12. Trivariate probit models of pre-purchase/purchase shopping channel choice: clothing purchases in Northern California

Patricia L. Mokhtarian and Wei (Laura) Tang

I. INTRODUCTION

Since becoming a reality in the late 1990s, online shopping has shown sturdy growth. Internet-based retail sales in the US constituted 1.1 percent of total retail sales in 2001 and 2.1 percent in 2004. By 2010, online retail, at $167 billion, accounted for 4.3 percent of total retail sales.[1] Predictions are that online retail sales (excluding travel) will rise to $334.7 billion in 2012.[2] Online purchases of the products of particular interest to the present study are also increasing. Specifically, the percentage of retail spending on apparel, accessories, footwear, and jewelry that occurred online jumped from 1.6 percent in 2001 to 8.4 percent in 2007.[3] Compared to traditional store shopping, the steadily rising trend of online retail sales and the confident predictions of how intensively online shopping (or e-shopping) will be adopted in the future make it increasingly important to understand more about the circumstances under which it is adopted, and its potential impacts on other activities, such as travel. Accordingly, there is considerable interest, within the retail industry and among researchers in marketing and transportation, in better understanding the nature of online shopping adoption, particularly in relationship to the traditional channels of store and catalog.

By now, numerous studies have analyzed (intended or actual) purchase (e.g. Bellman et al., 2000; Bhatnagar et al., 2000; Eastin, 2002; Rhee et al., 2009; Shang et al., 2005) or pre-purchase (search) behavior (e.g. Kulviwat et al., 2004; Rhee et al., 2009; Shim et al., 2001), but we are not aware of any empirical studies modeling the combined choices of pre-purchase and purchase modes. Yet it is important to understand those choices not as separate and independent, but rather as interrelated. Couclelis (2004) offers a valuable conceptual discussion of these interrelated choices. She considers three stages of the shopping process – pre-purchase, purchase, and post-purchase – and for simplicity considers two possible choices at each stage: local (store) or remote (internet). This leads to $2 \times 2 \times 2 = 8$ possible outcomes; Couclelis labels four of them (p. 49) "the traditional shopper (local/local/local), the cybernaut (remote/remote/remote), the good citizen (remote/local/remote) and the free rider (local/remote/ local)" (for further analysis of the latter category, see, e.g., van Baal and Dach, 2005; Huang et al., 2009).

Each of the eight possible outcomes has potentially different implications for transportation as well as for store and internet retailing. For example, Couclelis points out that the free rider (and indeed all four of the outcomes for which the actual purchase is remote rather than local) endangers the health of local retailers. The transportation implications of each pattern are less clear-cut; for example, remote purchases (except of digital goods that are downloaded) generate package delivery trips which may or may not save travel on net, depending on (1) whether the purchase would have taken place otherwise; (2) (if so) whether the store trip on which it would have been purchased was actually eliminated (perhaps other purchases were made in the store; perhaps the store was adjacent to another activity location that would have been visited anyway); and (3) (if so) the relative efficiencies of the eliminated store trip and the generated delivery trip (Mokhtarian, 2004).

In any case, however, it is important to better understand the combinations of choices that are occurring in the population, and the proposed study represents the first known empirical analysis of those combinations. Data limitations impose several constraints: we neglect the post-purchase dimension and the catalog channel, and focus on a single product type – clothing/shoes. Despite those limitations, we believe that both the methodology and the results will be of interest to researchers and practitioners in the field.

The rest of this chapter is organized as follows. The next section provides a brief review of some of the most relevant literature on multichannel shopping behavior. Section 3 presents the empirical context of the present study, including descriptions of the sample and the survey. Section 4 offers a descriptive analysis of the dependent variables in this study, while Section 5 reports on the trivariate probit model of the joint choices of pre-purchase and purchase channels. Section 6 provides further discussion and conclusions.

2. REVIEW OF RELATED LITERATURE

Interest in online shopping appears in the literature well back into the 1990s (e.g., Peterson et al., 1997) – essentially concurrently with the spread of the internet – with speculation about "teleshopping" taking place even earlier (e.g., Howard, 1985; Manski and Salomon, 1987). As late as 2005, however, Balasubramanian et al. (2005, p. 13) commented that "A specific issue that researchers have not tackled in sufficient detail is the choice and use of different channels at various stages of shopping." Nevertheless, in recent years this issue has been addressed by an escalating number of studies, and our review will necessarily be extremely selective and arbitrary (see Cao and Mokhtarian, 2005, for an earlier and more extensive review).

With respect to the pre-purchase stage of shopping, Rhee et al. (2009), following others, distinguish between casual browsing (somewhat undirected online activity without a specific intent to purchase) and searching (online activities directed toward fulfilling a specific purchase intention) – and, of course, the same type of contrast can be made for store shopping as well. This useful distinction is operationalized in our research by the "activeness of searching" explanatory variable, which is significant to the pre-purchase and purchase choices in one of our final models discusssed in Section 5.

A growing literature examines the multi-channel shopper explicitly. For example, Balasubramanian et al. (2005) offer a thoughtful conceptual analysis of the various stages of the shopping process. They identify five factors that are important at each stage of the process (economic goals, self-affirmation, symbolic meaning, social influence and experiential impact, and habit), and show how those factors can lead to the choice of different channels at different stages.

Among empirical studies, Soopramanien and Robertson (2007) blend the pre-purchase and purchase choices by subdividing people into (1) internet buyers, (2) those who browse but do not buy online, and (3) those who do neither, and modeling the choice among those three alternatives. One important way in which their approach differs from ours is that they conceive of a particular purchase as the single choice from among a mutually exclusive and collectively exhaustive set of the three possibilities just described, where the internet buyers' category is not further distinguished. In our study, by contrast, we conceive the choices of purchase channel (store versus internet), store as a pre-purchase channel, and internet as a pre-purchase channel to be three separate choices, and allow for the full set of (eight) combinations to be modeled. Another key difference is that we obtain parallel judgments on store as well as internet. This enables us to (1) directly compare separate perceptions of those two channels in our purchase model rather than making only an internet-versus-non-internet purchase comparison as Soopramanien and Robertson do, and (2) use those channel-specific perceptions in our separate pre-purchase models for store (no or yes) and internet (no or yes).

Schröder and Zaharia (2008) also combine the pre-purchase and purchase choices in their categorization of the shopping patterns of 525 customers of a multi-channel German retailer. They find that most customers use only a single channel, and that motivations differ between store-only, non-store (online and catalog) -only, and multi-channel (browse/search online and then purchase in a store) customers. Some key differences between their approach and ours is that we explicitly include the possibility of using both store and internet channels at the pre-purchase stage, and use discrete choice modeling rather than discriminant analysis and multivariate analysis of variance (MANOVA).

Cao (2012) models the choice of purchase channel for a recent search-good purchase of 540 internet users in the Minneapolis-St. Paul area, and finds the pre-purchase channels used at various stages (awareness, search, and trial) to be significant predictors (also see Cao et al., 2012). Focusing only on the purchase choice, Schoenbachler and Gordon (2002) present a conceptual model of the factors influencing whether an individual is (over time) a multi-channel buyer, single-channel buyer, or non-buyer. Although our sample contains a number of multi-channel buyers, the focus of the present study is on a single purchase. While it is of interest (and the subject of other analyses of this sample) to model frequencies of purchasing via each of multiple channels, it is also of interest to better understand the variables influencing a particular purchase – specifically, the choice of a particular bundle of pre-purchase and purchase channels with respect to a single item being bought. With that aim in mind, we turn now to the empirical context of the present study.

3. EMPIRICAL CONTEXT

3.1 The sample

The data used in this study were collected from an internet-based survey of Northern California residents (see Ory and Mokhtarian, 2007 for details). The purpose of the study is to identify potential population segments and then to investigate e-shopping behavior for each segment by analyzing relationships among the measured variables, rather than to report descriptive statistics of the sample distributions and expect them to reflect the corresponding population. Accordingly, the representativeness of the sample is not our primary concern, because the relationships of interest can be reliably measured even if the sample is not strictly representative (Babbie, 2010; Brownstone, 1998). It is more important to have adequate variability on the dimensions of interest and to have choice shares that are not too unbalanced.

To maximize the computer literacy and knowledge of e-shopping in the sample, two university communities were selected as study sites: Santa Clara and Davis. Both cities contain a large number of internet-literate residents, which helps to enrich the sample with a sizable portion of e-shopping adopters. One difference between the two sites is their regional locations: Santa Clara lies in the heavily urbanized Silicon Valley, while Davis is a smaller college town in the Sacramento metropolitan region.

Some 8,000 recruitment letters were mailed in June 2006 to randomly selected households in those two cities. Approximately 6,500 letters apparently reached their intended addressee and around 1,000 respondents went to the website to complete the survey. In addition, 72 respondents requested and returned a paper version of the survey that was offered as an option. Overall, the response rate was 16 percent, which we considered quite good for an internet survey of this length (117 web pages; the paper version has 19 pages) and complexity. Typical response rates for mail-out/mail-back surveys of the general population are 10–40 percent (Sommer and Sommer, 1997). We presume the higher end of that range to be unlikely for a survey as long as ours, with the additional barrier of being administered over the internet.

Eliminating surveys with incomplete responses on important questions and filling small amounts of missing data with category-specific means resulted in a working sample of 967 cases containing relatively complete data. Because the catalog channel was not well represented in the sample, we focused this study on the individual's purchase intention between store and internet. Also, the sample was split such that approximately half the respondents were asked about a recent book/CD/DVD/videotape ("book") purchase and half about a recent clothing/shoes ("clothing") purchase (referred to as the "key purchase"). These choices were made to represent frequently purchased items, while distinguishing between "experience" goods (those for which sensory perception and trial may be important, such as clothing) and "search" goods (those which are reasonably uniform and predictable, such as books; Peterson et al., 1997). Because the variables influencing purchase channel intention may substantially differ (or be differentially weighted) for different product types, we model behavior on the book and clothing subsamples separately, and in the present study, we only analyze the clothing cases. The initial sample size is 465; the final models have 452 and 390 cases due to missing data.

Table 12.1 presents several characteristics of the sample, including sample statistics for the variables significant in the final model. Average

Table 12.1 Selected characteristics of the sample (clothing cases, N=465)

Characteristic (sample size)	N (%)
Gender (female) (464)	278 (59.9)
Education (465)	
High school diploma or less	23 (4.9)
Some college or technical school	57 (12.3)
Two-year college associates degree	30 (6.5)
Four-year college/technical school degree	120 (25.8)
Some graduate school	59 (12.7)
Completed graduate degree(s)	176 (37.8)
Annual household income (438)	
Less than $15,000	18 (4.1)
$15,000 to $29,999	30 (6.8)
$30,000 to $49,999	54 (12.3)
$50,000 to $74,999	90 (20.5)
$75,000 to $124,999	152 (34.7)
$125,000 or more	94 (21.5)
	Mean (s.d.)
Sociodemographic traits	
Age (years) (452)	46.97 (15.08)
Number of workers (439)	1.60 (0.90)
Number of vehicles (464)	3.13 (1.40)
General attitudinal factors (465)	
Pro-exercise	0.02 (0.76)
Shopping enjoyment	0.07 (0.81)
Store enjoyment	0.11 (0.92)
Channel-specific perceptions: Store (465)	
Enjoyment	0.13 (1.19)
Convenience	−0.33 (0.94)
Post-purchase satisfaction	0.61 (1.01)
Cost savings	0.18 (1.00)
Channel-specific perceptions: Internet (465)	
Enjoyment	−0.15 (1.11)
Convenience	0.32 (0.96)
Post-purchase satisfaction	−0.39 (1.26)
Cost savings	−0.19 (1.15)
Purchase context	
Number of items purchased (465)	1.96 (0.79)
Activeness of searching (465)	2.50 (0.72)
Item was a gift (yes=1; no=0) (463)	0.078 (0.27)
Shopping experience	
Number of product types purchased in store (429)	9.85 (2.56)
Number of product types purchased online (429)	5.10 (2.88)
Store purchase frequency (clothing) (464) [a]	3.00 (0.69)
Internet purchase frequency (clothing) (465) [a]	2.01 (0.82)

Note: a. 1 = never; 2 = once or twice a year; 3 = several times per year; 4 = once a month or more.

characteristics include being middle-aged (47), more likely to be female (60 percent) than male, and having education beyond a four-year college or technical school degree. About three-quarters of the respondents have annual household incomes higher than $50,000. The attitudinal factor scores are explained in Section 3.2.2.

3.2 Survey contents

The survey started with a welcome question, followed by seven parts relating to general and channel-specific shopping attitudes, previous general purchasing experience by channel and a specific recent purchase, shopping frequency for specific product types, usage of information and communication technologies (ICT), and sociodemographics.

3.2.1 Dependent variables

As indicated, the present study aims to analyze the joint choice of pre-purchase and purchase channels, in the context of a recent clothing purchase (the "key purchase"). Specifying the purchase alternatives was straightforward: respondents were asked whether the item(s) was (were) purchased over the internet, in a store, or from a catalog, and we neglect the (relatively few) catalog purchases for this study.

Specifying the pre-purchase alternatives was more complex, since a series of questions asked (1) whether the purchase was an impulse, and if not, (2) the single means by which the respondent first became aware of the item, (3) the (possibly multiple) ways by which s/he directly tried or experienced the product, and (4) the (possibly multiple) other sources of information about the item. Each of the latter three questions presented a number of possible answers, including (depending on the question) store, internet, other people, catalog, other media (electronic distinguished from non-electronic), and none. The sample size did not permit an extensive classification distinguishing each of these stages, so we were forced to collapse them into a single pre-purchase stage. Given data limitations and the purposes of this study, we focused on the two pre-purchase channels of greatest interest, store and internet. A respondent was defined as choosing store for pre-purchase if she reported it in response to the awareness *or* trial *or* other-sources-of-information questions, and similarly for the internet. Thus, for their pre-purchase channel(s) individuals could have chosen store, internet, neither (including sources other than store or internet as well as nothing at all), or both.

Accordingly, the choice in this study is represented by a bundle of three binary variables: the store and internet pre-purchase variables, *PrePurS* and *PrePurI* (each taking on "yes" and "no" values), and the purchase variable, *PurCh* (taking on "store" and "internet" values). In the ensuing discussion, we will represent each of the eight possible alternatives with a three-character string consisting of the store and internet pre-purchase channel choices (in that order) followed by the purchase channel chosen, where "S" = store, "I" = internet, and "0" = not. Thus, for example, "00S" means "neither store nor internet were used pre-purchase, and store was used to purchase", while "S0I" means "store but not internet was used pre-purchase, and internet was used to purchase".

3.2.2 Potential explanatory variables

Developed from an extensive literature review (Cao and Mokhtarian, 2005), the potential explanatory variables measured by the survey fall into six main categories, each described below.

General shopping-related attitudes In Part A, the survey presented a series of 42 general shopping-related statements, with responses ordered on a five-point scale from "strongly disagree" (1) to "strongly agree" (5). Common factor analysis was used to extract 13 (obliquely rotated) factors (see Mokhtarian et al., 2009 for details), and standardized scores on these 13 factors were included as potential explanatory variables. Table 12.2 includes the strongly loading statements for each factor. While some of these factors (e.g. impulse-buying, materialism, shopping enjoyment) could apply about equally well to either shopping channel (and were developed primarily for models of shopping frequency), many of them (e.g. pro-technology, pro-environmental, caution, time consciousness, trustingness, pro-exercise and store enjoyment) could differentially affect shopping channel choices.

Channel-specific shopping experience In Parts B and E, a number of questions were asked with respect to prior experience with shopping by each of the channels store, internet, and catalog. Items significant in the final model include the number of product types out of 15 that the respondent purchased via a given channel within the past year (a measure of "breadth of use" of the channel), and the frequency of purchasing clothing via a given channel ("depth of use"). The latter variable is measured on a four-point ordinal scale (never, once or twice a year, several times per year, and once a month or more), which is treated as continuous for simplicity.

It is conceptually plausible to include such variables in the model, as past experience with a given channel could certainly be expected to influence present choices (Schoenbachler and Gordon, 2002; So et al., 2005). On the other hand, knowing that those who chose a given channel in the past are more likely to choose it in the future does not illuminate *why* that channel was chosen in the first place. Furthermore (as a reviewer pointed out), inclusion of such variables can create an endogeneity bias, because the past experience variables are probably correlated with unobserved variables affecting the utility of current choices. Accordingly, we present two joint choice models in Section 5: one containing shopping experience variables (consistent with other related studies; see Chang et al., 2005) and one excluding them, but we prefer the latter for both the behavioral and statistical reasons just noted.

Purchase context In survey Part C, several questions related to the key purchase were asked, such as whether the item was a gift, how much money was spent, how the item was obtained, the purchase location, and the availability of alternative channels for that specific purchase. These are possibly relevant explanatory variables giving important information on why the particular channel was adopted.

Channel-specific perceptions In Part D, respondents were asked to agree or disagree (on a five-point scale) with 28 channel-specific statements, assuming they were to make a purchase similar to the one discussed in Part C. To reduce the burden on the respondents, they were asked to complete such a set of statements for two of the three main shopping channels (store, internet, and catalog) – the channel chosen for the key purchase, and one alternative. Store was always assumed to be an alternative, so most respondents completed the store-internet pair, with the remainder reporting for store and catalog. As mentioned earlier, the store-catalog cases were excluded from the present analysis.

Common factor analysis was also conducted for this set of statements (details available in Tang, 2010). The statements were pooled across channel and factor-analyzed to find eight underlying dimensions (with scores computed for each dimension for each channel), as shown in Table 12.2 (where only the store version of each statement is shown for brevity). Although participants were asked to respond specifically with respect to a future purchase, the channel-specific perceptions embodied by the factors could logically affect pre-purchase choices as well. Accordingly, we test these factors as explanatory variables in both the pre-purchase (appearing individually) and purchase (appearing as the difference between store and internet factor scores) equations of the models presented in Section 5.

Use of ICT In Part F, the survey asked some general questions about the respondent's usage of the internet, as well as other information and communication technologies (ICT). The responses reflect the individual's

Table 12.2 General attitudes and channel-specific perceptual factors

General attitudes/personality traits/Values Factors[a b]

Factor	Survey Statement (Loading)
Pro-credit card	Credit cards encourage unnecessary spending (−0.573); I prefer to pay for things by cash rather than credit card (−0.514)
Pro-environmental	We should raise the price of gasoline to reduce congestion and air pollution (0.605); To improve air quality, I am willing to pay a little more to use a hybrid or other clean-fuel vehicle (0.556); Shopping travel creates only a negligible amount of pollution (−0.447); A lot of product packaging is wasteful (0.388); Whenever possible, I prefer to walk or bike rather than drive (0.354)
Pro-exercise	I follow a regular physical exercise routine (0.562); Whenever possible, I prefer to walk or bike rather than drive (0.540).
Impulse buying	I generally stick to my shopping lists (−0.586); When it comes to buying things, I'm pretty spontaneous (0.565); I like a routine (−0.289); If I got a lot of money unexpectedly, I would probably spend more of it than I saved (0.273)
Caution	"Better safe than sorry" describes my decision-making style (0.634); Taking risks fits my personality (−0.509); I like a routine (0.319); I am generally cautious about accepting new ideas (0.316); I prefer to see other people using new products before I consider getting them myself (0.265)
Materialism	For me, a lot of the fun of having something nice is showing it off (0.604); I would/do enjoy having a lot of expensive things (0.495); Buying things cheers me up (0.363); My lifestyle is relatively simple, in terms of material goods (−0.302)
Price consciousness	It's too much trouble to find or take advantage of sales and special offers (−0.648); It's important to me to get the lowest prices when I buy things (0.604)
Time consciousness	I'm often in a hurry to be somewhere else when I'm shopping (0.580); I'm too busy to shop as often or as long as I'd like (0.425).
Trendsetting	I often introduce new trends to my friends (0.604); I like to track the development of new technology (0.392)

Trustingness	People are generally trustworthy (0.469); I tend to be cautious with strangers (−0.408); I enjoy the social interactions shopping provides (0.343).
Store enjoyment	Even if I don't end up buying anything, I still enjoy going to stores and browsing (0.769); I like to stroll through shopping areas (0.752); Shopping helps me relax (0.586); Shopping is fun (0.529); For me, shopping is sometimes an excuse to get out of the house or workplace (0.427); Shopping is usually a chore for me (−0.389); Buying things cheers me up (0.293); Shopping is too physically tiring to be enjoyable (−0.285)
Shopping enjoyment	Shopping is too physically tiring to be enjoyable (−0.440); Shopping is usually a chore for me (−0.408); My lifestyle is relatively simple, in terms of material goods (−0.309); "Variety is the spice of life" (−0.267)
Pro-technology	Computers are more frustrating than they are fun (−0.735); The internet makes my life more interesting (0.582); I like to track the development of new technology (0.478); Technology brings at least as many problems as it does solutions (−0.444)
Channel-specific Perceptual Factors[c] (store version)	
Convenience	When it comes to buying clothing/shoes, I can find anything I want in stores (0.640); A lot of times, products I want are unavailable in stores (−0.636); The product information I need is easy to find in stores (0.615); Stores are open whenever I want to shop (0.518); When shopping in stores, it is easy to check the availability of products (0.475); The stores I want/need to shop at are conveniently located (0.447); All things considered, buying in stores saves me time (0.413); I often find shopping in stores to be frustrating (−0.345).
Product risk	I'm concerned that a product I purchase in a store will not perform as expected (e.g. quality, etc.), (0.469); When shopping in stores, I am able to experience products before buying, to the extent that I want to (−0.374); I am concerned that unfamiliar stores will fail to meet my expectations (0.334)
Enjoyment	Shopping in stores is boring (−0.768); I enjoy shopping in stores (0.760); I often find shopping in stores to be frustrating (−0.407); With respect to buying clothes/shoes, I am always on the lookout for a new store to check out (0.323)

Table 12.2 (continued)

Channel-specific Perceptual Factors[c] (store version)	
Financial/identity risk	It is risky to release credit card information to stores (0.838); I am uncomfortable about providing personal information to stores (0.627).
Efficiency/inertia	I value stores that allow me to fulfill many of my shopping needs in just one location (0.449); When it comes to clothing/shoes, I have a strong preference for shopping at one or a few particular stores (0.414); When shopping in stores, I am able to experience products before buying, to the extent that I want to (0.322)
Cost saving	All things considered, buying in stores saves me money (0.760); Considering taxes and other costs, clothes/shoes are usually more expensive when purchased in stores (−0.753).
Store brand independence	I prefer to shop at independent stores rather than national chains (0.561); With respect to buying clothes/shoes, I am always on the lookout for a new store to check out (0.389)
Post-purchase satisfaction	I often have to wait too long for a store to obtain the product I want to purchase (−0.594); Stores typically provide poor after-purchase customer service (−0.559); If necessary, it is easy to return a product purchased at a store (0.486); When shopping in stores, I am able to immediately obtain the products I purchase (0.412); It is difficult to compare products at stores (−0.316)

Notes:

a. Adapted from Mokhtarian et al. (2009). Based on oblique rotation of the common factor analysis solution (Rummel, 1970).
b. Pattern matrix loadings, reflecting the contribution each factor makes to the variance of each observed variable (higher-magnitude loadings reflecting a greater association between variable and factor). Only loadings greater than 0.25 in magnitude displayed.
c. Pattern matrix loadings greater than 0.30 in magnitude are displayed.

overall computer-use pattern, which can help explain the propensity to choose the internet shopping channel in particular. Although such variables are also commonly included in online shopping models (Chang et al., 2005), the same endogeneity problem could apply here as for the past experience variables (an unobserved propensity to use ICT could be correlated with these explanatory variables as well as with the dependent choice variables). However, inclusion of the pro-technology attitude and internet-specific perceptions obtained from Parts A and D of the survey should obviate this concern (moving such a propensity from unobserved to observed). In any case, none of the variables in this group were significant in the final models, perhaps precisely because of the internet channel perceptions that *were* significant.

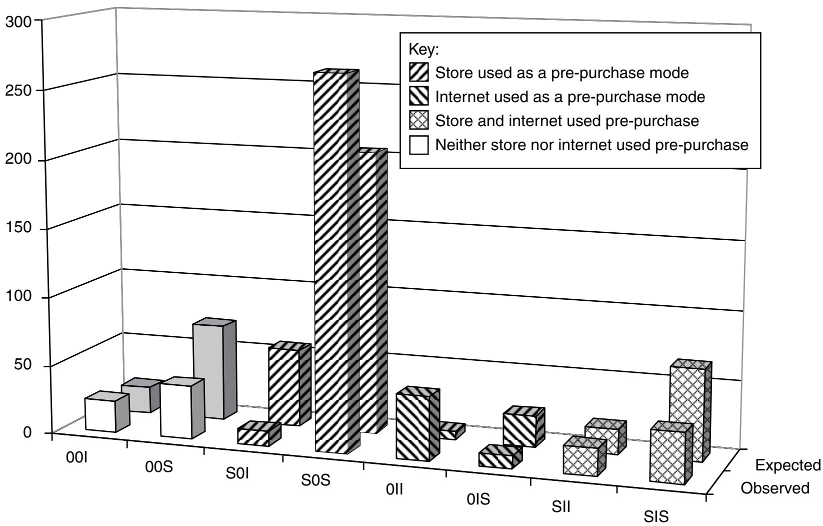
Sociodemographic characteristics Part G of the survey captured an extensive list of sociodemographic variables such as gender, age, employment status (part time or full time), available work arrangements, and educational background, as well as household information such as household income, household size, number of clothing and book stores near home and work, and so on.

4. DESCRIPTIVE ANALYSIS OF CHANNEL CHOICES

If channels were chosen independently at each stage of the shopping process, there would be no need for a joint analysis: each choice could be modeled separately with no loss of information. But to the extent that certain combinations occur more (or less) frequently than would be expected under independence, it becomes of interest to better understand those combinations. The marginal shares for our three dependent variables are as follows:

PrePurS: 74.4% yes, 25.6% no;
PrePurI: 24.3% yes, 75.7% no; and
PurCh: 78.3% store, 21.7% internet.

In view of the way our sample was drawn and the fact that the survey was largely online, these shares should not be taken as representative of the population as a whole (although they are in the expected direction, with store being more common than internet at each shopping stage). Specifically, the internet shares are far larger than would be the case in the general population. Given certain marginal shares, however, it is



Pre-purchase & purchase channel combinations	Observed		Expected		Description
	Number	*Percent*	*Number*	*Percent*	
00I	24	5.2	20	4.2	One-shot cybernaut
00S	40	8.6	71	15.2	One-shot traditional shopper
S0I	11	2.4	57	12.2	Free rider
S0S	277	59.6	205	44.1	Traditional shopper
0II	46	9.9	6	1.4	Cybernaut
0IS	9	1.9	23	4.9	Good citizen
SII	20	4.3	18	3.9	Hybrid cybernaut
SIS	38	8.2	66	14.2	Traditional and good citizen
Total	**465**	**100.0**	**465**	**100.0**	

Figure 12.1 Comparison of observed and expected distributions among the eight possible pre-purchase/purchase channel combinations

appropriate to test whether the eight possible combinations appear to occur independently or not. For example, even though the marginal probability of purchasing via the internet is overestimated by this sample, the probability of *pre*-purchasing via the internet *given* that the purchase was made online can still be properly represented, in which case independence tests can tell us something valid about the population relationships.

Under the null hypothesis of independence, we can estimate the expected number of cases falling into each of those eight possible combinations; for example, the expected number of cases in the SIS combination would be $465 \times 0.744 \times 0.243 \times 0.783 = 65.8$. Figure 12.1 shows the observed and expected numbers of cases for each combination, together with a brief description of each that draws on but modifies Couclelis's (2004) typology.

Consistently with Schröder and Zaharia (2008) and Cao (2012), we see that most (83 percent) of our sample consists of single-channel users (00I, 00S, S0S, and 0II). Two combinations stand out as occurring considerably more often than would be expected if choices were independent: S0S and 0II. These are the two outcomes corresponding to the channel loyalty or "stickiness" patterns: if a shopper uses store but not internet at the pre-purchase stage, he is more likely to purchase via store than would be expected under independence, and if he uses internet but not store at the pre-purchase stage, purchasing online becomes more likely. Conversely, combinations involving one channel at the pre-purchase stage and the other for purchase (S0I, 0IS) are less likely to occur than is predicted under independence; the same is true when both channels are used pre-purchase and the purchase channel is store (SIS), but not when the purchase channel is internet (SII). Finally, the store-purchase-only (00S) alternative is also less likely to occur than predicted.

Using log-linear analysis to compute Cochran-Mantel-Haenszel chi-squared test statistics,[4] we find (details not shown, to conserve space) that all interactions among our three binary variables are statistically significant ($p < 0.0001$), except that, controlling for purchase channel, the choices of pre-purchase channels are independent ($p = 0.502$). Thus, for example, if I purchase online I am more likely to pre-purchase online (Pr[pre-purchase online | purchase online] = 0.65) than in store (Pr[pre-purchase store | purchase online] = 0.31). But, given that I purchase online, I am no more likely to pre-purchase in a store if I pre-purchase online (Pr[pre-purchase store | pre-purchase online, purchase online] = 0.30), than I am if I do not pre-purchase online (Pr[pre-purchase store | do not pre-purchase online, purchase online] = 0.31).

5. TRIVARIATE PROBIT MODEL OF PRE-PURCHASE/PURCHASE CHANNEL CHOICES

As we have defined it, the choice of channels for pre-purchase activities and purchase constitutes three separate – even if not independent – decisions: at the pre-purchase stage, whether to use store or not and whether to use internet or not, and at the purchase stage, whether to use store or internet. Accordingly, it is natural to model the joint choice of pre-purchase and purchase channels with a three-equation binary response model, allowing the error terms to be correlated across equations. Using our knowledge about the pre-purchase choices to inform our predicted probabilities for the purchase choices (and conversely) increases the precision of our estimates (i.e. increases the efficiency of the coefficient estimators). Assuming a multivariate normal distribution for the error terms yields the trivariate probit (TVP) model (Chib and Greenberg, 1998). Estimation was performed with the Limdep 9.0/Nlogit 4.0 software package (Greene, 2007; Chapter N7).

5.1 Model overview

Table 12.3 summarizes the results of the final multivariate probit models with and without shopping experience variables. Although there is no universally reported measure of goodness of fit for such a system of equations, McFadden's R^2, or ρ^2, can be used for the goodness of fit of a multivariate probit model (e.g. Lansink et al., 2003). In this study, consistent with Ben-Akiva and Lerman (1985), ρ^2 is calculated by $1 - \ln[L(\beta)]/\ln[L(EL)]$, where $\ln[L(\beta)]$ and $\ln[L(EL)]$ are the values of the log-likelihood function evaluated at the estimated parameters of the final model and for equal shares, respectively. It varies between 0 and 1, with higher values being better.

The ρ^2s of our final models are 0.40 and 0.38, respectively. Re-estimating the final models without their constant terms[5] shows that the true variables alone produce ρ^2 values of 0.39 and 0.34, meaning that they account for most of the explanatory power of the models (97 percent and 89 percent, from the Hauser, 1978 perspective of decomposition of total information explained by the model). In essence, these variables reduce the importance of the constant terms in the models, thereby helping to explain (rather than just describe, as the constants-only model does) the sizable disparities among the market shares.

All correlation coefficients are strongly significant, with the expected signs: unobserved characteristics important to choosing a given pre-purchase channel (whether store or internet) are positively and very highly correlated with those important to choosing the same channel for the purchase (0.8 for store and 0.7 for internet, where the latter sign shows negative in Table 12.3 because the purchase channel variable takes on the lower value for internet), while unobserved factors are moderately negatively correlated (−0.5) between the two pre-purchase channels (i.e. variables increasing the propensity to choose store tend to decrease the propensity to choose internet, and conversely). This confirms that the three choices are not independent, and thus that it is more efficient to model them jointly rather than separately.

Variables from several different categories are significant in both models, including the general attitudes, channel-specific perceptions, the context of the purchase, and sociodemographics, as well as the shopping experience variables in the first model. All the variables in both models have satisfying interpretations, and all are significant at $p=0.05$ or better, except for two significant at 0.07 which are retained for their conceptual

Table 12.3 Trivariate probit model of pre-purchase and purchase channels for the recent clothing purchase

	Variable type	With shopping experience variables (N = 390)		Without shopping experience variables (N = 452)	
		Coeff.	P-value	Coeff.	P-value
Pre-Purchase Store (PrePurS)					
Constant		0.703	0.026	0.627	0.000
Number of product types purchased in store	experience	0.0840	0.003		
Store convenience	channel perception	0.202	0.005	0.180	0.006
Store post-purchase satisfaction	channel perception			0.168	0.010
Income	sociodemographic	−0.178	0.000		
Pre-Purchase Internet (PrePurI)					
Constant		−1.178	0.004	0.236	0.400
Pro-exercise	general attitude	0.228	0.014	0.236	0.005
Internet enjoyment	channel perception	0.258	0.002	0.234	0.001
Internet convenience	channel perception	0.225	0.012	0.178	0.040
Internet post-purchase satisfaction	channel perception			0.124	0.036
Number of items (of any kind) purchased	context			−0.202	0.013
Activeness of searching	context	0.309	0.013		
Age	sociodemographic	−0.00893	0.050	−0.0134	0.002
Purchase (PurCh)					
Constant		0.577	0.351	0.261	0.167
Number of product types purchased in store	experience	0.0723	0.017		
Store clothing purchase frequency	experience	0.306	0.012		
Internet clothing purchase frequency	experience	−0.541	0.000		
Shopping enjoyment	general attitude	−0.300	0.002	−0.185	0.022

Table 12.3 (continued)

	Variable type	With shopping experience variables (N=390)		Without shopping experience variables (N=452)	
		Coeff.	P-value	Coeff.	P-value
Purchase (PurCh)					
Store enjoyment	general attitude	0.165	0.029		
Pro-exercise	general attitude	−0.293	0.006	−0.166	0.032
Post-purchase satisfaction (store – internet)	channel perception			0.104	0.044
Convenience (store – internet)	channel perception			0.0885	0.066
Cost savings (store – internet)	channel perception			0.0616	0.070
Number of items (of any kind) purchased	context	0.255	0.019	0.274	0.003
Item was a gift	context	−0.440	0.036		
Activeness of searching	context	−0.238	0.032		
Correlation between $\varepsilon_{PrePurS}$ and $\varepsilon_{PrePurI}$		−0.509	0.000	−0.485	0.000
Correlation between $\varepsilon_{PrePurS}$ and ε_{PurCh}		0.793	0.000	0.759	0.000
Correlation between $\varepsilon_{PrePurI}$ and ε_{PurCh}		−0.718	0.000	−0.740	0.000
Number of parameters, K			23		20
Final log-likelihood, LL(β)			−487.157		−583.048
LL for final model without constant terms			−495.728		−617.937
LL for market-share (MS) model			−562.341		−638.814
LL for equally-likely (EL) model, LL(0)			−810.982		−939.908
ρ^2, adjusted ρ^2 (EL base)			0.399, 0.371		0.380, 0.358
ρ^2 of MS model (EL base)			0.307		0.320

relevance. We first interpret the model that includes experience variables, and then discuss how the model without those variables differs.

5.2 The model including experience variables

Only three variables are significant to the *pre-purchase choice of store*. Not surprisingly, the greater the number of product types previously purchased in a store, and the more convenient the store channel is perceived to be, the greater the probability of conducting a pre-purchase activity in a store. Interestingly, the higher one's income, the lower the probability of a store pre-purchase activity, probably reflecting a higher value of time.

Five variables are significant to the *pre-purchase choice of internet*. Again not surprisingly, the more enjoyable and convenient internet shopping is perceived to be, the more likely the respondent is to conduct pre-purchase activities online. "Activeness of searching" is an ordinal context variable taking on the value 1 if "I had not previously thought about buying such an item – I just came across it", 2 if "I had previously thought about buying such an item if I found it, but I was not actively looking for it on this occasion", and 3 if "I was actively looking for such an item on this occasion".[6] The fact that it is positively associated with pre-purchase activities online but not in stores points to the higher efficiency of "letting your fingers do the walking" (to borrow an old slogan for the telephone Yellow Pages directory of businesses) when it comes to purposive information gathering. However, the fact that it is insignificant in the pre-purchase store model, not negatively significant there, suggests that stores are still frequently a venue of active searching – about as often as they are not (whereas the internet attracts active searching substantially more often than not).

Somewhat unexpectedly, a pro-exercise general attitude is positively associated with internet pre-purchase activity (as well as with internet purchasing, discussed below, while not being significant to store pre-purchase activity). Our tentative prior hypothesis had been that exercise-oriented people would prefer store shopping for its greater physical activity. An alternative hypothesis, however, is that internet shopping saves time that can then be applied to more intensive exercise activities. Thus, the result is plausible. Finally for this group, age has the expected negative association with internet pre-purchase activity.

A rich set of nine variables is significant to the *purchase choice between store and internet*: three experience indicators, three general attitudes, and three context variables. Recalling that store is the higher-numbered alternative for the purchase decision, it is natural that the greater the breadth and depth of store purchase experience, the higher the probability of purchasing in a store, while the greater the depth of internet purchase experience (frequency of purchasing clothing), the higher the probability of purchasing online.

It is also natural that a general enjoyment of stores leads to a higher chance of purchasing in a store. Interestingly, enjoyment of shopping in general is associated with a higher probability of buying online. This result is consistent with the finding of Girard et al. (2003) that a recreational shopping orientation is positively associated with a preference for internet shopping. A different analysis (Circella and Mokhtarian, 2010), using the same data as the present study, found the same variable to be significant to internet shopping frequency for clothes, but not to store shopping frequency. Taken together, these results suggest that "shopaholics" indulge their enjoyment of shopping through purchasing more often online than others do, while not necessarily shopping any more or less often in stores than others do. The pro-exercise attitude also increases the probability of buying online, with the same interpretation as for the pre-purchase stage.

With respect to context variables, if the key purchase was a gift, it was more likely to have been bought online. This is quite plausible, as it is easy for the prospective recipient to "give a hint" regarding her wishes through e-mailing a web link or registering online for a specific desired item. The greater the number of items (of any kind) purchased "on this occasion" (not necessarily at the same retailer), the more likely the purchase took place in a store. This hints at a perceived efficiency of store shopping when a variety of purchases, possibly involving multiple retailers, needs to be made. It may not necessarily save time over making a similar set of purchases online, but together with the other advantages of store shopping (opportunity to feel and try, immediate possession, no shipping and handling costs, lower perceived risk of credit/identity theft), the economies of scale for store shopping may outweigh the advantages of online shopping in this type of situation.

Finally, the activeness of searching variable appears here too, with the same interpretation as for the pre-purchase stage: the more purposive the search, the higher the probability of buying online. This points to the convenience permitted by the internet, of being able to sift through a number of retailers' inventories when looking for a specific item. However (since, conversely, the more *impulsive* the purchase, the more likely to buy in a *store*), it also suggests that online retailers may need to develop more creative strategies for inducing customers to buy on impulse.

5.3 The model without experience variables

When the experience variables are excluded from the model, five other variables also drop out, eight remain (not counting the constants and correlation parameters), and six others enter. The new model lacks income, the activeness of searching (lost from both previous equations) and item-was-a-gift context variables, and the store enjoyment general attitude. Common to both models are the store-specific convenience perception (for store pre-purchase); the pro-exercise general attitude, internet-specific enjoyment and convenience perceptions, and age (for internet pre-purchase); and the shopping enjoyment and pro-exercise general attitudes and number-of-items-purchased context variable (for purchase). Interestingly, none of the three significant channel-specific perceptions dropped out from the first model, and five of the six new variables in the second model are also channel-specific perceptions. This indicates that these variables are rather robust indicators of utility, and the fact that the experience variables displace five perception variables when experience is allowed into the model indicates that these perception variables help to explain the choices comprising that past experience. Thus, although the model including shopping experience variables has a higher goodness of fit (as expected), the fact that it is not very much higher (0.40 versus 0.38) shows that the diverse array of other explanatory variables available is providing the bulk of the behavioral content of the model.

Three of the five new perception variables appear in the purchase equation (which lost three experience and three other variables from the first model). All have the expected positive sign, meaning that the more superior store is perceived to be than internet on the perceptual dimension in question, the more likely the purchase took place in a store. The convenience perception had already appeared in the internet pre-purchase equation; it now appears in the purchase equation as well, accompanied by two new perception variables. The cost savings variable is self-explanatory; the post-purchase satisfaction factor is based on disagreement with statements such as (Table 12.2) "I often have to wait too long to obtain the product I want to purchase" and "[Stores/ The internet] typically provide[s] poor after-purchase customer service", and agreement with statements such as "If necessary, it is easy to return a product purchased [at a store/over the internet]".

The channel-specific scores on the post-purchase satisfaction factor now appear in their respective pre-purchase equations as well. The ubiquity of this variable in all three equations from which experience has been excluded suggests that one's post-purchase satisfaction with a given channel (together with the natural variables convenience and cost) is an important component of one's experience with that channel. The sixth new variable to appear in the second model is the number of items purchased, where it is now negatively associated with the pre-purchase use of the internet, while remaining significant in the same direction (negatively associated with buying online, relative to store) in the purchase equation.

5.4 Variables not significant in either model

Although the models in general are quite interpretable, it is also of interest to review some variables that are *not* significant in either one. For example, among the general attitudes that might have been expected to be relevant, pro-technology, pro-environmental, caution, time consciousness, trendsetting, and trustingness did not enter either model. However, many of these dimensions are likely to be tapped by the channel-specific perceptions that did enter the models. For example, the presence of channel-specific convenience and post-purchase satisfaction may account for the impact of time consciousness, and internet-specific enjoyment (as well as internet purchase frequency) may serve a role similar to pro-technology. The latter may also be true for the ICT experience variables, none of which were significant in either model.

Three of the eight channel-specific perceptions were not significant: product risk, financial/identity risk, and efficiency/inertia (capturing a preference for sticking to one or a few retailers). Interestingly, the two risk variables are often advanced as a reason for shoppers to be reluctant to buy online, especially (in the case of product risk) for an experience good such as clothing. While the financial/identity risk was undoubtedly more salient in the early days of internet shopping, it may well be that as the practice has become mainstream (nearly 75 percent of U.S. internet users have bought products online, according to Table 1120 of the *2010 Statistical Abstract*, http://www.census.gov/compendia/statab/cats/information_communications.html, accessed August 13, 2010), this fear has largely dissipated – or perhaps is similarly salient to both channels, since, after all (as the popular media pointed out while online shopping was in its infancy), a store clerk (or household trash forager) can also steal a shopper's credit card number with relative ease. Product risk should indeed be higher for clothing online, but no more so than for catalog purchases (less so, in fact, given the greater richness of information available online), which has long been a viable channel for the clothing product type. And again, the channel-specific convenience and post-purchase satisfaction factors may partly be accounting for any perceived product risk differential between store and internet.

It is also interesting that age is the only sociodemographic variable significant in the second model, and age and income are the only two in the first model. A number of other significant relationships with sociodemographic variables could be postulated, but our speculation is that in many other empirical contexts such variables serve as limited markers for the kinds of attitudinal variables that are already included in our models. Thus, with variables like convenience, cost savings, and enjoyment in the models, the absence of variables such as gender and household size is not necessarily remarkable.

Several studies (e.g. Farag et al., 2006; Forman et al., 2009; Ren and Kwan, 2009) have examined the relevance to shopping behavior of the geographical context in which the shopper lives and works. The dataset used in the present study contained only two such indicators, namely three-point ordinal variables measuring how many clothing stores were within a 10-minute walk from the respondent's home and workplace, respectively. Neither of those variables was significant in our models; this could be because of their simplicity, or again because the impact of urban form might be accounted for by perceptions that were included, such as the convenience and store enjoyment factors.

5.5 Comparison of aggregate shares predicted by independent and joint models

Since estimating a trivariate probit model is a non-trivial exercise, it is reasonable to ask whether the improved ability to capture correlations among pre-purchase and purchase channels is worth the greater effort involved. Put another way, how badly wrong would we be, if we ignored those correlations and simply modeled each of the three choices independently? To answer this question, Table 12.4 compares the aggregate shares computed (using probability weights) from the independence model (comprising the three binary probit models) to those computed from the joint model (in both cases the one without the shopping experience variables). The upper block contains the shares for the three marginal binary choices, while the lower block treats the eight joint choices.

Turning first to the binary choice shares, we confirm that the three binary probit models essentially replicate the observed market shares of the marginal choices, as would be expected of any model with a constant term.[7] The trivariate probit model does not replicate the observed marginal shares quite as well (which is not surprising since it is essentially "considering" the eight joint alternatives rather than the three marginal choices in isolation from each other), but still, the shares computed from this model differ from the observed shares by at most 1.7 percent (for using the internet in a pre-purchase activity: observed share 24.6 percent; predicted share 24.1 percent). Thus, both models recover the marginal shares satisfactorily.

Not surprisingly, the more dramatic differences lie in the predicted

Table 12.4 Comparison of aggregate shares (N = 452)

				Independent			Joint		
		Observed frequency	Observed share	Prob (each alt)	Raw difference from observed	% difference from observed	Predicted share	Raw difference from observed	% difference from observed
Marginal probabilities									
Pre-purchase	no	117	0.2588	0.2588	0.0000	−0.0106	0.2590	0.0001	0.0468
Store	yes	335	0.7412	0.7412	0.0000	0.0037	0.7410	−0.0001	−0.0201
Pre-purchase	no	341	0.7544	0.7546	0.0002	0.0255	0.7586	0.0042	0.5536
Internet	yes	111	0.2456	0.2454	−0.0002	−0.0784	0.2414	−0.0042	−1.7119
Purchase	internet	100	0.2212	0.2216	0.0004	0.1642	0.2185	−0.0027	−1.2321
	store	352	0.7788	0.7784	−0.0004	−0.0466	0.7815	0.0027	0.3465
Joint probabilities									
00I		23	0.0509	0.0428	−0.0081	−15.8641	0.0524	0.0016	3.0550
00S		39	0.0863	0.1489	0.0626	72.5435	0.0952	0.0089	10.3076
S0I		11	0.0243	0.1061	0.0818	336.1101	0.0268	0.0025	10.1795
S0S		268	0.5929	0.4568	−0.1361	−22.9583	0.5842	−0.0087	−1.4756
0II		46	0.1018	0.0221	−0.0796	−78.2513	0.0912	−0.0106	−10.4143
0IS		9	0.0199	0.0450	0.0251	126.0005	0.0202	0.0003	1.3635
SII		20	0.0442	0.0505	0.0063	14.1822	0.0481	0.0038	8.6802
SIS		36	0.0796	0.1277	0.0481	60.3671	0.0819	0.0023	2.8655
Total/overall		452	1.0000	1.0000	0.0000	–	1.0000	0.0000	–

shares for the eight joint choices. The errors mirror those already evident in the descriptive analysis around Figure 12.1: the two "sticky" patterns of S0S and 0II are badly *under*predicted by the independence model (predicted shares 23 percent and 78 percent too low, respectively), while the 00S (73 percent), S0I (336 percent) and SIS (60 percent) patterns are especially badly *over*predicted. By contrast, predicted shares for the trivariate probit model are never more than 10.4 percent off; the S0S share is too low by only 1.5 percent. The trivariate probit model is dramatically superior here.

"So what?" the devil's advocate might ask. "Since the simpler model replicates the marginal shares just fine – even better than the more complicated model, in fact – do we really need the additional complexity of trying to predict the joint shares? Why not just adopt the simpler model?" The answer is twofold. First, as implied by the above discussion, comparing the marginal shares to the joint shares for the independence model shows that the nearly exact replication of the marginal shares is the net, in each case, of the sizable underpredictions for some joint alternatives being essentially counteracted by the sizable overpredictions for others. We have argued as the premise of this chapter that the pre-purchase/purchase *combinations* are of interest in their own right, and to the extent that this is true, it obviously improves our understanding of those combinations to have a model that is sensitive to their occurrence (as the independence model is not).

Second, however, suppose that for some reason we were only interested in the marginal shares. Then our future predictions even of those shares will probably be better with the joint model. If the joint distributions of observed and unobserved explanatory variables that are exhibited by the calibration data set remain stable, the simpler model will continue to work as well – for predicting the marginal shares. But if those distributions change (e.g., if unobserved characteristics begin to favor the internet more strongly than before), the more complex model is likely to be better able to predict the resulting new market shares. Again, this is because it is using data on each choice to inform the prediction not just regarding that choice, but regarding the other two choices as well. For example, the trivariate probit model "knows" that someone who becomes more likely to use the internet for pre-purchase activities will also tend to be more likely to purchase online. By contrast, the independence model does not share information across choices: the "knowledge" that an individual became more likely to use the internet for pre-purchase activities has no impact on the predicted marginal probability that she will purchase online. That probability can still change, as a function of changes in the observed and unobserved explanatory variables, it is just that the independence model cannot assess the *extent* to which it is predicted to change as well as the joint model can.

6. DISCUSSION AND CONCLUSIONS

To our knowledge, this study is the first to model the joint choice of pre-purchase (store no vs. yes and internet no vs. yes) and purchase (store vs. internet) channels, through the application of a trivariate probit model to a "recent" purchase of clothing made by a sample of more than 450 Northern California residents. The descriptive analysis clearly showed dependence across these three choices: in particular, the "sticky" combinations of {only-store pre-purchase + store purchase} and {only internet pre-purchase + internet purchase} occurred substantially more often than independent choices would predict. Chi-squared tests showed that the two pre-purchase channel choices of store and internet are each related to the purchase channel (store versus internet), but conditional on purchase channel, the two pre-purchase channel choices were independent of each other.

The models showed that this dependence is due to common variables both observed (e.g., channel-specific perceptions of convenience and post-purchase satisfaction appearing in all three submodels) and unobserved (strong correlations between unobserved variables favoring a given pre-purchase choice and those favoring the corresponding purchase choice). Thus, joint estimation is important. Although both the trivariate probit model (taking dependence of unobserved characteristics into account) and independent binary choice models replicated marginal shares reasonably well, the trivariate probit model was markedly superior with respect to recovering joint shares, i.e. shares of pre-purchase/purchase channel combinations.

The models contained a behaviorally rich set of explanatory variables. In addition to breadth and depth of experience variables and channel-specific perceptions (post-purchase satisfaction, cost savings, enjoyment, and convenience), significant variables included general shopping-related attitudes (pro-exercise, shop enjoyment, and store enjoyment), context variables, and sociodemographic traits (age and income). Because of the numerous meaningful perception/attitude measures available, the model without experience variables performed almost as well as the one with experience – the attitudinal variables, in essence, serving to explain the past experience – and avoided the endogeneity bias inherent in including such variables.

Several directions for future research are indicated. Using the data already collected, it is of interest to conduct a similar analysis on the subsample (excluded from this chapter) purchasing a book/CD/video, to explore how the relationships identified in the present study might differ by product type. With a larger sample obtained from new data collection, it would be desirable to extend this methodology to more complex patterns. For example, it could be useful to develop a joint model of channel choices for {awareness, information, trial, purchase}. Perhaps more importantly, it would be valuable to include catalog as a pre-purchase and purchase channel, since it can operate with store and internet in complex ways. This would generate a rarely seen multivariate model in which the (three) pre-purchase choices were binary (since each of the three channels could be chosen separately from the others), while the purchase choice was multinomial (since the purchase transaction would typically occur via one and only one of the three channels, where the catalog channel could be interpreted as "phone or mail").

It would also be useful to apply other model structures in this context. In particular, a structural equations model of relationships among personality traits, lifestyle and socioeconomic characteristics, channel-specific perceptions, and channel choice would help disentangle the multiple directions of causality that are suppressed in the essentially unidirectional models of the present study. Longitudinal studies could help further distinguish between state dependence and unobserved heterogeneity influences on choice.

Finally, it would be valuable to measure several additional variables. For example, it would be useful to capture perceptions of channels with respect to pre-purchase activities separately from those with respect to the purchase, as those perceptions could differ considerably. After all, the perception of financial/identity risk is not very important for online browsing only, whereas it can be quite important for online purchasing. Also, it is of interest to obtain more detailed appraisals of the geographical context in which the shopper lives and works, as our knowledge of how that context affects channel choice is still limited. Simple ordinal variables measuring how many stores were within a 10-minute walk from the respondent's home and workplace were not significant in our models, but these relatively primitive indicators of the surrounding retail environment could be greatly enriched in future studies. Ultimately, although it was not surprising to find correlated error terms across our three submodels, greater behavioral insight would result if we are able to isolate and observe those correlated predictors, leaving mostly independent variation in the unobserved factors associated with each choice.

ACKNOWLEDGEMENTS

Data collection was funded by the University of California Transportation Center. The present analysis was supported by a grant from the Sustainable Transportation Center at the University of California, Davis, which receives funding from the U.S. Department of Transportation and Caltrans (the California Department of Transportation) through the University Transportation Centers program. The authors wish to thank Xinyu (Jason) Cao, who participated heavily in the early stages of the survey design and also provided helpful comments on an earlier draft of this chapter; Susan Handy, who provided helpful input on the survey design and administration; Tara Puzin, who collected and organized Census statistics on the study areas; David Ory, who finalized the survey design, conducted the data collection and initial data cleaning activities, and performed the factor analysis on the general attitudes/personality traits/values; and Gil Tal, who conducted some preliminary analyses for the present study. We are also grateful to Bill Greene for his timely and informative help with some crucial Limdep and methodological questions. An anonymous reviewer also provided useful suggestions for improvement.

NOTES

1. Computed from "Latest Quarterly E-Commerce Report" at http://www.census.gov/retail/, accessed June 15, 2011.
2. Source: http://www.census.gov/compendia/statab/tables/09s1015.pdf, accessed July 13, 2009.
3. Source: http://www.census.gov/compendia/statab/cats/wholesale_retail_trade/online_retail_sales.html (computed from Tables 1054 and 1055), accessed April 8, 2011.
4. See, e.g., http://faculty.vassar.edu/lowry/abc.html, accessed August 27, 2010.
5. This is a heuristic approach, as one would ordinarily not accept as final a model without constant terms, and the coefficients of such a model would generally not be consistent estimators of the true values (Ben-Akiva and Lerman, 1985). Thus, we would not formally interpret such a model, but only use it to assess what proportion of the total "log-likelihood distance" (between the EL model and the perfect model with log-likelihood of 0) is traversed by a model containing only those variables.
6. These responses appeared in the opposite order in the original survey, but the resulting variable was reversed for greater ease of interpretation.
7. In contrast to the logit model, the probit model does not guarantee the exact replication of market shares, but it will generally be very close, and in this application the predicted share differs from the observed share by at most 0.16%, with that worst case occurring for the internet purchase share.

REFERENCES

Babbie, Earl (2010), *The Practice of Social Research*, 12th edn, Belmont, CA: Wadsworth Pub. Co.

Balasubramanian, Sridhar, Rajagopal Raghunathan, and Vijay Mahajan (2005), 'Consumers in a multichannel environment: product utility, process utility, and channel choice', *Journal of Interactive Marketing*, **19** (2), 12–30.

Bellman, S., G.L. Lohse, and E.J. Johnson (2000), 'Predictors of online buying behavior', *Communications of the ACM*, **42** (12), 32–38.

Ben-Akiva, Moshe and Steven R. Lerman (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, Cambridge, MA, USA: MIT Press.

Bhatnagar, A., S. Misra, and R.H. Rao (2000), 'On risk, convenience and internet shopping behavior', *Communications of the ACM*, **43** (11), 98–105.

Brownstone, David (1998), 'Multiple imputation methodology for missing data, non-random response, and panel attrition', Chapter 18 in T. Garling, T. Laitila, and K. Westin (eds), *Theoretical Foundations of Travel Choice Modeling*, Amsterdam: Elsevier, 421–449, available at http://www.its.uci.edu/its/publications/papers/ITS/UCI-ITS-WP-97-4.pdf (accessed 9 August 2012).

Cao, Xinyu (2012), 'The relationships between e-shopping and store shopping in the shopping process of search goods', *Transportation Research A*, **46** (7), 993–1002.

Cao, Xinyu and Patricia L. Mokhtarian (2005), *The Intended and Actual Adoption of Online Purchasing: A Brief Review of Recent Literature*, Research Report No. UCD-ITS-RR-05-07, Institute of Transportation Studies, University of California, Davis, available at http://pubs.its.ucdavis.edu/ (accessed 9 August 2012).

Cao, Xinyu (Jason), Zhiyi Xu, and Frank Douma (2012), 'The interactions between e-shopping and traditional in-store shopping: an application of structural equations model', *Transportation*, **39** (5), 957–974.

Chang, Man Kit, Waiman Cheung, and Vincent S. Lai (2005), 'Literature derived reference models for the adoption of online shopping', *Information & Management*, **42**, 543–559.

Chib, Siddhartha and Edward Greenberg (1998), 'Analysis of multivariate probit models', *Biometrika*, **85** (2), 347–361.

Circella, Giovanni and Patricia L. Mokhtarian (2010), 'Complementarity or substitution of online and in-store shopping: an empirical analysis from Northern California', Paper No. 10-0894 presented at the 89th online Annual Meeting of the Transportation Research Board, Washington, DC, January, available on the conference DVD and from the authors.

Couclelis, Helen (2004), 'Pizza over the internet: e-commerce, the fragmentation of activity and the tyranny of the region', *Entrepreneurship and Regional Development*, **16**, 41–54.

Eastin, M.S. (2002), 'Diffusion of e-commerce: an analysis of the adoption of four e-commerce activities', *Telematics and Informatics*, **19**, 251–267.

Farag, Sendy, Jesse Weltevreden, Ton van Rietbergen, Martin Dijst, and Frank van Oort (2006), 'E-shopping in the Netherlands: does geography matter?' *Environment and Planning B: Planning and Design*, **33**, 59–74.

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

Girard, Tulay, Pradeep Korgaonkar, and Ronnie Silverblatt (2003), 'Relationship of type of product, shopping orientations, and demographics with preference for shopping on the internet', *Journal of Business and Psychology*, **18** (1), 101–120.

Greene, W.H. (2007), *NLOGIT Version 4.0 Reference Guide*, Plainview, NY: Econometric Software, Inc.

Hauser, John R. (1978), 'Testing the accuracy, usefulness, and significance of probabilistic choice models: an information-theoretic approach', *Operations Research*, **26** (3), 406–421.

Howard, E.B. (1985), 'Teleshopping in North America', *Environment and Planning B: Planning and Design*, **12** (2), 141–150.

Huang, P., N.H. Lurie, and S. Mitra (2009), 'Searching for experience on the web: an empirical examination of consumer behavior for search and experience goods', *Journal of Marketing*, **73** (2), 55–69.

Kulviwat, S., C. Guo, and N. Engchanil (2004), 'Determinants of online search: a critical review and assessment', *Internet Research*, **14** (3), 245–253.

Lansink, A.O., M. van den Berg, and R. Huirne (2003), 'Analysis of strategic planning of Dutch pig farmers using a multivariate probit model', *Agricultural Systems*, **78** (1), 73–84.

Manski, Charles F. and Ilan Salomon (1987), 'The demand for teleshopping: an application of discrete choice models', *Regional Science and Urban Economics*, **17**, 109–121.

Mokhtarian, Patricia L. (2004), 'A conceptual analysis of the transportation impacts of B2C e-commerce', *Transportation*, **31** (3), 257–284, author's post-print available at http://repositories.cdlib.org/postprints/37/ (accessed 9 August 2012).

Mokhtarian, Patricia L., David T. Ory, and Xinyu Cao (2009), 'Shopping-related attitudes: a factor and cluster analysis of northern California shoppers', *Environment and Planning B: Planning and Design*, **36** (2), 204–228.

Ory, David T. and Patricia L. Mokhtarian (2007), *Description of a Northern California Shopping Survey Data Collection Effort*, University of California, Davis, Institute of Transportation Studies, Research Report No. UCD-ITS-RR-07-03, March, available at http://pubs.its.ucdavis.edu/publication_detail.php?id=1071 (accessed 9 August 2012).

Peterson, Robert A., Sridhar Balasubramanian, and Bart J. Bronnenberg (1997), 'Exploring the implications of the internet for consumer marketing', *Journal of the Academy of Marketing Science*, **25** (4), 329–346.

Ren, Fang and Mei-Po Kwan (2009), 'The impact of geographic context on e-shopping behavior', *Environment and Planning B: Planning and Design*, **36**, 262–278.

Rhee, Hyeun-Suk, Frederick J. Riggins, and Cheongtag Kim (2009), 'The impact of product type and perceived characteristics of the web on multifaceted online shopping behavior', *Journal of Organizational Computing and Electronic Commerce*, **19** (1), 1–29.

Rummel, R.J. (1970), *Applied Factor Analysis*, Evanston, IL: Northwestern University Press.

Schoenbachler, Denise D. and Geoffrey L. Gordon (2002), 'Multi-channel shopping: understanding what drives channel choice', *Journal of Consumer Marketing*, **19** (1), 42–53.

Schröder, Hendrik and Silvia Zaharia (2008), 'Linking multi-channel customer behavior with shopping motives: an empirical investigation of a German retailer', *Journal of Retailing and Consumer Services*, **15** (6), 452–468.

Shang, Rong-An, Yu-Chen Chen, and L. Shen (2005), 'Extrinsic versus intrinsic motivations for consumers to shop on-line', *Information & Management*, **42**, 401–413.

Shim, S., M.A. Eastlick, S.L. Lotz, and P. Warrington (2001), 'An online pre-purchase intentions model: the role of intention to search', *Journal of Retailing*, **77**, 397–416.

So, W.C.M., T.N.D. Wong, and D. Sculli (2005), 'Factors affecting intentions to purchase via the internet', *Industrial Management and Data Systems*, **105** (9), 1225–1244.

Sommer, Barbara and Robert Sommer (1997), *A Pracical Guide to Behavioural Research: Tools and Techniques*, 4th edition, New York: Oxford University Press.

Soopramanien, Didier G.R. and Alastair Robertson (2007), 'Adoption and usage of online shopping: an empirical analysis of the characteristics of "buyers", "browsers" and "non-internet shoppers"', *Journal of Retailing and Consumer Services*, **14** (1), 73–82.

Tang, Wei (Laura) (2010), *An Investigation of E-shopping for Clothing and Books, with a Focus on Taste Heterogeneity: Evidence from Northern California*, Doctoral dissertation, Transportation Technology and Policy program, University of California, Davis, March.

van Baal, Sebastian and Christian Dach (2005), 'Free riding and customer retention across retailers' channels', *Journal of Interactive Marketing*, **19** (2), 75–85.