JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **23**: 351–371 (2008)
Published online 29 February 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/jae.984

LEARNING AND FATIGUE DURING CHOICE EXPERIMENTS: A COMPARISON OF ONLINE AND MAIL SURVEY MODES

SCOTT J. SAVAGE[1] AND DONALD M. WALDMAN[1,2]*

[1] *Department of Economics, University of Colorado, Boulder, Colorado, USA*
[2] *Center for Health Economics Research and Evaluation, University of Technology, Sydney, Australia*

SUMMARY

This study investigates the effect of survey mode on respondent learning and fatigue during repeated choice experiments. Stated preference data are obtained from an experiment concerning high-speed Internet service conducted on samples of mail respondents and online respondents. We identify and estimate aspects of the error components for different subsets of the choice questions, for both mail and online respondents. Results show mail respondents answer questions consistently throughout a series of choice experiments, but the quality of the online respondents' answers declines. Therefore, while the online survey provides lower survey administration costs and reduced time between implementation and data analysis, such benefits come at the cost of less precise responses.

1. INTRODUCTION

Stated preference (SP) data are used extensively by economists, marketers, and policy makers to estimate individual's willingness to pay for multidimensional goods not traded in markets. SP data are often obtained from respondent choices in experiments administered through informal 'pencil and paper' surveys mailed to sample populations. When designing choice experiments, researchers pay careful attention to the number of alternatives the respondent can choose from, the number of attributes used to describe alternatives, appropriate wording of attribute descriptions, and the number of choice scenarios (or question replications) per respondent. Given the relatively high cost of developing and administering a statistically appropriate mail survey instrument, researchers often trade off aspects of choice task complexity with sample size and survey response quality. For instance, a relatively small sample of respondents may be asked to answer repeated choice questions to simultaneously reduce data collection costs and increase the number of observations available for estimation of marginal utilities. If respondents learn about their preferences and become more proficient at the choice task as they move through more question occasions, the quality of the data improves. Alternatively, multiple question occasions may induce fatigue or boredom. If respondents become tired or bored as they move through the repeated choice questions, the quality of the data deteriorates.

Given the tradeoffs described above it is not surprising that recent rapid growth in US Internet penetration has corresponded with greater interest in online or 'web' surveys administered through the Internet and other networks. Compared to traditional telephone and mail survey modes, online surveys have low marginal costs of providing completed surveys and reduced time between survey

* Correspondence to: Donald M. Waldman, Department of Economics, University of Colorado at Boulder, 256 UCB, Boulder, CO 80309, USA. E-mail: Waldman@colorado.edu

implementation, data collection, and analysis. Online surveys may also have a more refined visual appearance to which a variety of shapes and color may be added, can provide dynamic, three-dimensional interaction, and may facilitate extensive and difficult skip patterns (Dillman, 2000). 'Drop-down' menus and 'pop-up' boxes (with video and audio) can be added to provide additional, instantaneous explanation and help with choice questions without having to direct the respondent to a separate set of instructions on another page or accompanying booklet. These features are attractive because for a given budget researchers may be able to enhance data quality by designing better choice experiments, increasing sample size, decreasing/increasing choice replications per respondent, or some combination thereof. By providing easier, seamless access to enhanced descriptions of attributes and market scenarios, online technology can permit respondents to better understand and evaluate the tradeoffs they face in choice scenarios. In this respect, online surveys have the potential to enhance respondent learning and consistency in choice, thereby improving data quality. However, when stimuli from online technology place greater cognitive demands on individuals, fatigue and reduced consistency in choice can decrease data quality.

Several theoretical papers on rational choice predict that increased choice set complexity adversely affects consistency in choice, which implies greater error variance in random utility models (Heiner, 1983; De Palma *et al*., 1994). However, relatively few empirical studies investigate how choice task complexity and the number of replications per respondent affect data quality, and whether these effects vary by survey mode.[1] Louviere *et al*. (2000) and Hensher *et al*. (1999) describe a range of methods to better understand and estimate the deterministic and error components of respondent utility (e.g., statistical design theory, use of economic theory to guide econometric specification, combination of SP and revealed preference data), but note there is little rigorous empirical research to assist researchers in choosing the number of replications per respondent. Bradley and Daly (1994) find fatigue effects in SP choice experiments. Swait and Adamowicz (1996) use data from six different SP studies to show there is a level of choice task complexity where variance of stochastic utility is minimized, and a cumulative cognitive load after which the variance grows.[2] In contrast, Brazell *et al*. (1995) suggest fatigue effects may be minimal and in some cases learning occurs as respondents are exposed to more replications. Further, Brazell and Louviere (1997) show equivalent survey response rates and parameter estimates when they compare respondents answering 12, 24, 48 and 96 choice questions in a particular choice task. Carson *et al*. (1994) review a range of choice experiments and find respondents are typically asked to evaluate from one to 16 choice questions, with the average being around eight questions per respondent. Adjustments from the average account for choice task complexity (i.e., number of attributes and alternatives), incentives, mode elicitation and types of respondents. Louviere *et al*. (2000) argue that when the survey contains a number of questions that might burden the respondent

[1] A small but growing social science literature examines response rates and response quality in opinion-based online surveys. Burr *et al*. (2001) study clients of a Federal agency and find they report more satisfaction with services when responding by web vs. mail. Carini *et al*. (2003) study college experiences for first-year and senior students. They find relatively small mode effects except for items related to computing and information technology, which exhibit moderately more favorable responses when answered online. Berrens *et al*. (2003) show statistically significant but modest differences in political opinion questions across telephone and online surveys. VanBeselaere (2002) examines matched telephone and online survey data from the *Internet Survey of American Opinion* and finds that Internet respondents are not more likely to 'shirk' (i.e., short-cut their cognitive processes and provide responses that are not carefully thought out) than telephone respondents. Cameron *et al*. (2002) find no significant differences in marginal utilities from mail and telephone respondents for an environmental good.

[2] Cognitive load is typically defined by entropy and cumulative entropy. These measures are included in choice models to account for difficulty in choosing between alternatives in the current question, and the cumulative effort of answering previous choice questions (Swait and Adamowicz, 1996; Savage and Waldman, 2006).

besides the choice task itself, it may be prudent to reduce the choice task from eight to four choice questions to improve response rates and limit fatigue. More generally, several marketing and psychology studies find that the precision of respondents' choices declines moderately with repeated choice tasks because they become fatigued (Elrod *et al.*, 1992; McClelland *et al.*, 1993).[3]

This is the first study to empirically investigate the effect of survey mode on respondent learning and fatigue during repeated choice experiments. This investigation is done in a controlled setting where the (nearly) identical survey is administered to two samples. We then develop new econometric methods to estimate the parameters of the random utility models that determine the choices. SP data are obtained from a choice experiment concerning high-speed Internet service conducted on 357 mail respondents in the period September–October 2002 and 325 online respondents in the period January–February 2003, respectively. Sixty-four paired descriptions of Internet service are grouped into eight sets of eight questions and randomly distributed across all respondents. Since the same experimental design is applied to both survey modes, the study controls for factors such as context and cognitive load which may otherwise confound the analysis, and solely focuses on differences in learning and fatigue attributed to mode effects. Further, the study identifies and estimates the ratio of error standard deviations (i.e., variation in the unexplained component of utility) for mail to online respondents, and the ratio of error standard deviations for the first four questions to the last four questions of the eight question choice set.

Hypothesis tests reveal that error standard deviations are about equal in the first four choice occasions. However, online respondents suffer fatigue or boredom as they progress from questions one through four to questions five through eight of the choice task, while mail respondents are unaffected. The error standard deviation for mail respondents answering choice questions five through eight is about three-quarters the error standard deviation for online respondents, while the error standard deviations are about equal for the first four choice occasions. These results suggest that while the online mode lowers survey administration costs and reduces time between survey implementation and data analysis, these benefits may come at the cost of greater error standard deviation in estimation utility. Researchers administrating choice experiments through online surveys may want to consider increasing the number of observations but require fewer choice questions per respondent.

The paper is organized as follows. Section 2 describes the choice experiment and administration of the survey instrument via online and mail survey modes. Section 3 outlines the economic model and econometric methods used to estimate consumer preferences while allowing for different error standard deviation ratios across survey modes and choice questions. Parameter estimates and tests for equal error standard deviations across survey modes and questions are presented in Section 4. In Section 5 we discuss the implications of our results for survey research, and Section 6 provides concluding remarks.

2. EXPERIMENTAL DESIGN

We use an experimental approach to assess response behavior and data quality in household surveys. SP data are obtained from an Internet access choice experiment administered by mail

[3] In the experimental economics literature, Romeo and Sopher (1999) arugue that learning takes place in complex decision making experiments, and the level of the payoff can affect the speed of learning. When the subject has figured out the optimal choice, then a level of payoff is needed to outweigh other factors in the utility function (such as the desire to avoid boredom).

DOI: 10.1002/jae

and online survey modes. These data permit estimation of the error variance in the unexplained component of the household's utility function. With the notion that larger error variance is implied by fatigue so that poorer quality data results, while smaller variance is implied by learning (one's preferences) so that better data quality results, we conduct a test of whether respondents fatigue or learn as they move through repeated choice questions, and whether this effect varies between survey modes. Below, we discuss the design of the underlying Internet access choice experiment, the protocols used to draw samples for mail and online survey respondents, and the differences between survey instruments and sample demographics.

2.1. Internet Access Choice Experiment

The Internet access choice experiment was administered by household survey questionnaire. The questionnaire begins with questions that inform respondents about Internet access service and its attributes. This is followed by the choice tasks. The questionnaire concludes with a set of demographic questions.

In the cognitive build-up section respondents are asked 23 questions about their current access and use of information technology, and they are provided with information to form preferences about five Internet access attributes. Internet access is *always on* when no dial-up is required for Internet connection, and respondents can use the Internet and place telephone calls at the same time, or *not always on*. Cost is the fixed monthly price for access with unlimited usage ($10–85/month in multiples of $5). Speed is the time it takes to receive and send information to and from the home computer. Speed is described to respondents as either *very fast*, *fast*, or *slow* (the slowest speed is described as the same as a dial-up connection). Ease of installation reflects the time, cost, and complexity of ordering and installing a new Internet connection. Installation of an Internet service can be *immediate, within one week*, and *within several weeks. Very reliable* Internet access is never disrupted (i.e., there are no service outages); however, with *less reliable* Internet access users may occasionally experience slower speed, outages that require customer support, and their account may be transferred from one company to another. In addition to informing them about attributes, respondents are asked to state the attribute levels of their current service.

For the choice task, each respondent is presented with eight questions that describe a pair of Internet access options, A and B, that differ by five attributes.[4] Respondents indicate their preferred choice. They then indicate whether they would switch to the service they had selected if they were already online, or if they would adopt the service selected if they were not. Finally, respondents answered nine questions about their age, employment status, household size, education level, wage rate, gender, location, income, and race. The Appendix displays a choice question example.

2.2. Protocols for Drawing Online and Mail Samples

In an ideal experimental setting, a sample of the target population are randomly assigned to the treatment groups (here, either the mail or online survey mode) and the investigator observes the effect of the treatment on one or more variables of interest. However, implementation of perfect randomization is costly in this context. An alternative, lower-cost strategy involves

[4] Measures developed by Zwerina *et al*. (1996) and Huber and Zwerina (1996) generate an efficient non-linear optimal design. A fractional factorial design creates 64 paired descriptions of Internet service that are grouped into eight sets of eight questions, and randomly distributed across all respondents.

DOI: 10.1002/jae

drawing separate random samples for mail and online survey modes and merging the two surveys. Successful implementation of this strategy requires the survey protocols to be as similar as possible between online and mail survey modes.

PA Consulting Group administered the mail survey. The gross sample consisted of 1240 addresses. These were randomly drawn from a database of households collected from all US white-page telephone directories and proprietary sources of non-listed telephone households. Advance postcards were mailed on 3 September 2002 informing households that they would receive a survey in the mail in a few days, and requesting their cooperation in completing the survey. The initial package with questionnaire, a $2 bill as an incentive, and postage-paid return envelope were mailed on 6 September. Of the 1240 surveys sent out, 54 were undeliverable and four were deceased, reducing the gross sample to 1182. By the end of October 2002, 397 completed questionnaires were obtained with a unit response rate of 33.6%. Of these 397 completed questionnaires, 357 respondents answered all eight Internet access choice questions and provided information on their education and Internet experience, for an item response rate of 90.0%. The median completion time for each mail questionnaire was about 20 minutes.

Knowledge Networks Inc. (KN) administered the online survey. KN panel members are drawn by random digit dialing of listed and unlisted telephone households, with a success rate of about 45–50%. To ensure consistent delivery of online survey content, each panel member is provided with identical online hardware even if they currently have Internet access. Microsoft's WebTV is the hardware, consisting of a set-top box that connects to the TV and the telephone, a remote keyboard, and a pointing device.[5] For incentive, panel members are rewarded with points for participating in surveys, which can be converted to cash or used to pay the monthly Internet access charge.[6] KN contacted a gross sample of 799 panel members on 24 January 2003 informing them about the Internet access choice experiment. By 12 February 2003, 575 completed questionnaires were obtained with a effective unit response rate of 32.4–36.0% (i.e., 575/799 × 45–50%). 209 of the 575 questionnaires were excluded by us from this analysis because they had been randomly assigned an additional Internet access attribute as part of another study. Of the 366 completed questionnaires remaining for use in this study, 325 respondents answered all eight Internet access choice questions and provided information on their education and Internet experience, for an item response rate of 88.8%. The median completion time for each mail questionnaire was about 19 minutes.

2.3. Differences between Survey Questionnaires

The two survey questionnaires differ not only in the mode of data collection, but also in the demographic information solicited. Because KN maintains a database of panel demographics, online respondents answer only three demographic questions when concluding the survey, while mail respondents answer nine questions. Since these questions were posed *after* the choice questions, they should not result in any differences in response behavior between survey modes.

[5] WebTV has a 56K modem that connects the household to the Internet. The unit also has a hard drive to accommodate small downloads.

[6] Both online and mail respondents receive survey-specific incentives. As such, observed differences in response behavior (particularly, the distributions of error terms in the econometric analysis) are less likely to be confounded by differences in respondent's motivation. See Philipson (2001) and Philipson and Malani (1999) for discussion of how survey data quality is influenced by incentives and respondents' motivation.

DOI: 10.1002/jae

The two questionnaires also differ in the way they permit respondents to access information on the Internet access attribute descriptions. Because mail respondents have a hard copy, they can easily locate and refer to the page of attribute descriptions as they complete their choice task. Online respondents, however, may face physical and/or technical constraints which prevent them from seamlessly scrolling backward and forward from the attribute descriptions page to choice task pages. To better replicate the conditions facing mail respondents, online respondents can access attribute descriptions through a look-up feature by clicking on an attribute's underlined name in the choice question. When the respondent is finished viewing the attribute description, she clicks an icon that automatically returns her to the choice question screen she was viewing prior to accessing the additional information. This feature permits online respondents to move back and forward through screens to refresh their understanding of attributes in a similar manner to turning pages in the mail survey. Moreover, direct access to attribute descriptions simulates mail respondents' ability to copy or cut out the descriptions, and place them directly in front of them as they complete their choice questions.[7]

As an aside to this study we also use the look-up feature to analyze how sensitive respondent preferences are to the wording of attribute descriptions. More specifically, language is included in the instantaneous attribute descriptions of ALWAYS ON and INSTALLATION which may make them less attractive to online respondents relative to mail respondents.[8] Potential differences in the valuation of attributes are controlled for in Section 4 by estimating a utility function for the combined online plus mail sample that permits marginal utility parameters for ALWAYS ON and INSTALLATION to vary by online and mail respondent.

2.4. Sample Demographics

To be confident in the internal validity of our experimental design it may be important to account for any systematic variation in important demographic variables between mail and online samples. Table I presents a selection of demographics for the mail and web samples, along with similar data from the US Census Bureau (2003). Both samples are similar in their geographic coverage and respondents' employment status. The mail sample covers 44 states across the four main census regions and 68.9% of mail respondents were employed last month (at a location outside of the home). The online sample covers 46 states plus the District of Columbia across the four census regions and 66.8% of online respondents were employed last month. There are noticeable differences, however, in age, race, gender, education, income, and years online between samples. In general, mail respondents are older, a higher percentage white and male, more educated, wealthier, and have more Internet experience than the typical online respondent. To rule out the possibility that any survey mode effect found in the data is confounded by demographic-specific response behavior, we construct post-stratification weights for the econometric analysis in Section 4. The demographic

[7] The net effect of the look-up feature is not clear. One could argue that look-up options should lead to more focus, less fatigue, and a relative improvement in data quality. Alternatively, an anonymous referee notes the look-up feature can also create a demand effect; i.e., respondents may follow up those links more often than they really need to. This may lead to distraction and fatigue and a reduction in data quality. In either case, this discussion may be irrelevant for our analysis: only 6.3% of online respondents actually accessed the descriptions when answering the choice questions.

[8] A sentence is added to the ALWAYS ON description stating '... files on your home computer may be more exposed to "hacking" from outside intruders but you can make your computer reasonably secure by purchasing "firewall" software for about $35.' For INSTALLATION, the additional sentence is '... a technician from the phone or cable company may have to come out to your home once or several times.'

DOI: 10.1002/jae

variables used for post-stratification weighting are those that had significant differences: age, race, gender, education, income, and years online.

3. ECONOMETRIC METHODOLOGY

The data were collected by survey, where respondents answer a series of eight choice questions. In each choice occasion a pair of Internet access options is presented, labeled 'A' and 'B' in the survey (see Appendix, indicated in the analysis below by choice 1 and choice 2). The two access options differed by five attributes. Respondents indicate their preference for choice A or choice B. In addition, we enriched the information in these choices by having respondents indicate whether they would stay with their current Internet service (the 'status quo'), or switch to the service they had just selected, if they were already online, or if they would adopt the service selected if they were not already online. The parameters of the representative individual's utility function (the marginal utilities of the five attributes) are then estimated from the observed choices.

3.1. Random Utility Model and Likelihood

It is assumed that respondents maximize their household's conditional utility of the service option (conditional on all other consumption and time allocation decisions):

$$U_{ij}^{k_{ij}} = \boldsymbol{\beta}'\boldsymbol{x}_{ij}^{k_{ij}} + \varepsilon_{ij}^{k_{ij}}, i = 1, \ldots, n; j = 1, \ldots J, k_{ij} = 1, 2 \tag{1}$$

where $U_{ij}^{k_{ij}}$ is the utility of alternative k_{ij} chosen by individual i during occasion j.[9] The vector $\boldsymbol{x}_{ij}$ contains the observed attributes of the alternatives. It is assumed that the $\varepsilon_{ij}^{k_{ij}}$ are independent, and identically distributed mean zero normal random variables, uncorrelated with $\boldsymbol{x}_{ij}$, with constant unknown variance σ_ε^2.[10] The probability of choosing alternative 1, for example, is

$$\begin{aligned} P_{ij}^1 &= P(U^1 > U^2) \\ &= P(\boldsymbol{\beta}'\boldsymbol{x}_{ij}^1 + \varepsilon_{ij}^1 > \boldsymbol{\beta}'\boldsymbol{x}_{ij}^2 + \varepsilon_{ij}^2) \\ &= P\left(\varepsilon_{ij}^2 - \varepsilon_{ij}^1 < -\boldsymbol{\beta}'(\boldsymbol{x}_{ij}^2 - \boldsymbol{x}_{ij}^1)\right) \\ &= \Phi\left[-\boldsymbol{\beta}'(\boldsymbol{x}_{ij}^2 - \boldsymbol{x}_{ij}^1)/\sqrt{2}\sigma_\varepsilon\right] \end{aligned} \tag{2}$$

and similarly for alternative 2, where $\sqrt{2}\sigma_\varepsilon$ is the standard deviation of $\varepsilon_{ij}^2 - \varepsilon_{ij}^1$ and $\Phi(\cdot)$ is the univariate standard normal cumulative distribution function. Note that equation (2) comprises the

[9] This notation, especially the use of k_{ij} to indicate either a 1 or a 2, is a bit cumbersome at first, but will make precise many of the concepts below.

[10] We allow for correlation of errors for an individual when it comes to choices involving the status quo—see Section 3.2. For the hypothetical choices, there is no question of correlation since the effective errors that enter the likelihood are the *difference* in the two errors for any choice occasion, and the attribute sets are randomly assigned to choice 'A' or choice 'B'. That is, the relevant distribution theory for forming the likelihood is based on $\varepsilon_{i1}^1 - \varepsilon_{i1}^2$, for example (person i, first choice occasion—see equation (7)). In addition, any additive systematic component of the error is then eliminated. This is similar to the arguments of Heckman and Robb (1985) in their evaluation of social interventions.

Table I. Selected sample and population characteristics (%)

	US Census	Online sample	Mail sample	
Census region				
Northeast	19.1	15.3	16.5	
Midwest	22.9	28.1	34.6	
South	35.6	39.1	33.6	
West	22.4	17.5	15.2	
Sample size: *t*-statistic	—	366	381	1.62
Age				
18–24 years	13.3	12.3	3.9	
25–34 years	18.1	8.5	15.0	
35–44 years	21.8	23.7	19.4	
45–54 years	18.9	23.5	23.9	
55–64 years	11.9	19.4	20.7	
65 years and over	16.1	12.6	17.1	
Sample size: *t*-statistic	—	366	381	2.49***
Race				
Black	11.9	7.4	4.4	
White	83.2	79.2	91.2	
Other	4.9	6.3	2.9	
Sample size: *t*-statistic	—	366	385	4.87***
Gender				
Male	48.0	45.9	55.9	
Female	52.0	54.1	44.1	
Sample size: *t*-statistic	—	366	392	2.75***
Education				
Less than high school	15.8	18.9	4.6	
High school	33.0	24.6	16.2	
Some college	19.3	21.9	27.4	
Associate degree	7.8	9.0	11.8	
Bachelors degree and beyond	24.1	25.7	40.0	
Sample size: *t*-statistic	—	366	390	6.87***
Household Income				
Under $10,000	7.4	4.1	2.6	
$10,000–$24,999	18.4	11.8	8.3	
$25,000–$49,999	28.5	40.7	21.5	
$50,000–$74,999	20.0	24.6	29.7	
$75,000 or more	25.7	18.9	38.3	
Sample size: *t*-statistic	—	366	312	6.10***
Years online				
Fewer than 1 year	n.a.	4.1	4.6	
1–3 years	n.a.	36.0	19.3	
3–5 years	n.a.	25.8	23.9	
More than 5 years	n.a.	34.1	42.6	
Never been online	n.a.	0	9.6	
Sample size: *t*-statistic	—	364	394	6.12***
Employment status				
In labor force	65.3	66.8	68.9	
Not in labor force	34.7	33.2	31.1	
Sample size: *t*-statistic	—	365	351	0.60

Note: *t*-tests are for differences between means for the online and mail samples. *** indicates a significant difference between means at the 1% level. Years online is 'years using the Internet to go online at home, school, work, and other locations'.

usual probit model for dichotomous choice under the assumption the individual knows the random component and maximizes utility. The parameter vector β is identified only up to the scale factor $\sqrt{2}\sigma_{\in}$, and $\sigma_{\in}$ is not identified, since only the sign and not the scale of the dependent variable (the utility difference) is observed. If the J observations for each respondent are simply 'stacked' to produce a data set with Jn observations, the unit of observation is an i, j pair and the likelihood is the product of the Jn probabilities like equation (2):

$$L(k_{ij}, i = 1, \ldots, n, j = 1, \ldots, J | \boldsymbol{x}_{ij}^1, \boldsymbol{x}_{ij}^2; \beta, \sigma_{\in}) = \prod_{i=1}^{n}\prod_{j=1}^{J} P_{ij}^{k_{ij}} \tag{3}$$

3.2. Incorporating the Status Quo Question

After choosing k_{ij}, individuals answer a question stating whether alternative k_{ij} would be chosen over the status quo. Let the status quo be indicated by 0. There are now four kinds of observations. Let the binary variable Z_{ij}^1 indicate the choice of alternative 1 or 2 for individual i on occasion j, and let the binary variable Z_{ij}^2 indicate the chosen alternative or the status quo. These are defined by

$$Z_{ij}^1 = \begin{cases} 0 & \text{choose 1} \\ 1 & \text{choose 2} \end{cases} \quad Z_{ij}^2 = \begin{cases} 0 & \text{choose 1 or 2 over status quo} \\ 1 & \text{choose status quo over 1 or 2} \end{cases} \tag{4}$$

Note that there is an information asymmetry here: when the status quo is chosen over 1 or 2 ($Z_{ij}^2 = 1$), a complete ranking of the three alternatives has been determined; when 1 or 2 is chosen over the status quo ($Z_{ij}^2 = 0$), all that is known is that 1 or 2 is the most preferred alternative.

Utility for the status quo, U_i^0 under the model assumption (equation (1)) is given by

$$U_i^0 = \boldsymbol{\beta}'\boldsymbol{x}_i^0 + \in_i^0 \tag{5}$$

where $\in_i^0$ are disturbances and $\boldsymbol{x}^0$ are the attributes of the individual's current Internet access. The attributes of the status quo vary over individuals, but not over choice occasions, and the utility of the status quo is evaluated only once by each individual (U_i^0 and $\in_i^0$ are subscripted with i only). The $\in_i^0$ are assumed to be independent, identically distributed normal random variables with zero expectation and variance σ_0^2, uncorrelated with $\in_{ij}^{k_{ij}}$.

The probability of choosing alternative k_{ij} (1, 2) over alternative $3 - k_{ij}$ (2, 1) and then choosing alternative k_{ij} over the status quo ($Z_{ij}^2 = 0$) is the bivariate probability:

$$\begin{aligned} &P(U_{ij}^{k_{ij}} > U_{ij}^{3-k_{ij}}, U_{ij}^{k_{ij}} > U_i^0) \\ &= P\left(\in_{ij}^{3-k_{ij}} - \in_{ij}^{k_{ij}} < -\boldsymbol{\beta}'(\boldsymbol{x}_{ij}^{3-k_{ij}} - \boldsymbol{x}_{ij}^{k_{ij}}), \in_i^0 - \in_{ij}^{k_{ij}} < -\boldsymbol{\beta}'(\boldsymbol{x}^0 - \boldsymbol{x}_{ij}^{k_{ij}})\right) \\ &= \Phi_2[-\boldsymbol{\beta}'(\boldsymbol{x}_{ij}^{3-k_{ij}} - \boldsymbol{x}_{ij}^{k_{ij}})/\sqrt{2}\sigma_{\in}, -\boldsymbol{\beta}'(\boldsymbol{x}^0 - \boldsymbol{x}_{ij}^{k_{ij}})/\sqrt{\sigma_0^2 + \sigma_{\in}^2}; \rho] \end{aligned} \tag{6}$$

where ρ is the correlation between $\in_{ij}^{3-k_{ij}} - \in_{ij}^{k_{ij}}$ and $\in_i^0 - \in_{ij}^{k_{ij}}$:

$$\rho = \frac{\sigma_{\in}^2}{\sqrt{2\sigma_{\in}^2(\sigma_0^2 + \sigma_{\in}^2)}} = \frac{\sigma_{\in}}{\sqrt{2(\sigma_0^2 + \sigma_{\in}^2)}} \tag{7}$$

DOI: 10.1002/jae

and Φ_2 is the standard bivariate normal cumulative distribution function. Similarly, the probability of choosing alternative k_{ij} over alternative $3 - k_{ij}$ and then choosing the status quo over alternative $k_{ij}(Z_{ij}^2 = 1)$ is

$$\begin{aligned} & P(U_{ij}^{k_{ij}} > U_{ij}^{3-k_{ij}}, U_{ij}^{k_{ij}} < U_i^0) \qquad (8)\\ &= P(\in_{ij}^{3-k_{ij}} - \in_{ij}^{k_{ij}} < -\boldsymbol{\beta}'(\boldsymbol{x}_{ij}^{3-k_{ij}} - \boldsymbol{x}_{ij}^{k_{ij}}), \in_i^0 - \in_{ij}^{k_{ij}} > -\boldsymbol{\beta}'(\boldsymbol{x}^0 - \boldsymbol{x}_{ij}^{k_{ij}}))\\ &= \Phi_2[-\boldsymbol{\beta}'(\boldsymbol{x}_{ij}^{3-k_{ij}} - \boldsymbol{x}_{ij}^{k_{ij}})/\sqrt{2}\sigma_\in, \boldsymbol{\beta}'(\boldsymbol{x}^0 - \boldsymbol{x}_{ij}^{k_{ij}})/\sqrt{\sigma_0^2 + \sigma_\in^2}; -\rho] \end{aligned}$$

where the symmetry of the normal distribution has been utilized.

One normalization is required: let $\sigma_\in = 1/\sqrt{2}$. Define $\lambda^2 = \sigma_0^2/\sigma_\in^2 = 2\sigma_0^2$. Then equation (8) can be written as

$$\begin{aligned} & P(U_{ij}^{k_{ij}} > U_{ij}^{3-k_{ij}}, U_{ij}^{k_{ij}} < U_i^0) = \qquad (8')\\ & \Phi_2\left[-\boldsymbol{\beta}'(\boldsymbol{x}_{ij}^{3-k_{ij}} - \boldsymbol{x}_{ij}^{k_{ij}}), \frac{\boldsymbol{\beta}'(\boldsymbol{x}^0 - \boldsymbol{x}_{ij}^{k_{ij}})}{\sqrt{(1+\lambda^2)/2}}; -\frac{1}{\sqrt{2(1+\lambda^2)}}\right] \end{aligned}$$

and similarly for equation (6). The additional parameter to be estimated is λ. When $\lambda = 1$, $\sigma_\in^2 = \sigma_0^2$ and the A versus B question and the question comparing A or B to the status quo have equal weight in the likelihood. When $\lambda < 1$ the question relating to the status quo contains more information, as there is more variability in the errors for the A vs. B question ($\sigma_\in^2 > \sigma_0^2$), and conversely. Let $\boldsymbol{x}_{ij}^{rp} = (\boldsymbol{x}_{ij}^r - \boldsymbol{x}_{ij}^p)$ for $r, p = 0, 1$. Then the probabilities of the four data types are

$$\begin{aligned} P(Z_{ij}^1 = 0, Z_{ij}^2 = 0) &= \Phi_2\left[-\boldsymbol{\beta}'\boldsymbol{x}_{ij}^{21}, -\boldsymbol{\beta}'\boldsymbol{x}_{ij}^{01}/\lambda; \frac{1}{2\lambda}\right]\\ P(Z_{ij}^1 = 0, Z_{ij}^2 = 1) &= \Phi_2\left[-\boldsymbol{\beta}'\boldsymbol{x}_{ij}^{21}, \boldsymbol{\beta}'\boldsymbol{x}_{ij}^{01}/\lambda; -\frac{1}{2\lambda}\right]\\ P(Z_{ij}^1 = 1, Z_{ij}^2 = 0) &= \Phi_2\left[\boldsymbol{\beta}'\boldsymbol{x}_{ij}^{21}, -\boldsymbol{\beta}'\boldsymbol{x}_{ij}^{02}/\lambda; \frac{1}{2\lambda}\right]\\ P(Z_{ij}^1 = 1, Z_{ij}^2 = 1) &= \Phi_2\left[\boldsymbol{\beta}'\boldsymbol{x}_{ij}^{21}, \boldsymbol{\beta}'\boldsymbol{x}_{ij}^{02}/\lambda; -\frac{1}{2\lambda}\right] \qquad (9) \end{aligned}$$

The likelihood is the product of these Jn probabilities:

$$L(Z_{ij}^1, Z_{ij}^2, i = 1, \ldots, n, j = 1, \ldots, J|\boldsymbol{x}_{ij}^1, \boldsymbol{x}_{ij}^2, \boldsymbol{x}^0; \boldsymbol{\beta}, \lambda) = \prod_{i=1}^{n}\prod_{j=1}^{J} P(Z_{ij}^1, Z_{ij}^2) \qquad (10)$$

which, upon substitution of equations (9) can be written

$$\begin{aligned} & L(Z_{ij}^1, Z_{ij}^2, i = 1, \ldots, n, j = 1, \ldots, J|\boldsymbol{x}_{ij}^1, \boldsymbol{x}_{ij}^2, \boldsymbol{x}^0; \boldsymbol{\beta}, \lambda) = \qquad (11)\\ & \prod_{i=1}^{n}\prod_{j=1}^{J} \Phi_2\left\{(-1)^{1-Z_{ij}^1}\boldsymbol{\beta}'\boldsymbol{x}_{ij}^{21}, (-1)^{1-Z_{ij}^2}[(1 - Z_{ij}^1)\boldsymbol{\beta}'\boldsymbol{x}_{ij}^{01} + Z_{ij}^1\boldsymbol{\beta}'\boldsymbol{x}_{ij}^{02}]/\lambda; (-1)^{Z_{ij}^2}\frac{1}{\lambda}\right\} \end{aligned}$$

The likelihood as it is written in equation (11) does not take into consideration the fact that the formation of that part of the likelihood involving the comparison of the chosen alternative to the status quo involves the error difference $\varepsilon_i^0 - \varepsilon_{ij}^{k_{ij}}$, where $k_{ij} = 1$ or 2 (depending upon the choice), and from choice occasion to choice occasion these error differences are correlated. This correlation is induced by the common occurrence of ε_i^0, since respondents need evaluate their utility of the status quo only once. This point is generally missed in conjoint analysis. An econometric innovation of this study is to treat the person, and not the person–choice occasion, as the unit of observation, so that we may explicitly model this correlation. The likelihood is now written

$$L(Z_{ij}^1, Z_{ij}^2, i = 1, \ldots, n, j = 1, \ldots, J | \boldsymbol{x}_{ij}^1, \boldsymbol{x}_{ij}^2, \boldsymbol{x}^0; \boldsymbol{\beta}, \lambda) = \prod_{i=1}^{n} P(Z_{i1}^1, Z_{i1}^2, Z_{i2}^1, Z_{i2}^2, \ldots, Z_{ij}^1, Z_{ij}^2) \tag{12}$$

The probability in equation (12) would appear to be computationally intractable, as it involves a 16-fold ($2 \times J = 8$) integration of the multivariate normal density function. Fortunately, this is not the case, as the correlation between $\varepsilon_i^0 - \varepsilon_{ij}^1$ and $\varepsilon_i^0 - \varepsilon_{ij}^2$, for example, is a result of the common occurrence of ε_i^0. This means that we can follow a familiar conditioning argument to express the probability in equation (12) as the integral of the product of eight bivariate probabilities, integrated against the univariate normal density (see Waldman, 1985). But the cost of this generality is in programming and computer time, as the likelihood must be maximized by simulation or with quadrature methods. We used Hermite polynomial quadrature (Abramowitz and Stegun, 1964, p. 890).

3.3. Two Sources of Data

Now consider two subsets of the data, e.g., the first four and the last four questions, or the online and mail surveys. Suppose that the error in one subset is ε and in the other ε'. We can allow for the possibility that the variance in the choice questions is different, due to the different position (questions 1–4 or 5–8) or presentation (mail survey or online survey) of the choice questions. Define

$$\gamma^2 = \frac{\sigma_\varepsilon^2}{\sigma_{\varepsilon'}^2} \tag{13}$$

where $\sigma_{\varepsilon'}^2 = V(\varepsilon')$. For observations in the second data set, σ_ε^2 in equations (6)–(8) is replaced by $\sigma_{\varepsilon'}^2$. Using the definition from equation (13) and the normalization we have

$$\sigma_{\varepsilon'}^2 = \frac{\sigma_\varepsilon^2}{\gamma^2} = \frac{1}{2\gamma^2} \tag{14}$$

so that equation (8) can be written

$$P(U_{ij}^{k_{ij}} > U_{ij}^{3-k_{ij}}, U_{ij}^{k_{ij}} < U_i^0) = \Phi_2 \left[\frac{-\boldsymbol{\beta}'(\boldsymbol{x}_{ij}^{3-k_{ij}} - \boldsymbol{x}_{ij}^{k_{ij}})}{1/\gamma}, \frac{\boldsymbol{\beta}'(\boldsymbol{x}^0 - \boldsymbol{x}_{ij}^{k_{ij}})}{\sqrt{(1/\gamma^2 + \lambda^2)/2}}; -\frac{1}{\sqrt{2(1 + \gamma^2\lambda^2)}} \right] \tag{15}$$

DOI: 10.1002/jae

and similarly for equation (6).

The question of the identification of γ requires elaboration, and we consider a simple analogy to our split sample to motivate the discussion. Suppose there are two, independent random variables: $Y_1^* \sim N(\mu_1, \sigma_1^2)$, and $Y_2^* \sim N(\mu_2, \sigma_2^2)$. With observations y_1^* and y_2^* on Y_1^* and Y_2^*, all parameters and hence $\gamma = \sigma_1/\sigma_2$ are estimable. But suppose the data are censored, so that all that is available is

$$y_j = \begin{cases} 0 & Y_{1j}^* < 0 \\ 1 & Y_{1j}^* \geq 0 \end{cases} \qquad j = 1, 2$$

Estimation from either sample (alone) produces estimates of only the ratios $\widehat{\mu_j/\sigma_j} = n_{1j}/n_j$, where n_{1j} equals the number of times $y_j = 1$, and n_j is the sample size. If it is assumed that $\mu_1 = \mu_2 = \mu$ (say), the two sample moments n_{11}/n_1 and n_{12}/n_2 produce estimates of μ/σ_1 and μ/σ_2, and the ratio

$$\gamma = \frac{\sigma_1}{\sigma_2} = \frac{\mu/\sigma_2}{\mu/\sigma_1}$$

may be estimated by

$$\widehat{\gamma} = \frac{\widehat{\mu/\sigma_2}}{\widehat{\mu/\sigma_1}}$$

To continue the example, suppose the simple regression model applies so that there exists a set of constants x such that

$$Y_j^* \sim N(\mu_j + \beta_j x, \sigma_j^2)$$

The additional moments, $\sum y_i x$, $j = 1, 2$, serve to identify the β_j. Again, one normalization is required to identify γ, logically either $\mu_1 = \mu_2$ or $\beta_1 = \beta_2$. If both constraints are imposed, the model is overidentified, and the overidentifying restrictions can be tested with a likelihood ratio. In the empirical work below we consider various combinations of sensible parameter restrictions to identify γ.

The estimation and interpretation of γ is the primary focus of this paper. Estimates are used to test the equality of the standard deviations between the mail and online survey modes, and to test the equality of the standard deviations between question position one through four and question position five through eight.

4. EMPIRICAL RESULTS

4.1. Introduction

Data were obtained for 325 online and 357 mail respondents who answered all eight choice questions, the follow-up status quo question regarding the desirability of their current Internet service, and who provided demographic data for number of years online (*YEARS*) and education level (*EDUC*). Since each pair of binary choices for each choice occasion represents information on preferences, the size of the combined mail plus online sample is $nJ = (325 + 357) \times 8 = 5456$. In this section we use these data to estimate the marginal utilities of Internet service attributes, the βs, and the resulting willingness-to-pay for improvements in these attributes. We will also estimate λ, the ratio of the standard deviation of the errors in evaluating the status quo service to the errors in evaluating the hypothetical services.

4.2. Marginal Utility Parameter Estimates and Willingness-to-Pay

In Table II we report maximum likelihood estimates of the parameters of the random utility model, for various specifications of the model and subsamples of the data. Models (i) and (ii) contain the results for the mail and online samples, respectively. The observed attributes $\boldsymbol{x}_{ij}$ are the five access attributes, the interaction between SPEED and *YEARS*, and the interaction between SPEED and *EDUC*.[11]

Estimated coefficients for the attributes of Internet access generally have the expected sign and are statistically significant.[12] Coefficients also suggest plausible valuations of Internet attributes.

Table II. Parameter estimates

Parameter group	Model: Sample:	(i) Mail	(ii) Online	(iii) Combined	(iv)
Service attributes					
ALWAYS ON		−0.105	0.010	−0.050	−0.096
		(−5.218)	(0.620)	(−3.330)	(−5.564)
ALWAYS ON (online)		—	—	—	−0.007
		—	—	—	(23.38)
SPEED		−0.037	−0.012	0.025	0.056
		(−0.997)	(−0.531)	(1.075)	(4.069)
COST		−0.018	−0.016	−0.017	−0.017
		(−17.95)	(−13.39)	(−20.36)	(−23.38)
INSTALLATION		0.040	−0.016	0.028	0.060
		(2.385)	(−0.873)	(3.042)	(4.346)
INSTALLATION (online)		—	—	—	−0.010
		—	—	—	(−2.434)
RELIABILITY		−0.283	−0.390	−0.319	−0.325
		(−12.99)	(−15.15)	(−18.21)	(−21.93)
Interactions					
SPEED × *YEARS*		−0.029	−0.049	−0.043	−0.055
		(−4.510)	(−4.629)	(−6.687)	(−7.592)
SPEED × *EDUC*		−0.015	0.002	−0.015	−0.012
		(−2.600)	(0.324)	(−3.742)	(3.642)
Variance parameters					
λ		1.599	1.371	1.581	1.524
		(6.468)	(4.669)	(8.094)	(31.69)
Mean Log-L		−9.288	−8.895	−9.142	−9.110
n		357	325	682	682

Note: Asymptotic t-ratios in parentheses. For attributes, the test is $H_0 : \beta = 0$, and for λ and γ, the test is $H_0 : \lambda = 1$ or $H_0 : \gamma = 1$.

[11] In another manuscript we propose a theoretical model for the demand for Internet access which shows that Internet ability affects the demand for speed. These two variables are represented here by education and the number of years online. See Savage and Waldman (2006).

[12] The expected signs are negative, as increasing attribute levels imply a less desirable service. The coefficient of ALWAYS ON for the online subsample is not statistically significantly different from zero. This was not unexpected, as the instantaneous attribute description of ALWAYS ON, available to the online respondents only, highlights problems with

DOI: 10.1002/jae

For example, the willingness-to-pay for a service that is always on is estimated for the mail respondents to be about \$ 5.85. This is calculated by dividing the marginal utility of ALWAYS ON by the marginal utility of income, estimated by the coefficient on COST. The marginal utility of SPEED is the combination of its stand-alone coefficient and the interaction coefficients multiplied by the values of the interaction variables. For example, for the mail respondents, the marginal utility of SPEED is

$$\frac{\partial U}{\partial \text{SPEED}} = -0.037 - 0.029 \times YEARS - 0.015 \times EDUC$$

The willingness-to-pay for speed calculated at the mean values of *YEARS* and *EDUC* is about \$13.60 for online respondents and \$11.14 for mail respondents.

Recall the definition of the ratio of the variance of the errors in valuing the hypothetical choices to the variance of the error in evaluating the status quo: $\lambda^2 = \sigma_0^2/\sigma_\varepsilon^2$. Models (i) and (ii) provide estimates of λ that are greater than one for both the mail (1.60) and online (1.37) samples. The hypothesis that $\lambda = 1$ is rejected in both samples ($t = 6.47, 4.67$, respectively). All respondents appear to have more consistency in choice when comparing hypothetical choices than when comparing a hypothetical choice to a real alternative. A plausible explanation for this finding is that experience with an actual Internet service is more multidimensional than is portrayed by our set of attributes, while the hypothetical choice presented by the survey instrument highlights the attributes in a cleaner manner.[13]

In model (iii) mail and online data are combined. Marginal utility estimates for ALWAYS ON, COST, RELIABILITY, SPEED $\times$ *YEARS* and SPEED $\times$ *EDUC* have negative signs predicted by theory, while the estimated marginal utility of INSTALLATION is positive. A likelihood ratio test of the hypothesis that marginal utilities are equal across modes is rejected ($\chi^2_{(8)} = 56.3$).

Rejection of the hypothesis of equal marginal utility parameters is plausible when online respondents face slightly different attribute descriptions for some attributes. Accordingly, model (iv) reports estimates for the combined mail and online data with survey mode-specific variances *and* separate parameters for ALWAYS ON and INSTALLATION. Inspection of estimates suggests the same general results as in models (i), (ii) and (iii), but now the χ^2 statistic for the test that marginal utilities are equal across modes has fallen to a marginally insignificant 12.66.

4.3. Learning and Fatigue

In this section we make use of the econometric methodology in Section 3.3 for specifying separate variances for different subsets of the data to make statements about the extent of learning and fatigue in choice experiments. The primary focus will be to quantify the precision of the responses. By analogy to the linear regression model, where the standard error of an estimator is proportional to the standard deviation of the model error, we will measure the relative precision of responses from two sources of data or subsets of the data by the ratio of the standard deviations of their

hackers and has the potential to lower online respondent's valuation of this attribute. The coefficient of INSTALLATION is positive and significant in the mail sample, and negative and insignificant in the online sample. Similarly, the extra description for INSTALLATION focuses on the inconvenience of phone or cable installers making several visit to the home during the installation process and may decrease online respondents' valuation of INSTALLATION.

[13] The parameter λ is consistently estimated to be greater than one and the null hypothesis that $\lambda = 1$ is rejected in all models in Tables II and III. We report its estimate and the test statistic for testing $H_0 : \lambda = 1$, but do not discuss it further.

DOI: 10.1002/jae

errors. Note this is not comparing competing estimators. Instead, we are estimating the ratio of standard deviations either from two methods of eliciting information (mail and online survey modes), or from two sets of responses (first four and last four questions) in a particular survey. Thus we estimate the standard deviation ratio and test the null hypothesis that the ratio is equal to one.

The following notation for the ratios of standard deviations (the γ of equation (13)) is adopted:

- $\gamma_{4/8}^{M}$ for questions 1–4 to questions 5–8, mail respondents;
- $\gamma_{4/8}^{W}$ for questions 1–4 to questions 5–8, online respondents;
- $\gamma_{M/W}^{1-4}$ for mail respondents to online respondents, questions 1–4; and
- $\gamma_{M/W}^{5-8}$ for mail respondents to online respondents, questions 5–8.

The superscript denotes the sample ($M \Rightarrow$ mail, $W \Rightarrow$ web (online), 1–4 or 5–8 $\Rightarrow$ combined mail and online) and the subscript denotes the comparison ($4/8 \Rightarrow$ first four versus last four questions; $M/W \Rightarrow$ mail versus online). In the first two bullets the comparisons are within a mode. Estimates indicate whether respondents learn or tire in the course of answering a sequence of choice questions, for a specific survey mode. In the last two bullets the comparisons are between modes. Estimates here will provide information about which survey type elicits better information, for which sequence of questions.

To examine learning and fatigue, we specify different variances for different subsets of the data. Here we interpret the standard deviation of the error as the magnitude of the unexplained component of utility. Unequal standard deviations could result for several reasons, as discussed in the Introduction. If respondents learn about their preferences as they move through the question replications, they become more proficient at the choice task. The information in the attribute levels then determines their utility more precisely, resulting in a decrease in the standard deviation of the error. Alternatively, respondents could become tired as they move through the choice occasions, and make mistakes in their choice of service. Or they could become bored with the demands of the survey, and expend less mental energy picking their preferred alternative. Either fatigue or boredom would result in a larger role for the error in determining their choice, that is, a larger standard deviation of that error.

In initial estimation on the combined mail and online samples there is a deterioration in responses, with the ratio of standard deviation of the first four to the last four questions estimated to be 0.881, significantly different from one ($t = 2.28$). But combining the data may mask differing results for the two survey modes. Table III contains maximum likelihood estimates of γs unique to a survey mode. Model (v) displays results for mail respondents, and model (vi) for online respondents. The important hypothesis to be tested is that error standard deviations for the first four and last four questions are equal. The hypothesis of equal standard deviations,

$$H_0 : \gamma_{4/8} = \sqrt{\frac{\sigma_{1-4}^2}{\sigma_{5-8}^2}} = 1$$

is not rejected for the mail survey ($\widehat{\gamma}_{4/8}^{M} = 1.019; t = 0.23$), but *is* rejected for the online survey ($\widehat{\gamma}_{4/8}^{W} = 0.745; t = 3.54$). The estimated ratio of error standard deviations of 0.745 implies that online respondents suffer fatigue or boredom as they progress through the survey and have about

DOI: 10.1002/jae

Table III. Parameter estimates for mail and Internet by choice question

Parameter group	Model: Questions: Sample:	(v) Q1–8 Mail	(vi) Q1–8 Online	(vii) Q1–4 Both	(viii) Q5–8 Both	(ix) Q1–8 Both	(x) Q1–8 Both
Service attributes							
ALWAYS ON		−0.102	0.001	−0.090	−0.076	−0.054	−0.096
		(−5.111)	(0.055)	(−3.551)	(−2.916)	(−3.55)	(4.86)
ALWAYS ON (Online)		—	—	—	—	—	−0.009
		—	—	—	—	—	(1.30)
SPEED		0.094	0.023	0.118	−0.033	0.097	0.043
		(2.407)	(0.584)	(3.602)	(−0.528)	(2.83)	(1.40)
COST		−0.018	−0.018	−0.017	−0.022	−0.017	−0.017
		(−16.63)	(−10.90)	(−13.32)	(−12.15)	(−16.69)	(22.22)
INSTALLATION		0.044	−0.017	0.060	−0.040	0.032	0.051
		(2.640)	(−0.530)	(3.511)	(−2.149)	(2.57)	(3.50)
INSTALLATION (Online)		—	—	—	—	—	−0.026
		—	—	—	—	—	(1.48)
RELIABILITY		−0.287	−0.441	−0.325	−0.371	−0.319	−0.333
		(−12.84)	(−12.13)	(−11.02)	(−12.40)	(15.46)	(18.36)
Interactions							
SPEED × *YEARS*		−0.054	−0.065	−0.062	−0.055	−0.060	−0.054
		(−5.210)	(−4.773)	(−4.919)	(−3.606)	(6.30)	(6.07)
SPEED × *EDUC*		−0.023	0.002	−0.011	−0.024	−0.017	−0.012
		(3.81)	(0.33)	(1.88)	(3.02)	(3.64)	(2.39)
Variance parameters							
λ		1.569	1.584	1.691	2.056	1.588	1.653
		(6.018)	(5.986)	(6.616)	(7.550)	(17.50)	(17.96)
$\gamma^{M}_{4/8}$		1.019	—	—	—	1.037	0.984
		(0.230)	—	—	—	(0.441)	(0.233)
$\gamma^{W}_{4/8}$		—	0.754	—	—	0.735*	0.733*
		—	(−3.544)	—	—	—	—
$\gamma^{1-4}_{M/W}$		—	—	1.026	—	1.071	1.008
		—	—	(0.293)	—	(0.860)	(0.090)
$\gamma^{5-8}_{M/W}$		—	—	—	0.743	0.760*	0.751*
					(3.508)		
Mean Log-L		−9.264	−8.883	−4.767	−4.682	−9.117	−9.088
n		357	325	682	682	682	682
J		8	8	4	4	8	8

* This estimate is the ratio of two estimates, hence the standard error is not calculated as a result of maximizing the log-likelihood. The delta method could be used, but both numerator and denominator estimates are precise, so we omit that calculation.

one-third more error variation in their evaluation of utility for questions 5 through 8 than for questions 1 through 4. This result is robust to specification, and is corroborated by alternative comparisons, discussed below. This finding gives us a way to compare the two survey modes. While the online survey is less costly, there is a tradeoff in that responses are less precise.

Models (vii) and (viii) combine the mail and online data in order to look at mode comparisons for subsets of the questions. The hypothesis that error standard deviations are equal across modes for questions 1 through 4 is not rejected ($\gamma^{1-4}_{M/W} = 1.026; t = 0.29$), but *is* rejected for questions 5 through 8 ($\gamma^{5-8}_{M/W} = 0.743; t = 3.54$). The estimated ratio of error standard deviations for mail versus online of 0.743 is consistent with similar estimates in models (v) and (vi) implying greater error variation for online respondents for the later questions. This result implies that neither survey mode has an advantage over the other initially, but something about the online interface causes respondents to tire or lose interest in the survey, and the quality of their responses declines. Estimates of $\gamma^{M}_{4/8}$, $\gamma^{W}_{4/8}$, $\gamma^{1-4}_{M/W}$, and $\gamma^{5-8}_{M/W}$ can also be obtained from the combined mail/online sample using all eight questions.[14] Combining the two samples brings to bear the maximum overidentifying information when it comes to estimation of the γs. Model (ix) is the most restrictive model in the sense that the marginal utilities for all respondents for all eight questions are constrained to be the same. Model (x) is less restrictive with the marginal utilities for speed, cost, and reliability constrained equal. These estimates are in the last two columns of Table III, and even with the added restrictions the estimates of the γs are very similar to the less restrictive models (v)–(viii).

4.4. Parameter Heterogeneity

If there is heterogeneity in the preference parameters (McFadden and Train, 2000), the traditional (nonrandom) parameter models fit above are misspecified. The implication is that this additional source of variation in choices would be attributed to the model errors ($\varepsilon^{k_{ij}}_{ij}$ and ε^{0}_{i}), and cause the estimates of $\gamma_{M/W}$ to be biased. To investigate this possibility, we fit the random parameters probit model to the mail and online data sets, looking at the preference parameter variation individually for each attribute. To be precise, we respecified equations (1) and (5) of the random utility models as

$$\begin{aligned} U^{k_{ij}}_{ij} &= (\beta_l + v_l)x^{k_{ij}}_{ijl} + \boldsymbol{\beta}'_{(l)}\boldsymbol{x}^{k_{ij}}_{ij(l)} + \varepsilon^{k_{ij}} \\ U^{0}_{ij} &= (\beta_l + v_l)x^{0}_{ijl} + \boldsymbol{\beta}'_{(l)}\boldsymbol{x}^{0}_{ij(l)} + \varepsilon^{0} \end{aligned} \tag{16}$$

where the vectors $\boldsymbol{\beta}_{(l)}$ and $\boldsymbol{x}^{k_{ij}}_{ij(l)}$ are missing the lth element, for $l = 1, \ldots, 5$. The random component of the parameter is v_l, with standard deviation σ_v. Table IV displays estimates of σ_v for each attribute and for both survey modes, their average over l, and the square root of the average variance.

Table IV. Preference parameter heterogeneity

	Mail	Online
ALWAYS ON	0.646*	0.477
SPEED	0.534	0.565
COST	0.382	0.410
INSTALL	0.262	0.376
RELIABLE	0.404	0.389
$\frac{1}{5}\sum\widehat{\sigma}_l$	0.446	0.443
$\sqrt{\frac{1}{5}\sum\widehat{\sigma}^2_l}$	0.465	0.449

* All estimates are precise, with standard errors less than 1/10 parameter estimates.

[14] We use the transpose mark, as in $\widehat{\gamma}'^{W}_{4/8}$ to signify these estimates.

DOI: 10.1002/jae

Although the standard deviations of the coefficients vary from mail to online mode, the differences are not great, and they do not vary systematically, in that for some attributes (ALWAYS ON, COST, RELIABLE) the mail coefficients have a larger variance, while for others (SPEED, INSTALL) the online coefficients are more variable. The mean of the standard deviations differs by only 0.003 (or about two-thirds of 1% of the mean), and the square root of the mean variances differs by 0.016. While the existence of any parameter variability will tend to mask differences in the standard deviations of the two modes, these small differences are not likely to change the conclusions reached above.

We also tried a variety of additional specifications to assess robustness. The basic results were not sensitive to choice experiment cutoff (e.g., rather than 1–4 vs. 5–8, 1–3 vs. 4–8 and 1–5 vs. 6–8); the possibility of a U-shaped relationship between variance and choice experiment, that is, the possibility that respondents *first* learn and later tire; and other parametric modeling of the variance.

Finally, tests of equality of means of some demographic variables in Table I reveal statistically significant differences in age, race, gender, education, income, and years online. Therefore, it is possible that some of the differences in error variances we find *between* samples ($\gamma^{1-4}_{M/W}$ and $\gamma^{5-8}_{M/W}$, reported in Table III) may be a result of these demographic differences, and not a result of the survey mode. To remedy this possible contamination of our results, we applied post-stratification weights to the two samples and there was very little difference in parameter estimates.

5. IMPLICATIONS OF THE RESULTS

Estimates of the ratio of standard deviations from the unweighted data are summarized in Table V. Mail respondents do not learn or fatigue when they answer repeated choice questions ($\widehat{\gamma}^{M}_{4/8} = 1.019$, $\widehat{\gamma}'^{M}_{4/8} = 0.984$), while online respondents suffer fatigue ($\widehat{\gamma}^{W}_{4/8} = 0.745$, $\widehat{\gamma}'^{W}_{4/8} = 0.733$). The advantage of mail over online survey mode is in the last four question replications. In the first four choice experiments, the quality of responses is about equal ($\widehat{\gamma}^{1-4}_{M/W} = 1.026$, $\widehat{\gamma}'^{1-4}_{M/W} = 1.008$), but in the last four choice experiments the standard deviation of the mail survey responses is only about three-quarters those from the online survey ($\widehat{\gamma}^{5-8}_{M/W} = 0.743$, $\widehat{\gamma}'^{5-8}_{M/W} = 0.751$). Since the surveys are identical, except for visual presentation and tools for completing the survey, model results provide *prima facie* evidence that visual and interactive stimuli from online delivery are placing greater cognitive demand on respondents. Accordingly, online respondents suffer fatigue or boredom, reducing consistency in choice, which causes data quality to deteriorate. Nevertheless, online surveys have substantial advantages with respect to lower costs for completed surveys (approximately \$30 per completed survey for online, compared to \$58 for mail), and reduced time for survey implementation and delivery of the final data set for analysis (20 days for online, compared to 53 days for mail). Given these tradeoffs, researchers administrating online choice experiments can attain a Pareto improvement by reducing the number of choice questions (relative to mail) to enhance administrative efficiency, while bearing no cost in terms of the efficiency of model estimation. In effect, savings in time and cost per completed survey are used to increase the number of sample observations from more respondents, but with fewer choice questions per respondent.

Table V. Estimates of γ

		Within mode	Combined
Mail 1–4/5–8	$\widehat{\gamma}^{M}_{4/8}$	1.019	0.984
Online 1–4/5–8	$\widehat{\gamma}^{W}_{4/8}$	0.745	0.733
Mail 1–4/Online 1–4	$\widehat{\gamma}^{1-4}_{M/W}$	1.026	1.008
Mail 5–8/Online 5–8	$\widehat{\gamma}^{5-8}_{M/W}$	0.743	0.751

6. CONCLUSIONS

Estimation results suggest that while the online survey provides benefits in terms of lower survey administration costs and reduced time between survey implementation and data analysis, these benefits may come at the cost of respondent fatigue and greater error standard deviation in the estimation of utility. For instance, in this experiment, sole use of online survey data for utility estimation reduces costs from about \$58 per completed mail survey to about \$30 per completed online survey, but variation in the error of online respondents' utility is about 25% higher for choice questions 5 through 8. Researchers administrating choice experiments through online surveys may want to consider increasing the number of observations with a larger sample and fewer choice questions per respondent, if they are to efficiently use their research budgets.

Although the results of this study apply to one experiment, they clearly suggest a need for further research into the effects of online technology and online surveys on the congnitive processes of respondents, and how these effects impact the consistency of choice during repeated choice experiments. This is particularly important as researchers continue to use the Internet to solicit opinions on major political decisions, public policy, and for collecting data to estimate consumer's valuation of the environment, public transport, recreational amenities, health treatments, and telecommunications.

DOI: 10.1002/jae

APPENDIX: CHOICE QUESTION EXAMPLE

Q1a. Check the Internet access option you would prefer (even if you do not view either A or B as ideal, tell us which you would most prefer):

Attribute	A	B
Always on	Always on	Not always on
Speed	Slow	Fast
Cost	\$25/month	\$45/month
Installation	Immediate	Immediate
Reliability	Less reliable	Very reliable
Mark the option you prefer:	☐	☐

Q1b. If you currently have Internet access at home, consider the always on, speed, cost and reliability features of your service. Would you switch to the access option (A or B) you choose above? (mark one answer)
Yes No
If you do not currently have Internet access at home, would you adopt the access option (A or B) you choose above? (mark one answer)
Yes No

ACKNOWLEDGEMENTS

The authors acknowledge research support from Cisco Systems, Knowledge Networks Inc. and the Interdisciplinary Telecommunications Program, University of Colorado, Boulder. Thanks for helpful comments from Don Dillman, Jordan Louviere, Gary McClelland, Edward Morey, Jayne Toman, Randall Walsh, Brad Wimmer, the associate editor and anonymous referees, and seminar participants at INFORMS, Denver, 2004, the University of New South Wales, Sydney University, the University of Tasmania, the University of South Australia, Monash University, the University of Nevada, and the University of Colorado. Jennifer Thacher provided research assistance. The usual disclaimer applies.

REFERENCES

Abramowitz M, Stegun J. 1964. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. *Applied Mathematics Series* 55, National Bureau of Standards: Washington, DC.
Berrens R, Bohara A, Jenkins-Smith H, Silva C, Weimer D. 2003. The advent of Internet surveys for political research: a comparison of telephone and Internet surveys. *Political Analysis* **11**: 1–22.
Bradley M, Daly A. 1994. Use of logit scaling approach to test for rank-order and fatigue effects in stated preference data. *Transportation* **21**: 167–184.
Brazell J, Louviere J. 1997. Respondent's help, learning and fatigue. Presented at the *1997 INFORMS Marketing Science Conference*, University of California, Berkeley.

DOI: 10.1002/jae

Brazell J, Gray-Lee J, Louviere J, Dallaert B, Pullman P. 1995. Nobody will do that! An empirical investigation of survey length effects on response rate and reliability in choice model experiments. Presented at the *1995 INFORMS Marketing Science Conference*, 2–5 July, Sydney, Australia.

Burr M, Levin K, Becher A. 2001. Examining web vs. paper mode effects in a federal government customer satisfaction study. Presented at the *Annual Conference of the American Association for Public Opinion Research*, Montreal.

Cameron T, Poe G, Ethier R, Schulze W. 2002. Alternative non-market value-elicitation methods: are the underlying preferences the same? *Journal of Environmental Economics and Management* **44**: 391–425.

Carini R, Hayek J, Kuh G, Kennedy J, Ouimet J. 2003. College student responses to web and paper surveys: does mode matter? *Research in Higher Education* **44**: 1–19.

Carson R, Mitchell R, Haneman W, Kopp R, Presser S, Ruud P. 1994. Contingent valuation and lost passive use: damages from the Exxon Valdez. Resources for the Future discussion paper, Washington, DC.

de Palma A, Myers G, Papageorgiou Y. 1994. Rational choice under an imperfect ability to choose. *American Economic Review* **84**: 419–440.

Dillman DA. 2000. *Mail and Internet Surveys: The Tailored Design Method* (2nd edn). Wiley: New York.

Elrod T, Louviere J, Davey K. 1992. An empirical comparison of ratings-based and choice-based conjoint models. *Journal of Marketing Research* **29**: 368–377.

Heckman JJ, Robb R. 1985. Alternative methods for evaluating the impact of interventions. In *Longitudinal Analysis of Labor Market Data* Heckman JJ, Singer B (eds). Cambridge University Press: Cambridge, UK; 156–245.

Heiner R. 1983. The origin of predictive behavior. *American Economic Review* **73**: 560–595.

Hensher D, Louviere J, Swait J. 1999. Combining sources of preference data. *Journal of Econometrics* **89**: 197–221.

Huber J, Zwerina K. 1996. The importance of utility balance in efficient choice designs. *Journal of Marketing Research* **33**: 307–317.

Louviere J, Hensher D, Swait J. 2000. *Stated Choice Methods: Analysis and Application*. Cambridge University Press: Cambridge, UK.

McClelland G, Judd C. 1993. Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin* **114**: 376–390.

McFadden D, Train K. 2000. Mixed MNL models for discrete response. *Journal of Applied Econometrics* **15**: 447–470.

Philipson T. 2001. Data markets, missing data, and incentive pay. *Econometrica* **69**: 1099–1111.

Philipson T, Malani A. 1999. Measurement errors: a principal investigator-agent approach. *Journal of Econometrics* **91**: 273–298.

Romeo C, Sopher B. 1999. Learning and decision costs in one-person games. *Journal of Applied Econometrics* **14**: 335–357.

Savage S, Waldman D. 2006. High technology goods in household production functions: the demand for internet bandwidth. Mimeo, Department of Economics, University of Colorado at Boulder.

Swait J, Adamowicz W. 1996. The effect of choice experiment and task demands on consumer behavior: discriminating between contribution and confusion. Staff paper 96–09, Department of Rural Economy, University of Alberta.

US Census Bureau. 2003. American FactFinder. http://factfinder.census.gov/servlet/BasicFactsServlet?_basicfacts=1&_multl=8 545 266&_geo2=010&_current=&_action=_subjectSelected&_child_geo_id=&_lang=en [21 March 2003].

VanBeselaere C. 2002. Survey response quality: shirking behavior in Internet and telephone surveys. Mimeo, Department of Social Science, California Institute of Technology.

Waldman DM. 1985. Computation in duration models with heterogeneity. *Journal of Econometrics* **28**: 127–134.

Zwerina K, Huber J, Kuhfeld W. 1996. A general method for constructing efficient choice designs. In *Marketing Research Methods in the SAS System*, 2002 Version 8 edn. SAS Institute: Cary, NC.

DOI: 10.1002/jae