however, as they most surely do online (Anderson 2006), then the negative relationship between search and prices arises through a different mechanism. Namely, when searching for attributes consumers become more particular as to the exact specifications of the product they want, and become less sensitive to prices as a result. Facilitating attribute search allows consumers to indulge in niche, or “long tail” items, and equilibrium prices rise for highly differentiated items.

We test the implications of our model of attribute search using a data set describing online search behavior of visitors to Safeway.com. We exploit variation in the number of brands offered in different categories online to examine attribute search in response to changes in assortment depth in a setting that allow us to isolate the intra-retailer margin of consumer search. Our clickstream data encompasses consumer-level observations of items purchased, amounts paid, and several alternative measures of search intensity. Two of these measures – the duration spent on the website, and the number of pages viewed – form very good proxy

measures for search intensity.<sup>3</sup> With these measures, we are able to test the relationship between product variety and search intensity, and then examine the implications of search intensity for retail prices in a reduced-form search model. Once consumers determine how intensively they will search for attributes on the intra-retailer margin, observed prices reflect an equilibrium between shoppers' willingness to pay for specific items, and retailers' expectations to be rewarded for making the search process easier. Because search and prices are determined together, we estimate a simultaneous model of search and price-determination using GMM, instrumenting for endogenous search intensity.

Controlling for unobserved heterogeneity in our panel-data allows us to control for many of the confounding factors between product variety provision, search intensity, and equilibrium prices. Our empirical findings strongly support our main hypotheses, namely that search intensity declines with the depth of the product assortment, while equilibrium retail prices rise with assortment depth.

Much of the existing empirical literature focuses on either the relationship between product variety and search (Iyengar and Lepper 2000; Diehl and Poynor 2010; Kuksov and Villas-Boas 2010) or between search and prices (Mehta, Rajiv, Srinivasan 2003). Here, we examine the two together by empirically examining the relationship between variety and equilibrium prices, as moderated by consumer's optimal search behavior. Our findings paint an entirely different picture for food retailing than that expected by more casual empiricism. That is, because the conventional wisdom assumes prices fall with search costs, and search costs are lower online than offline, retail food prices are expected to fall as online grocery shopping becomes the norm. However, we show that the opposite is more likely to happen – consumers will not have to search as intensively for food items that they prefer, retail price elasticities will fall, and prices will rise. Retailers use the online channel to differentiate more

---

<sup>3</sup>de Los Santos, Hortacsu, and Wildenbeest (2012) employ similar data from comScore, Inc. to test the appropriateness of sequential relative to a non-sequential model of search in a multi-retailer model of book shopping. They conclude that the data are more consistent with a non-sequential model of search. While our data is similar to theirs, our objective is rather to study intra-retailer search among several categories, and not inter-retailer search for the same item.

efficiently, and raise margins by selling to more discriminating consumers.

In the next section, we derive a simple theoretical model of variety, search, and retail prices in which consumers search for attributes they prefer in an attribute space inhabited by differentiated food products. We summarize our data and describe our empirical strategy in more detail in a third section, while we present and interpret our econometric results in a fourth. In the fifth section, we draw some of the more important implications of our findings for the performance of the retail food sector, and for food prices more generally, and suggest avenues for future research in this area.

## 2 Economic Model of Consumer Search

In this section, we develop a simple conceptual model of search, and then use this model to derive testable hypotheses regarding how the net benefits of search are expected to change as retailers increase variety. The model is framed by a three stage game in which retailers select product variety and position their products in attribute space in stage 1, retailers select prices for products in the category in stage 2, and consumers select between retailers based on product variety and prices and search over attributes at the chosen retailer  $j$  in stage 3.

We consider a Butters (1977)-type process for consumer search in attribute space. Consumers have no ability to target search towards particular product segments and choose a level of “search intensity”, which represents the level of effort spent searching for the most desired brand. Following Bakos (1997) and Innes and Hamilton (2014), we assume consumer’s attribute preferences are uniformly distributed around the circumference of a unit circle. Each consumer has unit demand for the product that matches most closely with her preferred product attributes, and retailers array their products in attribute space to facilitate matches between consumers and brands.

Consider, first, a full-information model of consumer purchasing behavior in which consumers have full information on the product attributes and prices available at all retail



locations. Consumers selecting among products in a given retail store incur a “matching cost” of  $\delta$  per unit of distance in attribute space between the location of their most preferred product and the location of the nearest available brand. Retailers are able to reduce the distance between each consumer and her most desired product by increasing the depth of the product assortment ( $n_s$ ), where  $s = 1, 2, \dots, S$  is an index of retail store. That is,  $n_s$  is the number of products on the intraretailer margin. Given an ex ante uniform distribution of consumer preferences for product attributes, retailers optimally locate products symmetrically on the circle. Thus, competition between retailers in product assortments is thus isomorphic with competition between retailers in attribute space.

Measuring product variety continuously on the unit circle, the expected consumer matching cost when shopping at retailer  $s$  with assortment depth  $n_s$  is

$$2n_s \int_0^{1/2n_s} \delta x dx = \frac{\delta}{4n_s}.$$

Suppose prices are constant and equal across products in the category, for instance different flavors of yogurt at a supermarket.<sup>4</sup> Letting  $v$  represent the gross value each consumer receives from consuming her most preferred set of product attributes, expected utility from shopping at retailer  $s$  offering assortment depth  $n_s$  and prices  $p_s$  is:

$$u_s(p_s, n_s) = v - p_s - \frac{\delta}{4n_s}.$$

On the interretailer margin, retailers compete in prices ( $p_s$ ) and product variety ( $n_s$ ) subject to consumer transportation costs for traveling between retailers. Before turning to this stage, we next consider the role of attribute search in shaping retailer price and variety decisions.

Let  $\phi_i$  denote the search intensity of consumer  $i$  in attribute space, which we interpret as the probability that consumer  $i$  discovers a given product on the retailer’s shelf. It follows that consumer  $i$  searches the attribute space with intensity level  $\phi_i$  to find her most desired

---

<sup>4</sup>The assumption of constant prices across variants in the category clarifies the role of attribute search within a product category. Consumer search for lower prices between brands on the intraretailer margin has no strategic implications for consumer prices, because retailers act like monopolists on the intraretailer margin.

brand with probability  $\phi_i$ . Similarly, the probability that the consumer fails to find her most desired brand as the outcome of her search, but discovers her second-favorite brand is given by  $\phi_i(1 - \phi_i)$  and the probability that she fails to find any of her  $k$  most desired brands and finds her  $(k + 1)^{th}$  favorite brand is given by  $\phi_i(1 - \phi_i)^k$ . Taking into account the location of products in attribute space, expected utility (gross of search cost) for consumer  $i$  who chooses search intensity  $\phi_i$  is

$$E(u(p, n, \phi_i)) = \phi_i(v - p) \sum_{k=1}^n (1 - \phi_i)^{k-1} - \phi_i \frac{\delta}{4n} \sum_{k=1}^n (2k - 1)(1 - \phi_i)^{k-1}.$$

Summing terms yields

$$E(u(p, n, \phi_i)) = (1 - (1 - \phi_i)^n) \left[ (v - p) - \frac{\delta}{4n\phi_i}(2 - \phi_i) \right] + \frac{\delta}{2}(1 - \phi_i)^n. \quad (1)$$

For reasonable parameter values such that  $(1 - \phi_i)^n$  is small, a good approximation to expected utility in equation (1) is:

$$E(u(p, n, \phi_i)) = (v - p) - \frac{\delta(2 - \phi_i)}{4n\phi_i}. \quad (2)$$

Next, let  $c(\phi_i)$  denote the costs of searching for attributes, where  $c_\phi > 0$ , and  $c_{\phi\phi} \geq 0$ . Search costs rise in the intensity of search as the probability of finding a suitable match requires greater cognitive effort for larger consideration sets. Upon entering a store, consumers facing posted prices of  $p_j$  and product variety  $n_j$  at retailer  $j$  select a search intensity to maximize expected utility in (2), net of the cost of search,

$$E(u(p_j, n_j, \phi_i)) - c(\phi_i) = v - p_j - \frac{\delta(2 - \phi_i)}{4n_j\phi_i} - c(\phi_i).$$

The optimal search intensity is then defined as the solution to:

$$\frac{\delta}{2n_j\phi_i^2} = c_\phi(\phi_i). \quad (3)$$

In settings with hierarchical search, it follows by implicit differentiation of (3) that:

$$\frac{\partial \phi_i}{\partial n_j} = \frac{-\delta}{2n_j^2\phi_i(2c_\phi + \phi_i c_{\phi\phi})} < 0. \quad (4)$$

The intensity of consumer attribute search decreases with product variety in the market equilibrium. The reason is that consumers are more likely to find a product that meets their needs with greater variety to choose from, decreasing the return to search. Put differently, greater product variety reduces the attribute distance between neighboring products in the category, so that searching to improve the match from the  $n$ th-preferred to the  $(n-1)$ th-preferred brand has a smaller consequence for utility.

Notice that attribute search does not depend on retail prices on the intraretailer margin. This is because consumers must search the attribute space to find their most desired product irrespective of the price level in the category. The process of attribute search on the intraretailer margin results in a particular match at that retailer that can subsequently be compared with the expected product match and prices available at rival retailers on the interretailer margin. Consumers prefer to shop with retailers that offer low prices and deep product assortments, but are unaware of the particular product they will find available at the chosen retailer until they enter the store.

Next consider retailer behavior on the interretailer margin of competition. For analytic convenience, suppose consumers are uniformly distributed on a Hotelling line between two competing retailers and face unit transportation costs of  $t$ . Letting  $\phi(n_s)$  denote the consumer search-response to product variety level  $n_s$  by a representative consumer, consumer utility can be written as

$$v_s(p_s, n_s) \equiv E(u(p_s, n_s, \phi(n_s))) - c(\phi(n_s)).$$

Given the menu of expected product variety and prices available at retailer  $s$ ,  $s = 1, 2$ , the location of the critical consumer on the unit line between retailers is

$$\theta^*(p_1, n_1; p_2, n_2) = \frac{t + v_1(p_1, n_1) - v_2(p_2, n_2)}{2t}$$

Profit for retailer 1 is given by

$$\pi_1 = \theta^*(p_1, n_1; p_2, n_2)(p_1 - c_1) - fn_1.$$

The first-order condition in the pricing stage is given by

$$\frac{\partial \pi_1}{\partial p_1} = \theta^*(.) - \frac{1}{2t}(p_1 - c_1) = 0,$$

and similarly for retailer 2. In the symmetric equilibrium with fixed consumer territories,  $\theta^* = 1/2$ , the model yields conventional Hotelling prices,  $p - c = t$ ; however, retail margins in the present model depend on variety provision from the prior stage of the game. Noting that  $\theta^*(p_1, n_1; p_2, n_2)$  depends on prior outcomes for the product assortment selected by both retailers, retail prices in the final stage of the game respond to retail product lines according to

$$\begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} \partial p_1 \\ \partial p_2 \end{bmatrix} = \frac{\delta}{4n_1^2} \left( \frac{2 - \phi_1}{\phi_1} \right) \begin{bmatrix} -1 \\ 1 \end{bmatrix} \partial n_1,$$

which yields

$$\frac{\partial p_1}{\partial n_1} = \frac{\delta}{12n_1^2} \left( \frac{2 - \phi_1}{\phi_1} \right) = \frac{-\partial p_2}{\partial n_1} > 0.$$

Now consider the variety choice stage of the game. The optimal variety choice of retailer 1 satisfies

$$\frac{\partial \pi_1}{\partial n_1} = \frac{1}{2t} \left( \frac{\partial v_1}{\partial n_1} + \frac{\partial v_1}{\partial p_1} \frac{\partial p_1}{\partial n_1} - \frac{\partial v_2}{\partial p_2} \frac{\partial p_2}{\partial n_1} \right) - f = 0.$$

Substituting terms from above and imposing symmetry ( $p_1 = p_2 = p; n_1 = n_2 = n$ ) yields

$$\frac{\delta(p - c)}{24tn^2} \left( \frac{2 - \phi}{\phi} \right) = f. \quad (5)$$

The price implications following a search-response to a change in variety are given by implicit derivation of (5) with  $\phi = \phi(n)$  given by equation (3). Considering a multilateral increase in product variety at both retailers in the symmetric market equilibrium, the price effect of search is given by

$$\frac{\partial p}{\partial n} = \frac{2(p - c)}{\phi(2 - \phi)n} \left[ \phi(2 - \phi) + n \left( \frac{\partial \phi}{\partial n} \right) \right],$$

where the consumer search response to greater variety is given by equation (4). Given that consumer attribute search intensity decreases with the depth of product assortments, equilibrium prices can rise or fall on the interretailer margin depending on the magnitude of the effect of product variety on attribute search.

In the case of linear search costs over attributes,  $c_{\phi\phi} = 0$ , prices rise with assortment depth. Specifically,

$$\frac{\partial p}{\partial n} = \frac{(p - c)(3 - 2\phi)}{(2 - \phi)n} > 0.$$

The mechanism underlying this effect is straightforward. Greater variety causes consumers to search less as consumers can find acceptable matches more easily. Retailers thus are able to forestall attribute search by providing deeper product assortments, enhancing pricing power on the interretailer margin. In the empirical model below, we test the prediction of the theory that product variety reduces search intensity using a reduced-form model of search behavior and price response.

### 3 Empirical Test of Variety and Search Intensity

#### 3.1 Overview

Empirical tests of search theory are relatively rare, because search behavior in traditional retail channels is difficult to observe. Recently, however, researchers have begun to exploit web technology that enables direct observation of search behavior online (Koulayev 2010; Honka 2010; Honka and Chintagunta 2013). For our empirical test, we use data describing grocery purchases made by online shoppers at a major supermarket website. By directly observing search behavior, we are able to conduct empirical tests of our theory more effectively than the "structural" literature in which search behavior is only inferred (Mehta, Rajiv, and Srinivasan 2003).

#### 3.2 Online Search Data

Our data is drawn from the comScore, Inc. Web Behavior Panel (WBP).<sup>5</sup> comScore is the most prominent source of online purchase data in the US, and their WBP consists of over 50,000 households, tracked over a 3-year period. For our purposes, we were only interested in panelists' grocery shopping behavior, so our data focus only on visits to the only retailer with

---

<sup>5</sup>de Los Santos, Hortacsu, and Wildenbeest (2012) also use WBP data, but from a different set of retailers, and categories.

a national presence in online / local delivery: Safeway.com. The total number of households that reported more than 10 visits to Safeway.com was 36. WBP subjects agree to allow comScore install software on their primary home computer that tracks each keystroke when they are online. By following this "clickstream" data, comScore analysts know the websites visited on each session on the computer, the pages visited in each domain, the duration spent on each page, and whether a page visit results in a purchase, or additional search. From an economic perspective, WBP data allows direct measurement of each shopper's consideration set, how seriously each element of the consideration set was vetted, and which element was ultimately chosen. Our sample of WBP data, therefore, comprises some 5,267 clickstream observations and 148 completed shopping baskets from 389 different product categories.

To test the hypothesis that follows from the theoretical model above, we require a measure of search intensity, and a measure of assortment depth. Assortment depth (VAR) is calculated as the sum of all UPCs (universal product code, or unique items) purchased in each product category across all panel members in a particular week. That is, if panelists purchased a total of 120 unique cereal products in the fourth week of the sample period, then we treat  $n$  as 120. We derive three measures of search intensity from the WBP data. First, we observe the number of pages viewed by each panelist on each shopping occasion, or browsing session in this case. The number of pages viewed per session (PAGES) is a rough measure of search intensity because the more pages viewed, the more products were considered. However, this measure does not take into account variation in the number of products sought by each shopper, and variation in shopping needs across browsing sessions. Therefore, we calculate a second measure of search intensity (SEARCH) by dividing the number of pages viewed by the number of products purchased on that trip. The resulting measure is likely to more closely capture the number of products considered relative to each purchase. Third, we have a measure of the amount of time spent per session. Again normalizing the session duration by the number of products purchased (DURATION) provides a reasonable proxy for the amount of mental processing time used in assessing each of the

products that were purchased. If shoppers did indeed take the time to compare nutritional labels, or read ingredient lists, then DURATION would be much higher relative to a habitual purchase wherein no consideration of attributes took place.

We also have a measure of the unit price paid for each item purchased (PRICE). In order to ensure that the price variable is as comparable as possible across sample subjects, we define the price paid as the average unit price (\$ per relevant unit of measure for that category) across all purchases made by each household, on each purchase occasion, in each category.

Exploiting assortment, search, and price variation across different product categories in a panel-data environment is key to identifying equilibrium behavior on the intra-retailer margin. While others in this literature examine search across retailers (Brynjolfsson and Smith 2000; Clay, Krishnan, and Wolff 2001), they are unable to disentangle the intra-retailer from the inter-retailer, or competitive, effects of search. In our data, shoppers search more or less intensively across different product categories within the same retailer. If a particular consumer searches less in deeper categories, and is evidently willing to pay higher prices for the convenience of finding the ideal product match, then our hypotheses will be confirmed. We document the extent of category-level assortment and search variation in the next section. If our theoretical predictions are correct, then search intensity should be negatively correlated with the level of prices. This expectation is intuitive as orthodox theory suggests that as search increases, the elasticity of demand rises as the consumer becomes aware of more substitutes, and the level of prices falls. However, we describe a the mechanism that links changes in variety more specifically to prices through search intensity. That is, if consumers search for product attributes, then more variety makes this search less costly, search intensity falls, and retail prices rise as a result.

### 3.3 Stylized Facts and Data Summary

In this Section, we provide a profile of the online grocery shoppers in the WBP data relative to a sample taken from a commonly used panel of offline shoppers (Nielsen Homescan). Based on the evidence presented in Degeratu, Wangaswamy, and Wu (2000), Andrews and Currim (2004), and Chu, Chintagunta, and Cebollada (2008), online and offline shoppers differ in important ways; however, it is not immediately obvious that the shoppers in our data share similar attributes to those in other, earlier studies. In our sample, the average online shopper spends \$180.23 per visit, is 52 years of age, and has annual income of \$88,300. For comparison purposes, the average Homescan shopper in 2009 is 50.3 years old, spends \$41.60 on groceries on each shopping trip, and has annual income of \$51,554. The finding that online buyers have more income, and spend more on each shopping occasion is consistent with the literature on online shoppers (Dejeratu, Wangaswamy, and Wu 2000).<sup>6</sup> A more complete profile of the average WBP shopper is provided in table 1 below.

[table 1 in here]

Variation in assortment, and search behavior, across categories is key to our identification assumption. In table 1, we show that, on a category-basis, the mean number of products offered in each category was 5.83, with a standard deviation of 7.51. Although this is considerably lower than the observed number of SKUs in a typical category, either online or offline, recall that our measure of assortment depth is the range of products actually purchased by WBP members over the sample period, and that the comScore category definition is very narrow. As a result, the effective assortment size is more realistically measured than simply including all products stocked at any given time. Note also that the number of pages viewed per category is 80.62 with a standard deviation of 49.69, and the average duration is 45.33 with a standard deviation of 27.88. Taken together, these three pieces of summary data suggest that our data show ample variation in assortment, and search behavior, whether

---

<sup>6</sup>We cannot compare search intensities between online and offline shoppers, because Homescan does not contain a comparable measure of search intensity to either our PAGES or DURATION variables.



measured by the number of pages viewed, or the time spent on each page.

We calculate correlation coefficients among variety, search intensity, and prices, using our alternative measures of search intensity. These results are shown in table 2 below. In the upper matrix of table 1, we show correlation coefficients between variety, search intensity, and prices where search intensity is defined as SEARCH above. In the lower matrix, we define search intensity instead as DURATION, or the amount of time taken to purchase each item. Using the SEARCH definition of intensity, we find a statistically significant negative relationship between variety and search intensity, as predicted by our theory. Retailers are able to limit search behavior by providing greater variety. Moreover, we find a negative, and significant, relationship between SEARCH and prices paid, meaning that when consumers searched less to find the desired product, they paid higher prices. Perhaps not surprisingly, the simple correlation between variety and prices, without considering search, is significant and positive (Trindade 2012; Richards and Hamilton 2014). That is, deeper assortments tend to be associated with higher retail prices in online grocery categories. In the lower matrix, we find a similar set of results when defining search intensity as the time taken to find each product. Namely, a deeper product assortment is associated with lower search intensity and higher prices. However, when measuring search intensity with the DURATION variable, the intervening linkage between search and prices is no longer statistically significant at a 5% level. Simple correlation coefficients, however, do not take into account other factors that may influence the incentive to search, or the clear endogeneity of search. In the next section, we develop a more complete empirical test of our theory of search and retail prices.

[table 2 in here]

### **3.4 An Empirical Model of Variety, Search, and Retail Prices**

In our theoretical model above, we implicitly assume retailer variety decisions are equilibria to a game played among retailers prior to households' search decisions, and retailers' pricing decisions. This assumption is both reasonable and descriptive as retail stocking decisions are

made according to monthly or quarterly cycles according to contracts with suppliers, while pricing decisions are made weekly for each product in the store. Moreover, Brynjolffson, Hu, and Simester (2011) argue that deeper assortments online are facilitated by technological and supply-chain advances that are independent of any household's decision to search. Consequently, retail variety is exogenous to the consumer's decision. Conditional on the amount of variety in each store, the consumer then decides how much to search. Based on the level of revealed search activity, retailers then set equilibrium prices based on their expectations of how shoppers will respond. Our empirical model of variety, search, and retail prices reflects this logic as it consists of two equations: One describing search intensity as a function of retail variety and other factors, and a second that relates search intensity to equilibrium retail prices.

In the retail price equation, however, search intensity is endogenous. Consequently, we instrument for search with variables that are likely to be correlated with the cost of search, and hence the amount of search "supplied" by households, and yet mean independent of retail prices. Because variety is thought to be related to retail prices, it is not an appropriate instrument, even though it is an important determinant of search cost. Most search cost models assume that search behavior is determined instead by travel cost, and cognitive cost, or the dollar-metric amount of time spent thinking about how one product compares to another (Mehta, Rajiv, and Srinivasan 2003). Cognitive cost, in turn, is determined by the innate ability of the decision maker to recall and compare accurately, and his or her opportunity cost of time. Of course, we do not have data on the former, so we instrument search intensity with measures of the opportunity cost of time. Income, age, and the number of children in the household are each measures of the value a household is likely to place on time. Income is a direct and somewhat obvious measure of the opportunity cost of time as it reflects, or should reflect, the value of next-highest use of the shopper's time. For much the same reason, search cost should fall with age. As the shopper moves beyond their peak

earning years, and into retirement, the lower the opportunity cost of time spend shopping.<sup>7</sup>

Household size is expected to be positively related to search costs for two reasons. First, if tastes are randomly distributed among household members, and one shopper is nominated to buy for the entire household, then larger households place a greater decision-making burden on the shopper. For example, in our data, we often observe multiple brands purchased within a single category. Within-household taste heterogeneity is often advanced as a reason for such observed multiple discreteness (Dube 2004). Second, it is reasonable to assume that the shopper is an individual who is also tasked with other household-management duties. Larger households mean greater competition for the shopper’s time, and a higher implicit cost of avoiding these other responsibilities. Finally, the presence of children is expected to be positively related to search costs, both for the same reasons as the household-size effect, and the expectation that adult household-members impose a lower time burden on the shopper than do children.

More formally, the estimated model of search intensity and retail prices is written as:

$$SEARCH_{ht} = \beta_{0h} + \beta_1 AGE + \beta_2 INC + \beta_3 HHS + \beta_4 CHL + \beta_{5h} VAR + \varepsilon_{1ht}, \quad (6)$$

$$PRICE_{ht} = \gamma_{0h} + \gamma_1 AGE + \gamma_2 INC + \gamma_3 HHS + \gamma_4 CHL + \gamma_{5h} SEARCH + \varepsilon_{2ht},$$

where *SEARCH* is the measure of search intensity described above, *AGE* is the age of the shopper, *INC* is household income, *HHS* is the number of individuals in the household, *CHL* is a binary indicator measuring the presence of children in the household, *VAR* is the total number of stock-keeping units offered in each category, *h* indexes households, *t* indexes shopping occasions, and  $\varepsilon_{iht}$  are iid error terms for each equation.<sup>8</sup> Because of the recursive nature of our model, we estimate each equation separately, instrumenting for endogenous search in the manner described above.

---

<sup>7</sup>A counterargument could be made for the fact that advancing age implies increasing scarcity of remaining time, therefore increasing the marginal value. We expect the argument advanced in the text to dominate.

<sup>8</sup>Note that the model for the *DURATION* definition of search intensity is identical.

We exploit the panel nature of our data by estimating both fixed and random-coefficient versions of the model. In the variety-and-search equation, we allow the constant term and response to variety to be randomly distributed on the expectation that unobserved heterogeneity will play an important, yet independent, part of each effect. Specifically, we assume  $\beta_{0h} = \beta_0 + \sigma_{\beta_0}\nu_1$ , where  $\nu_1 \sim N(0, 1)$  and  $\beta_{5h} = \beta_5 + \sigma_{\beta_5}\nu_2$ , where  $\nu_2 \sim N(0, 1)$  and  $\sigma_{\beta_i}$  are scale parameters associated with the random effects. Our expectations are well justified. First, allowing the constant term to vary over households reflects the fact that some people simply like to shop more than others (or dislike shopping less). All else constant, these individuals will search more intensively either motivated by a search for the "transaction utility" associated with finding a deal (Thaler 1985), or because they happen to be more particular in their attribute demands. how much individuals search regardless of variety, and the variety effect itself. Second, we also expect the variety effect to differ over households for reasons that are not reflected in our data. While our theory is based on the behavior expected from a representative shopper, we acknowledge that there may be some individuals in the data for whom variety presents a greater burden – the "choice overload" hypothesis advanced in the literature (Klemperer and Padilla 1997; Iyengar and Lepper 2000; Diehl and Poynor 2010) – so that search intensity actually rises in variety because they are unable to make a decision. By allowing this parameter to vary over households, our model therefore provides an indirect test of the overload hypothesis.

Similarly, we also allow the constant and the search-effect to be randomly distributed over households in the search-and-price equation. Regardless of the amount of search activity undertaken, we expect to find that some shoppers are simply willing to pay more for the same product relative to others out of brand loyalty, and aversion to substitute brands, or simply out of indifference to price changes. With respect to the search-effect, our theory suggests that search and prices are negatively correlated because consumers who are able to find their desired product more easily will be willing to pay more for products that more nearly meet their preferred attribute set. Our hypothesis refers to the mean effect from all

shoppers even though the random parameter admits behavior that may deviate among some members of the sample.

## 4 Results and Discussion

In this section, we first examine the empirical evidence on our hypothesis regarding variety and search behavior, and then the relationship between search and prices. In each case, we first present results from a simpler, fixed-coefficient version of the model, and then compare the fit of this model to our maintained, random coefficient specification.

In the fixed-coefficient specification, we estimate the relationship between search and variety, where search is defined as the number of pages per item purchased (*SEARCH*). These results are shown in table 3 below. Notice that the point estimate of the variety effect on search is indeed negative, as hypothesized, although the effect is not significantly different from zero. When we allow this effect to be randomly distributed over sample households, however, the mean effect becomes nearly twice as large, and significantly different from zero.<sup>9</sup> That is, the deeper the assortment in each category, the less the shopper searches before making a purchase, roughly 1 less page for every 10 products added to the assortment. It is important to note, however, that the standard deviation of the variety effect is over half as large as the mean effect (0.05 versus 0.09) so, given the assumption, of normality, there are clearly members of the sample who shop more when presented with greater variety. Further, the size and significance of the scale of the constant term suggests that there is a wide range of online shopping behaviors represented in the data. While some panel members appear to shop extensively for groceries online, others appear willing to avail themselves of the opportunity to shop very little, if at all. Online shopping is amenable to both types of shoppers. Among the other variables of interest, we find that shoppers who are older, more

---

<sup>9</sup>The random coefficient specification is preferred relative to the fixed-coefficient model according to a likelihood ratio test as the test statistic of 16,608.8 is greater than the critical chi-square value (with 2 degrees of freedom) of 5.991.

wealthy, and who have children tend to search less, as expected, but shoppers representing larger households tend to shop more. In this case, we expect the taste-heterogeneity effect described above dominates the relative lack of available shopping time.

[table 3 in here]

When search is defined instead as the amount of time taken to make each purchase (*DURATION*), a slightly different picture emerges. We still find that the random coefficient model is preferred to the fixed-coefficient alternative ( $\chi^2 = 5,581.7$ ), but the random coefficient estimates differ qualitatively with respect to the effect of age on the propensity to search. Whereas our estimates with the pages definition of search intensity (*SEARCH*) show that older shoppers tend to search less intensively, the random coefficient model suggests that they search more. Although older shoppers may be less comfortable with an online platform, and therefore less willing to shop, it is also true that their opportunity cost of time is lower, so may be more willing to spend time searching. Most importantly, however, the negative relationship between variety and search intensity is robust to whether search is defined as in the previous model, or using the *DURATION* definition of search intensity. Moreover, the relatively large variance of the variety coefficient is also consistent with the previous model, and suggests that there is substantial unobserved heterogeneity in the incentives to search when presented with a deeper assortment. Whether less intensive search implies higher equilibrium prices, however, remains to be confirmed.

[table 4 in here]

Tables 5 and 6 present the price effect of online search using the *SEARCH* definition of intensity and the *DURATION* definition, respectively. In both models, we instrument for endogenous search behavior.<sup>10</sup> The estimates in table 5 show that, whether in the fixed or random-coefficient specification, search and prices are negatively related, as our theory suggests. In other words, if consumers are able to find their desired product more easily,

---

<sup>10</sup>In the first-stage instrumental variables regression, the *F*-statistic is 317.8 for the pages definition of search, and 219.6 for the duration definition, so neither sets of instruments are weak in the sense of Staiger and Stock (1997).

they will be willing to pay higher prices, in equilibrium, than if they have to search more intensively. It is important to remember that ours is a model of attribute search, so consumers are not necessarily searching for low prices, but rather products that are closer to their ideal. How do we know consumers are not searching for low prices only? If our sample of online shoppers were indifferent to attributes, they would simply buy the lowest-price item in each category. Because shoppers do not always buy the lowest-price item, and because there is significant heterogeneity in tastes, we know they are comparing attributes online. We test this more directly by comparing the fixed-coefficient to the random-coefficient specification. In this model, the fixed-coefficient specification implicitly assumes that there is no heterogeneity among sample members in terms of their willingness to pay for grocery items online, whether independent of their search behavior (the constant term), or as a result of their search behavior (the search term). Although the mean effects are similar between the fixed and random-coefficient specifications, we reject the assumption of no heterogeneity both through t-tests of the scale parameters associated with each variable, and a likelihood-ratio test of the validity of restricting both parameters to be non-random ( $\chi^2 = 1,125.6$ ). Our estimates also show how the willingness-to-pay for online groceries varies with observed shopper characteristics, holding search behavior constant. Namely, the estimates in table 5 show that older shoppers with children are willing to pay less for online groceries than others in the sample, but the income and household-size effects are not statistically different from zero. Combined with our findings from the search-and-variety model, these estimates suggest that older online shoppers not only search more, but pay less once they find what they want.

[table 5 in here]

We also estimate the search-and-price model using the *DURATION* definition of search intensity in order to examine the robustness of our finding. As the estimates in table 6 show, the results are very similar. Search and prices paid are again negatively related in when search intensity is defined in terms of the time taken to find a desired product, and

there is substantial heterogeneity in the search-and-price relationship. Age and the presence of children are again found to be the only significant mitigating factors, both of which are inversely related to the willingness to pay for online groceries. Corroborating our previous results with a different definition of search is perhaps not surprising, because the primary factor in increasing the duration of search online is the number of products examined, or the number of pages of information viewed. We do not claim that these two definitions of search intensity are independent, but rather alternative ways of measuring the same activity.

Our findings have a number of important implications for the direction of the online retailing industry. First, as others have found, we find support for the notion that the expansion of online food retailing is not likely to be as pro-competitive as principles of economics would have us believe (Brynjolfsson and Smith 2000; Brown and Goolsbee 2002; Chevalier and Goolsbee 2003; Brynjolfsson, Dick, and Smith 2010). Because online food retailers will, ultimately, be able to sell a far larger variety of items than offline retailers, we expect to see higher variety lead to lower search intensity among online shoppers, and higher retail prices as consumers are better able to find items that match their desired attribute profiles.

Second, an emerging literature on online search behavior provides evidence that online demand elasticities for grocery items tend to be lower than their offline counterparts (Degeratu, Wangaswamy, and Wu 2000; Andrews and Currim 2004; Chu, Chintagunta, and Cebollada 2008). The usual reasoning for such a finding is that online consumers tend to be more time constrained, have higher incomes, purchase both larger item-sizes and a greater number of items on each shopping occasion, are more brand loyal (Danaher, Wilson and Davis 2003), purchase brands more frequently out of habit, or are simply willing to pay more for the convenience of avoiding physical stores. We depart from this literature to explicitly consider the possibility that the larger amount of non-price attribute information available online alters the intensity of attribute search for differentiated products. In purely conceptual research, Alba, et al. (1997) attribute the difference between online and offline elasticities to the



volume of price and attribute information available online. Because online consumers have access to more information on non-price attributes, they are more likely to choose brands based on desired attribute composition, making price a less important consideration in the purchase decision. Similarly, Lynch and Ariely (2000) conduct experiments on wine sales, and find that providing more information on non-price attributes does indeed reduce price sensitivity. We empirically examine a similar question – the linkage between variety, search intensity, and equilibrium prices – and control for unobserved heterogeneity in a panel-data environment to account for many of these confounding factors. In doing so, we arrive at the same place. Namely, our empirical findings strongly support our main hypotheses that search intensity declines with the depth of the product assortment, while equilibrium retail prices rise with assortment depth.

Third, if online shoppers are able to find items that meet very specific attribute demands more easily, then they can benefit in a welfare sense, even if they pay more for the product that they end up purchasing.

Fourth, our findings suggest another mechanism that underlies the common finding that variety and retail prices are positively related (Trindade 2012; Richards and Hamilton 2014). This outcome is more commonly attributed to the softening of retail price competition if retailers compete in variety – a non-price competitive tool – instead of prices. Our study suggests that the mechanism may instead be operate through consumer search behavior and not more traditional means of strategic behavior. Fourth, our findings contradict the theory of the “long tail” in online retailing, albeit indirectly. While the long-tail hypothesis suggests that retailers will be able to sell niche products for higher margins if consumers are better able to search online than offline, our results suggest that more value-added will be generated from packing existing attribute spaces more densely – not forcing consumers to look for radically differentiated products, but items that are more sharply defined within the attribute space spanned by existing products.

## 5 Conclusions

In this article, we examine the likely effect of the growth of online grocery shopping on consumers' search behaviors, and, by extension, equilibrium prices. While the online grocery industry in the US is currently relatively small, the amount of investment by Amazon, FreshDirect, and others as well as the emergence of online grocery shopping in the UK, suggest that online grocery shopping may be the rule rather than the exception in the near future. To answer the question regarding how retail food prices are likely to be affected by the advent of online shopping, we consider the key differentiating factor between online and offline shopping: The ease of search and the potential variety of products offered online.

Our theoretical model of variety, search, and equilibrium prices shows that when consumers search for attributes, increasing variety reduces equilibrium search intensity, which raises retail prices. If we think of consumers searching in an attribute space of limited size for their ideal product, the more densely that attribute space is packed with products, the easier it will be to find the one that comes closest. Because consumers do not have to search hard to find their ideal product, they become less price sensitive, and retail prices rise.

We examine these hypotheses using a unique data set that describes the shopping, and purchasing, behaviors of a panel of online grocery shoppers. Simple summary statistics show that shoppers in our online panel resemble those described elsewhere in the literature: More wealthy than offline shoppers, from larger households, and willing to purchase larger baskets of groceries on each purchase occasion. Further summary analysis shows a significant negative correlation between retail variety and online search intensity, and negative correlation between one of our measures of search intensity and retail prices. More careful econometric analysis, in which we control for the obvious endogeneity of search behavior, confirms the hypothesized negative relationship between variety and search, and between search and retail prices. Consequently, our findings show that higher retail prices associated with deeper retail assortments may not be due to strategic mechanisms as is commonly believed, but rather due to retailers' endogenous response to consumer search behavior. By manipulating

variety strategically, retailers can forestall search, and raise retail prices in equilibrium. In this regard, our findings mirror those of Richards, Allender and Hamilton (2013).

Our findings would be strengthened if they were confirmed in a larger data set, comprising more households, more purchase occasions, and perhaps more direct measures of search activity. While our data includes measures of page views, and time spent viewing each page, it is technologically possible to capture more detail on shoppers' clickstreams. Clickstream data captures exactly what links were followed by the machine each session, providing a more detailed picture of what was considered, and how seriously it was considered. Further, the U.S. market currently provides limited scope for online grocery research, because there really is only one firm that offers online grocery shopping on a national level (Safeway.com). While there are many other regional and local firms in the market (Peapod, Fresh Direct, Relay Foods, etc.) there is no large-scale, consistent, data base that captures the type of behavior required to do the type of analysis conducted here. Finally, a deeper data set, gathered in the future when the industry is more mature, would allow us to develop a more comprehensive set of instruments. Clearly, search behavior is endogenous in models of consumer search, so quality instruments are necessary to identify equilibrium outcomes.

## References

- [1] Alba, J., and J. Lynch. 1997. "Interactive Home Shopping: Consumer, Retailer, and Manufacturer Incentives to Participate in Electronic Marketplaces." *Journal of Marketing* 61:
- [2] Anderson, S. P. and R. Renault. 1999. "Pricing, Product Diversity and Search Costs: A Bertrand-Chamberlin-Diamond Model." *RAND Journal of Economics* 30: 719-735.
- [3] C. Anderson. 2006. *The Long Tail: Why the Future of Business is Selling Less of More* New York: Hyperion.
- [4] Andrews, R. L., and I. S. Currim. 2004. "Behavioural Differences Between Consumers Attracted to Shopping Online versus Traditional Supermarkets: Implications for Enterprise Design and Marketing Strategy." *International Journal of Internet Marketing and Advertising* 1: 38-61.
- [5] Bakos, Y. 1997. "Reducing Buyer Search Cost: Implications for Electronic Marketplaces." *Management Science* 42: 1613-1630.
- [6] Boatwright, P., and J. C. Nunes. 2001. "Reducing Assortment: An Attribute-based Approach." *Journal of Marketing* 65: 50-63.
- [7] Brown, J. and A. Goolsbee. 2002. "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry." *Journal of Political Economy* 110: 481-507.
- [8] Brynjolfsson, E. and M. Smith. 2000. "Frictionless Commerce? A Comparison of Internet and Conventional Retailers." *Management Science* 46: 563-585.
- [9] Brynjolfsson, E., A. A. Dick and M. D. Smith. 2010. "A Nearly Perfect Market? Differentiation vs. Price in Consumer Choice." *Quantitative Marketing and Economics* 8: 1-33.

- [10] Brynjolfsson, E., Hu, Y., and D. Simester. 2011. "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales." *Management Science* 57: 1373-1386.
- [11] Butters, G. R. 1977. "Equilibrium Distributions of Sales and Advertising Prices." *The Review of Economic Studies* 44: 465-491.
- [12] Cachon, G. P., C. Terwiesch and Y. Xu. 2008. "On the Effects of Consumer Search and Firm Entry in a Multiproduct Competitive Model." *Marketing Science* 27: 461-473.
- [13] Chen, P., and L. Hitt. 2003. "Understanding Price Dispersion in Internet-Enabled Markets." Working paper, Carnegie Mellon University, Pittsburgh.
- [14] Chevalier, J. and Goolsbee, A. 2003. "Measuring Prices and Price Competition Online: Amazon and Barnes and Noble." *Quantitative Marketing and Economics* 1: 203-222.
- [15] Chu, J., P. Chintagunta, and J. Cebollada. 2008. "Research Note - A Comparison of Within-Household Price Sensitivity Across Online and Offline Channels." *Marketing Science* 27: 283-299.
- [16] Clay, K., Krishnan, R., and Wolff, E. 2001. "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry." *Journal of Industrial Economics* 49: 521-540.
- [17] Cloud, K. 2014. "Forecast: Online Sales will Near \$100B by 2019." *Packaged Facts*. October 31. (<http://www.theshelbyreport.com/2014/10/31/forecast-online-grocery-sales-will-near-100b-by-2019>). Accessed on December 27, 2014.
- [18] Danaher, P. J., I. W. Wilson, and R. A. Davis. 2003. "A Comparison of Online and Offline Consumer Brand Loyalty." *Marketing Science* 22: 461-476.
- [19] Degeratu, A. M., A. Rangaswamy, and J. Wu. 2000. "Consumer Choice Behavior in Online and Traditional Supermarkets: The Effects of Brand Name, Price, and Other Search Attributes." *International Journal of Research in Marketing* 17: 55-78.

- [20] Diehl, K. and C. Poynor. 2010. "Great expectations?! Assortment Size, Expectations, and Satisfaction." *Journal of Marketing Research* 47: 312–322.
- [21] Dubé, J. P. 2004. "Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks." *Marketing Science* 23: 66-81.
- [22] Food Marketing Institute. Supermarket Facts. (<http://www.fmi.org/research-resources/supermarket-facts>). Accessed October 13, 2014.
- [23] Hong, H., and M. Shum. 2006. "Using Price Distributions to Estimate Search Costs." *RAND Journal of Economics* 37: 257-275.
- [24] Honka, E. 2010. "Quantifying Search and Switching Costs in the U.S. Auto Insurance Industry." (<http://home.uchicago.edu/ehonka/Paper EHonka100310.pdf>).
- [25] Honka, E. and P. Chingtagunta. 2013. "Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry." Paper presented at the International Industrial Organization Conference, Boston, MA. May.
- [26] Innes, R., and S. F. Hamilton. 2014. "Slotting Allowances and Product Variety in Oligopoly Markets." Working paper, California Polytechnic State University, San Luis Obispo, CA.
- [27] Iyengar, S. S., and M. R. Lepper. 2000. "When Choice is Demotivating: Can One Desire too Much of a Good Thing?" *Journal of Personality and Social Psychology* 79: 995–1006.
- [28] Klemperer, P., and A. J. Padilla. 1997. "Do Firms' Product Lines Include too Many Varieties?" *RAND Journal of Economics* 28: 472–488.
- [29] Koulayev, S. 2010. "Estimating Demand in Search Markets: The Case of Online Hotel Bookings (December 14, 2009). FRB of Boston Working Paper No. 09-16. (<http://ssrn.com.ezproxy1.lib.asu.edu/abstract=1523926>).

- [30] Kuksov, D. 2004. "Buyer Search Costs and Endogenous Product Design." *Marketing Science* 23: 490–499.
- [31] Kuksov, D. and J. M. Villas-Boas. 2010. "When More Alternatives Lead to Less Choice." *Marketing Science* 29: 507-524.
- [32] Lynch Jr, J. G., and D. Ariely. 2000. "Wine Online: Search Costs Affect Competition on Price, Quality, and Distribution." *Marketing Science* 19: 83-103.
- [33] Mehta, R., S. Rajiv and K. Srinivasan. 2003. "Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation." *Marketing Science* 22: 58-84.
- [34] MWPVL. International Supply Chain Experience: Amazon Global Fulfillment Center Network. ([http://www.mwpvl.com/html/amazon\\_com.html](http://www.mwpvl.com/html/amazon_com.html)). Accessed October 13, 2014.
- [35] Oppewal, H., and K. Koelemeijer. 2005. "More Choice is Better: Effects of Assortment Size and Composition on Assortment Evaluation." *International Journal of Research in Marketing* 22: 45-60.
- [36] Point of Purchase Advertising International. 2012. *2012 shopper engagement study media topline report*.
- [37] Progressive Grocer. 2011. "PG's 78th Annual Report of the Grocery Industry." (<http://www.progressivegrocer.com/inprint/article/id1684/crunching-the-numbers/>). Accessed on August 23, 2013.
- [38] Richards, T. J. W. Allender and S. F. Hamilton. 2013. "Rivalry in Price and Location by Differentiated Product Manufacturers." *American Journal of Agricultural Economics* 95: 650-668.
- [39] Richards, T. J. and S. F. Hamilton. 2014. "Variety Pass-Through: An Examination of the Ready-To-Eat Cereal Market." Forthcoming in *Review of Economics and Statistics*.

- [40] Staiger, G. and J. H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65: 557-586.
- [41] Tappata, M. 2009. "Rockets and Feathers: Understanding Asymmetric Pricing." *RAND Journal of Economics* 40: 673-687.
- [42] Thaler, R. 1985. "Mental Accounting and Consumer Choice." *Marketing Science* 4: 199-214.
- [43] Train, K. 2003. *Discrete Choice Methods with Simulation* Cambridge University Press: Cambridge, UK.
- [44] Trindade, A. 2012. "Price and Variety in Supermarkets: Can Store Competition Hurt Consumers?" Working paper, Department of Economics, Northwestern University, Evanston, IL.
- [45] Wildenbeest, M. R. 2011. "An Empirical Model of Search with Vertically Differentiated Products." *RAND Journal of Economics* 42: 729-757.



Table 1: Summary of comScore Web Behavior Panel Data

| Variable                | Units  | Mean   | Std. Dev. | Minimum | Maximum |
|-------------------------|--------|--------|-----------|---------|---------|
| Age                     | Yrs.   | 51.95  | 13.89     | 22.50   | 70.00   |
| Income                  | \$,000 | 88.30  | 34.93     | 10.00   | 125.00  |
| Household Size          | #      | 2.72   | 1.52      | 1       | 6       |
| White                   | %      | 86.10% | 34.60%    | 0.00%   | 100.00% |
| Black                   | %      | 4.01%  | 19.61%    | 0.00%   | 100.00% |
| Asian                   | %      | 1.80%  | 13.31%    | 0.00%   | 100.00% |
| Children?               | %      | 49.93% | 50.07%    | 0.00%   | 100.00% |
| Basket Size             | \$     | 180.23 | 53.78     | 37.59   | 342.05  |
| Pages per Household     | #      | 69.01  | 63.51     | 3       | 259     |
| Duration per Household  | Mins.  | 41.75  | 37.40     | 1       | 189     |
| Assortment per Category | #      | 5.83   | 7.51      | 1       | 55      |
| Pages per Category      | #      | 80.62  | 45.33     | 3       | 283     |
| Duration per Category   | Mins.  | 49.69  | 27.88     | 2       | 213     |

Table 2: Correlations Between Variety, Search Intensity, and Prices

|            | Variety            | Search (P)         | Prices             |
|------------|--------------------|--------------------|--------------------|
| Variety    | 1.000              | -0.065<br>(-4.741) | 0.044<br>(3.188)   |
| Search (P) | -0.065<br>(-4.741) | 1.000              | -0.034<br>(-2.498) |
| Prices     | 0.044<br>(3.188)   | -0.034<br>(-2.498) | 1.000              |

  

|            | Variety            | Search (D)         | Prices             |
|------------|--------------------|--------------------|--------------------|
| Variety    | 1.000              | -0.054<br>(-3.917) | 0.044<br>(3.188)   |
| Search (D) | -0.054<br>(-3.917) | 1.000              | -0.018<br>(-1.328) |
| Prices     | 0.044<br>(3.188)   | -0.018<br>(-1.328) | 1.000              |

Note: Variety, search, and prices all defined on a category basis; t-ratios are in parentheses.

Table 3: Variety and Search (P) Regression Results

|                  | Non-Random |         | Random     |         |
|------------------|------------|---------|------------|---------|
|                  | Estimate   | t-ratio | Estimate   | t-ratio |
| Age              | -0.399     | -10.234 | -0.975     | -90.735 |
| Income           | -0.319     | -17.196 | -0.094     | -16.644 |
| Household Size   | 0.810      | 2.201   | 8.063      | 87.560  |
| Children         | -7.618     | -6.195  | -12.385    | -40.639 |
|                  | Non-Random |         | Random     |         |
| Constant         | 89.271     | 38.273  | 111.269    | 277.895 |
| Variety          | -0.053     | -1.606  | -0.094     | -9.381  |
| Scale Parameters |            |         |            |         |
| Constant         |            |         | 39.106     | 414.220 |
| Variety          |            |         | 0.0509     | 9.302   |
| Variance         |            |         |            |         |
| $\sigma$         |            |         | 11.978     | 558.173 |
| LLF              | -28,559.48 |         | -20,525.07 |         |
| AIC              | 8.519      |         | 7.797      |         |

Table 4: Variety and Search (D) Regression Results

|                  | Non-Random |         | Random     |         |
|------------------|------------|---------|------------|---------|
|                  | Estimate   | t-ratio | Estimate   | t-ratio |
| Age              | -0.329     | -5.103  | 0.921      | 42.883  |
| Income           | -0.634     | -20.690 | -0.873     | -95.982 |
| Household Size   | 7.525      | 12.359  | 0.335      | 1.755   |
| Children         | -14.134    | -6.948  | -26.843    | -51.444 |
|                  | Non-Random |         | Random     |         |
| Constant         | 124.624    | 32.297  | 120.125    | 119.135 |
| Variety          | -0.123     | -2.265  | -0.145     | -8.325  |
| Scale Parameters |            |         |            |         |
| Constant         |            |         | 70.723     | 330.758 |
| Variety          |            |         | 0.034      | 2.863   |
| Variance         |            |         |            |         |
| $\sigma$         |            |         | 19.363     | 563.539 |
| LLF              | -25,908.15 |         | -23,117.31 |         |
| AIC              | 8.325      |         | 8.782      |         |

Table 5: Search (P) and Price Regression Results

|                  | Non-Random  |         | Random Parameter |         |
|------------------|-------------|---------|------------------|---------|
|                  | Estimate    | t-ratio | Estimate         | t-ratio |
| Age              | -0.025      | -5.920  | -0.022           | -6.261  |
| Income           | -0.003      | -1.264  | -0.001           | -0.567  |
| Household Size   | 0.135       | 3.495   | 0.037            | 0.996   |
| Children         | -0.791      | -6.096  | -0.453           | -4.200  |
|                  | Non-Random  |         | Random           |         |
| Constant         | 6.127       | 22.069  | 6.138            | 26.387  |
| Search (P)       | -0.004      | -2.618  | -0.004           | -2.818  |
| Scale Parameters |             |         |                  |         |
| Constant         |             |         | 0.880            | 20.353  |
| Search (P)       |             |         | 0.001            | 1.389   |
| Variance         |             |         |                  |         |
| $\sigma$         |             |         | 3.389            | 223.718 |
| LLF              | -14,038.011 |         | -13,475.238      |         |
| AIC              | 7.099       |         | 5.121            |         |

Note: Search is the fitted value from an instrumental variables regression

Table 6: Search (D) and Price Regression Results

|                  | Non-Random  |         | Random      |         |
|------------------|-------------|---------|-------------|---------|
|                  | Estimate    | t-ratio | Estimate    | t-ratio |
| Age              | -0.024      | -5.893  | -0.023      | -6.271  |
| Income           | -0.004      | -1.799  | -0.001      | -0.539  |
| Household Size   | 0.159       | 4.061   | 0.052       | 1.416   |
| Children         | -0.822      | -6.323  | -0.396      | -3.745  |
|                  | Non-Random  |         | Random      |         |
| Constant         | 6.231       | 23.365  | 6.108       | 26.864  |
| Search (D)       | -0.004      | -4.034  | -0.003      | -3.674  |
| Scale Parameters |             |         |             |         |
| Constant         |             |         | 0.826       | 19.167  |
| Search (D)       |             |         | 0.000       | 0.455   |
| Variance         |             |         |             |         |
| $\sigma$         |             |         | 3.391       | 222.972 |
| LLF              | -14,033.345 |         | -13,475.541 |         |
| AIC              | 7.098       |         | 5.120       |         |

Note: Search is the fitted value from an instrumental variables regression