# Consumer Preference Elicitation of Complex Products using Fuzzy Support Vector Machine Active Learning

Dongling Huang and Lan Luo

March 2015

Forthcoming: *Marketing Science*

*Special Issue: "Big Data"*

---

Dongling Huang (email: huangd3@rpi.edu) is an Assistant Professor of Marketing at the Lally School of Management, Rensselaer Polytechnic Institute, Troy, NY 12180. Lan Luo (email: lluo@marshall.usc.edu) is an Associate Professor of Marketing at the Marshall School of Business, University of Southern California, Los Angeles, CA 90089. The authors greatly appreciate seminar attendants at MIT, Columbia, the University of Texas at Austin, the University of British Columbia, the INFORMS Marketing Science Conference, and the AMA Advanced Research Techniques (ART) Forum for their constructive comments. The authors also thank the USC Marshall Center for Global Innovation for the financial support it has provided for this research.

**Abstract**

As technology advances, new products (e.g., digital cameras, computer tablets) have become increasingly more complex. Researchers often face considerable challenges in understanding consumers' preferences for such products. The current research proposes an adaptive decompositional framework to elicit consumers' preferences for complex products. The proposed method starts with a collaborative-filtered initial part-worths, followed by an adaptive question selection process where fuzzy support vector machine active learning algorithm is used to adaptively refine the individual-specific preference estimate after each question. Our empirical and synthetic studies suggest that the proposed method performs well for product categories equipped with as many as 70 to 100 attribute levels, which is typically considered prohibitive for decompositional preference elicitation methods. In addition, we demonstrate that the proposed method provides a natural remedy for a long-standing challenge in adaptive question design by gauging the possibility of response errors on the fly and incorporating it into the survey design. This research also explores in a live setting how responses from previous respondents may be used to facilitate active learning of the focal respondent's product preferences. Overall, the proposed approach offers some new capabilities that complement existing preference elicitation methods, particularly in the context of complex products.

Keywords: product development, support vector machines, machine learning, active learning, adaptive questions, conjoint analysis.

## 1. Introduction

As technology advances, new products (e.g., digital cameras, computer tablets) have become increasingly more complex. Researchers often face considerable challenges in understanding consumers' preferences for such products (e.g., Green and Srinivasan 1990; Hauser and Rao 2004). Conventional preference elicitation methods such as conjoint analysis often become infeasible in this context, because the number of questions required to obtain accurate estimates increases rapidly with the number of attributes (and/or attribute levels). Historically, researchers have mostly relied on compositional approaches to handle preference elicitation of such products (e.g., Srinivasan 1988; Scholz et al. 2010; Netzer and Srinivasan 2011). Adaptive question selection algorithm has also been proposed for complex product preference elicitation, due to its ability to rapidly reveal consumer's product preferences with relatively few questions (e.g., Netzer and Srinivasan 2011). While significantly enhancing our abilities to understand consumers' preferences for complex products, the extant research has yet to address the following challenges.

*First*, although compositional approaches have been primarily used, such methods may encounter obstacles such as unrealistic settings, inaccurate attribute weighting, and etc. (e.g., Green and Srinivasan 1990; Sattler and Hensel-Borner 2000) *Second*, despite its high efficiency in uncovering consumers' product preferences, adaptive question selection method is often subject to response errors that could potentially misguide the selection of each subsequent question. *Lastly*, with the exception of Dzyabura and Hauser (2011) in the context of consideration heuristics elicitation, this line of research has yet to explore the possibility of utilizing other respondents' data to facilitate active learning of the focal respondent's product preferences.

The current research proposes an adaptive decompositional framework in response to these challenges. The proposed method starts with a collaborative-filtered initial part-worths, followed by an adaptive question selection process where fuzzy support vector machine (SVM) active learning algorithm is used to adaptively refine the individual-specific preference estimate after each question. Compared to extant preference elicitation methods, our research offers the following new capabilities:

- Our adaptive decompositional approach is computationally efficient for preference elicitation of complex products on the fly, as the algorithm primarily scales with the sample size of the training data, rather than the dimensionality of the data vector.
- While extant research either neglects response errors in adaptive question selection or sets possibility of error instance as a priori, our algorithm gauges the possibility of response errors on the fly and incorporates it into adaptive survey design.
- Although most adaptive question selection methods only utilize information from the focal respondent, we use responses from previous respondents in a live setting via collaborative filtering to facilitate active learning of the focal respondent's product preferences.

We illustrate the proposed method in two computer-based studies involving digital cameras (with 30+ attribute levels) and computer tablets (with 70+ attribute levels). Our empirical investigation demonstrates that the proposed method outperforms the self-explicated method, the adaptive Choice-Based Conjoint method, the traditional Choice-Based Conjoint method, and an upgrading method similar to Park et al. (2008) in its ability to correctly predict validation choices. Our synthetic data experiments further demonstrate that the proposed method is able to rapidly and effectively elicit individual-level preference estimates even when the product category is equipped with more than 100 attribute levels. We also use synthetic data

experiments to compare the scalability, parameter recovery, and predictive validity of the proposed algorithm with that of Dzyabura and Hauser (2011) for consideration questions and of Abernethy et al. (2008) for choice questions. We show that the proposed question selection algorithm may be used either in conjunction with or as a substitute for the algorithms by Dzyabura and Hauser (2011) and Abernethy et al. (2008) to uncover consumers' preferences for complex products. Overall, our empirical and synthetic studies suggest that the proposed approach offers a promising new method to complement existing preference elicitation methods.

The remainder of the paper unfolds as follows. In Section 2, we discuss the relationship of this research to the extant literature. In Section 3, we present details of the proposed adaptive question design algorithm. In Section 4, we describe our two empirical applications. Details of our synthetic studies are presented in Section 5. The final section concludes the paper by summarizing key results and offering directions for future research.

## 2. Relationship to Extant Literature

The algorithm employed in our proposed framework is closely related to the machine learning literature with origins in computer science. An important application of machine learning is classification, in which machines "learn" to recognize complex patterns, to distinguish between exemplars based on their different patterns, and to make intelligent predictions on their classes. Many marketing problems require accurately classifying consumers and/or products (e.g., consumer segmentation; identification of desirable vs. undesirable products). Therefore, marketing researchers have started to embrace machine learning methods for estimation of classic marketing problems in recent years (e.g., Cui and Curry 2005; Evgeniou et al. 2005; Evgeniou et al. 2007; Hauser et al. 2010).

Built upon this stream of literature, the current paper introduces support vector machine based active learning into adaptive question design. Arguably the most popular statistical machine learning method in the past decade (Toubia, Evgeniou, Hauser 2007), support vector machine methods are well-known for high-dimensional classification problems (e.g., Vapnik 1998; Tong and Koller 2001). In particular, we employ fuzzy SVM method to adaptively select each subsequent question for each individual on the fly. As a weighted variant of the soft margin SVM formulation (the soft margin SVM was initially introduced by Cortes and Vapnik 1995), the fuzzy SVM method assigns different weights to different data points to enable greater flexibility of error control. In recent years, the class of fuzzy SVM methods has gained notable popularity in the SVM literature, mainly due to its effectiveness in reducing the effect of noises/errors in the data (e.g., Lin and Wang 2002, 2004; Wang, Wang, and Lai 2005; Shilton and Lai 2007; Heo and Gader 2009).

When used for preference elicitation of complex products, this algorithm exhibits a number of advantages over extant methods. First, among others, one desirable property of the SVM-based active learning algorithm is that the optimization used to facilitate adaptive selection of each subsequent question can be transformed to a dual convex optimization problem (Tong and Koller 2001). Within our context, the primal problem (Equations 2 and 4) is also constructed to be convex. Therefore, the proposed algorithm not only offers an explicitly defined unique optimum but also is easily solvable by most software for problems with dimensions that are likely to be of interest to marketers. Indeed, the SVM-based classification is primarily scaled by the size of the training data (the number of questions presented to each consumer in our context), rather than the dimensionality of the data vector (Dong, Krzyzak, and Suen 2005). Consequently, the SVM-based active learning is particularly suitable for the problem at hand. In contrast,

several alternative adaptive methods (such as the adaptive fast polyhedral methods by Toubia et al. 2003, Toubia, et al. 2004, Toubia, et al. 2007) are scaled by the dimensionality of the product vector, which may become more computationally cumbersome as the dimension of product attributes/attribute levels increases. Moreover, while the Hessian-based adaptive methods (e.g., Abernethy et al. 2008; Toubia et al. 2013) require discrete transformations when used for discrete attributes, the SVM active learning method is flexible enough to directly accommodate both discrete and continuous product attributes.

Furthermore, another unique advantage particularly related to the fuzzy SVM active learning method is that it enables us to gauge the possibility of response errors on the fly and to incorporate it into adaptive question selection. In the context of adaptive question design, response errors may be conceptualized as the random error component in the consumer's utility function (e.g., Toubia et al. 2003). Empirical data suggest that response errors are approximately 21% of total utility (Hauser and Toubia 2005). Because each response error has the potential to set the adaptive question selection to the wrong path and negatively impact selection of all subsequent questions, the presence of such errors poses a long-standing challenge to the adaptive question design literature (e.g., Hauser and Toubia 2005; Toubia et al. 2007). To date, response errors have either been neglected (e.g., Toubia, et al. 2004; Netzer and Srinivasan 2011) or set as a priori possibility for all individuals and all questions (e.g., Toubia et al. 2003; Toubia, et al. 2007; Abernethy et al. 2008; Dzyabura and Hauser 2011; Toubia et al. 2013). We demonstrate that the proposed method can be used not only to gauge possible response errors on the fly but also to reduce the effects of such noises in adaptive question selection.

Lastly, inspired by Dzyabura and Hauser (2011) who suggest previous-respondent data may be used to improve elicitation of consideration heuristics, the proposed method utilizes

responses from previous respondents via collaborative filtering to facilitate active learning of the focal respondent's product preferences. The concept of collaborative filtering has been applied in various contexts such as prediction of TV show preferences and movie recommendation systems (Breese et al. 1998). We illustrate that such technique can be incorporated in adaptive question design using actual respondents in a live setting.

## 3. Adaptive Question Selection for Complex Product Preference Elicitation

### 3.1. Overall Flow

Figure 1 depicts the overall flow of our adaptive question design. We start with prompting the consumer to configure a product that he is most likely to purchase, taking into account any corresponding feature-dependent prices. Based on collaborative filtering between the focal and previous respondents' self-configured profiles, we obtain an individual-specific initial part-worths vector (Section 3.2). Next, we provide each consumer with an option of selecting "must-have" and "unacceptable" product features. We utilize information obtained from such features to construct the pool of candidate profiles to be evaluated in the consideration stage, as it is infeasible to evaluate all profiles using active learning without excessive delays in between survey questions within our context (Section 3.3). Conditional on the initial part-worths and the candidate pool of profiles obtained for each individual, we employ a two-stage consider-then-choose process where the consideration stage asks the respondent whether he would consider a product profile (Section 3.4) and the choice stage asks the respondent to choose among competing product profiles (Section 3.5). In both stages, we employ fuzzy SVM active learning for adaptive question design, so that each subsequent question is individually customized to refine the consumer-specific preference estimate while accounting for possible response errors on the fly.

<Insert Figure 1 about here>

We explain the underlying rationale of our overall framework below. The primary goal of the proposed method is to estimate a part-worths vector for each respondent *j*. Given that the scale of the part-worths vector is arbitrary (Orme 2010), before the respondent answers any question, we may visualize the feasible region of the part-worths being all the points on the surface of a hypersphere with unit norm (i.e., $\mathbf{w}_j \in \mathbf{W} \middle| \left\|\mathbf{w}_j\right\| = 1$). Such a feasible region is referred to as version space in Tong and Koller (2001) and Herbrich et al. (2001) and a polyhedron in Toubia et al. (2003, 2004, and 2007). Conceptually, each answer given by the respondent provides constraint(s) that makes this feasible region smaller.

Given the large number of attributes/attribute levels associated with complex products and the limited number of questions we can ask each respondent, an informative first question would enable us to efficiently construct the initial region of the feasible part-worths (Section 3.2). Similarly, by constructing a candidate pool with the majority of profiles satisfying the "must-have"/"unacceptable" criteria, we can maximize our learning about the focal respondent's product preferences by asking whether he would consider a profile based on his favorability towards other product features (Section 3.3). Essentially, the first two steps of our overall framework aim to construct a suitable foundation for the adaptive question selection later on.

We then employ a consider-then-choice framework to elicit each respondent's product preference (Sections 3.3. and 3.4). Specifically, our algorithm aims to uncover a set of part-worths estimates that are consistent with consumer's answers to these questions. Historically, researchers have often used conjunctive rules to capture consumer's decision rules in the consideration process (e.g., Hauser et al. 2010; Dzyabura and Hauser 2011). And it has been less common to use part-worths to characterize consumer's responses to consideration questions. Our

synthetic data experiments reveal that, even when the true consideration model is driven by conjunctive decision rules, the part-worths estimate from our algorithm exhibits good ability to predict whether the respondent would consider a profile (Section 5.1). Indeed, even if heuristic decision rules are adopted by some respondents, particularly in the consideration stage, such preferences will be partially captured in our individual-specific part-worths estimates that aim to optimally predict responses from these consumers. Therefore, while not explicitly portraying these respondents' consideration heuristics, our part-worths estimates serve as an approximation of the decision heuristics used by such individuals.[1]

Within this setup, we present an algorithm where we use a set of part-worths estimates to characterize a respondent's answers to both consideration and choice decisions. In this context, we aim to select each subsequent consideration/choice question such that we can reduce the feasible region of the part-worths as rapidly as possible. Intuitively, one good way of achieving this goal is to choose a question that halves such a region (Tong and Koller 2001; Toubia et al. 2003, 2004, and 2007). To achieve this goal, we adapt the active learning approach proposed by Tong and Koller (2001) to select each subsequent consideration/choice question on the fly. Similar to Toubia et al. (2003, 2004, and 2007), this algorithm relies on intermediate individual-level part-worths estimates to adaptively select each next question. When the consumer makes no response errors, such an approach would rapidly shrink the feasible region of the part-worths. Nevertheless, response errors are often inevitable in practice. To reduce the effects of response errors, we adapt the fuzzy SVM estimation algorithm (Lin and Wang 2002) in both consideration

[1] It is worth noting that, in practice, some consumers may not use the same utility function for consideration and choice decisions. For example, an individual might emphasize different sets of attributes in the consideration phase versus in the choice phase. In Section 5.1, we discuss how our algorithm can be used in conjunction with the algorithm by Dzyabura and Hauser (2011) to accommodate a consider-then-choice framework where conjunctive rules are used to examine answers to consideration questions and conjoint part-worths are used to capture product preferences reflected in choice questions. In such cases, different utility functions may also be employed to model consideration and choice decisions separately.

and choice stages to obtain the intermediate par-worths. Under this algorithm, each part-worths estimate is obtained as an interior point within the current feasible region of part-worths, via a simultaneous optimization that balances between data-imposed constraints and weighted classification violation. Consequently, when selecting each subsequent question based on such intermediate part-worths estimates, the negative impact of response errors in the process of adaptive question selection would be alleviated.

We also conjecture that our current multi-stage framework may help reducing the effect of response errors. In adaptive question design, early response errors are considerably more detrimental than errors that occur toward the end of the adaptive survey (e.g., Hauser and Toubia 2005; Toubia et al. 2007). Therefore, when presented with the less-demanding self-configuration and consideration questions prior to the more-challenging choice questions, respondents may be less likely to incur response errors early on.[2] Pseduo code of our algorithm is provided in Web Appendix A. Screenshots of the survey interface from our computer tablet applications are presented in Web Appendix B.

### 3.2. Collaborative-Filtered Initial Part-worths Vector

Our framework starts with asking each respondent to configure a product profile that he is most likely to purchase, taking into account any corresponding feature-dependent prices.[3] Such self-configured product profile provides substantial information about a consumer's product

[2] One potential concern is that such a flow might prime respondent to answer questions in a manner consistent with their previous responses, even if erroneous. Our conjecture is that, given the high dimensionality of the product vector, consumers are less likely to accurately recall errors and intentionally adjust responses during consideration and choice stages. To alleviate this concern in the first two stages of our framework, our survey interface checks whether the respondent lists a product feature as unacceptable when he previously configured a product with this feature. When such scenario occurs, the respondent is prompted to go back and double check his inputs. If the respondent is primed to the extent that prior errors are not self-recognizable, this cross check mechanism would not be effective.

[3] Following Johnson and Orme (2007), we include feature-dependent prices in the self-configuration task to increase the realism of this task (otherwise respondents may self-configure the most advanced product profile with the lowest price). In the subsequent consideration and choice questions, we follow Orme (2007) by adopting a summed price approach with a plus/minus 30% random price variation in both empirical studies.

preferences (Dzyabura and Hauser 2011; Johnson and Orme 2007). Within our context, the information contained in the self-configured profile is utilized as follows.

First, we generate an initial individual-specific part-worths vector based on collaborative filtering between the focal and previous respondents' self-configured product profiles. The basic intuition is that we may learn about the focal respondent's product preferences by examining the preferences of previous respondents who have configured similar product profiles. For example, if two respondents self-configured two identical computer tablets, they are likely to share some commonality in their overall preferences towards computer tablets. Analogous to the role of informative prior in the Bayesian literature, consumer-specific initial part-worths obtained through collaborative filtering has the potential to enhance active learning of the focal respondent's product preference at the outset of our adaptive question selection. Specifically, after the focal respondent *j* configures a profile that he is most likely to purchase, the following equation is used to obtain the respondent's initial part-worths vector $\tilde{\mathbf{w}}_j^0$ on the fly:

(1) $$\tilde{\mathbf{w}}_j^0 = \frac{1}{j-1}\sum_{j'=1}^{j-1} s(j, j') * \tilde{\mathbf{w}}_{j'}$$

with $$s(j, j') = \frac{\mathbf{c}_j^{'} * \mathbf{c}_{j'}}{\|\mathbf{c}_j\| * \|\mathbf{c}_{j'}\|}$$

where $\tilde{\mathbf{w}}_{j'}$ is the estimated part-worths of previous respondent $j'$ with $j' = 1, \ldots, j-1$; $\mathbf{c}_j$ denotes the vector that represents product features of the self-configured profile for the focal respondent *j*; $\mathbf{c}_{j'}$ represents the corresponding vector from previous respondent $j'$; and $s(j, j')$ measures the degree of cosine similarity between vectors $\mathbf{c}_j$ and $\mathbf{c}_{j'}$.

The cosine similarity measure is widely used to capture similarity between two vectors in informational retrieval and collaborative filtering literature (e.g., Salton and McGill 1986; Breese

1998 et al.). Given that our method employs aspect type coding with attribute-level dummies, this measure is bounded between 0 and 1. $s(j, j')=0$ if there is no overlap between $\mathbf{c}_j$ and $\mathbf{c}_{j'}$; $0<s(j, j')<1$ if there is partial overlap between $\mathbf{c}_j$ and $\mathbf{c}_{j'}$; and $s(j, j')=1$ if $\mathbf{c}_j = \mathbf{c}_{j'}$. The resulting initial part-worths is then used to identify the next set of profiles shown to the consumer.

Second, we utilize information contained in the configurator to set the two initial anchoring points in our fuzzy SVM active learning algorithm. As a classification method, a well-posed SVM problem entails training data from both classes. In our context, we first give the self-configured product profile (i.e., the respondent's favorite) a label of "1" (meaning that the consumer would consider it). Next, we select a profile among the ones that are the utmost different from the self-configured profile and give it a label of "-1". As such a profile is not unique, the "opposite" profile is randomly chosen among the ones that do not share any common feature with the self-configured profile (our synthetic data experiments suggest that, in over 99% of cases, consumers would not consider such an opposite profile). After the next set of profiles is queried based on the collaborative-filtered initial part-worths, we combine the two anchoring points with the newly labeled profiles in the training data to ensure that the SVM problem is well-posed. When previous respondents are absent, only the two anchoring points are used to obtain the initial part-worths vector for the focal respondent.

### 3.3. Identify Must-Have and/or Unacceptable Product Features

After the configuration task, we proceed to provide each respondent with an option of selecting some "must-have" and "unacceptable" product features. Within the context of complex products, it is often infeasible to evaluate all profiles using active learning without excessive delays in between survey questions (Dzyabura and Hauser 2011). One major advantage of identifying the

"must-have" and "unacceptable" product features is that such information can be leveraged to construct the pool of candidate profiles to be evaluated in the consideration stage.[4]

In particular, among all possible product profiles to be queried, we develop an individual-specific pool containing mostly (e.g., 90%) profiles that satisfy the "unacceptable" and "must-have" criteria (denoted as $N_j^1$), with remaining profiles randomly chosen from the ones that do not satisfy these criteria (denoted as $N_j^2$).[5] The rationale of having the majority of profiles in the candidate pool satisfying the "must-have"/ "unacceptable" criteria is that we can maximize our learning about the focal respondent's product preferences by asking whether he would consider a profile based on his favorability towards other product attributes. The remaining profiles are chosen to account for the possibility that some individuals may identify "desirable" / "undesirable" features as "must-haves" / "unacceptables" (Johonson and Orme 2007). As long as the size of $N_j^2$ is sufficiently large, our adaptive algorithm will update the estimated part-worths vector so that profiles not satisfying the initial criteria may also be queried in subsequent survey questions.

[4] Note that one potential caveat of this approach is that the inclusions of "must-have" and "unacceptable" features might prime the respondent into a conjunctive-style decision making. An alternative would be to use the uncertainty sampling method used by Dzyabura and Hauser (2011) to construct the candidate pool of product profiles, with the tradeoff that it might be challenging to accurately identify the most uncertain profiles during the first few queries.

[5] Synthetic data experiments reveal that, as long as the total number of candidate profiles (i.e., $N_j = N_j^1 + N_j^2$) is sufficiently large, we are able to recover a part-worths estimate that is close to the true part-worths under our active learning method. In the empirical applications, we set $N_j$ to be 20,000. Our synthetic studies suggest that this is sufficiently large to recover the true part-worths while keeping the question selection to be less than 0.25 second in between questions at the consideration stage. Similar approaches are used to determine the number of profiles to be considered in the choice stage.

### 3.4. Consideration-Based Fuzzy SVM Active Learning

We next present the consideration-based fuzzy SVM active learning algorithm. We first describe the algorithm used to estimate the individual-specific part-worths vector on the fly. Then we elaborate the algorithm used to adaptively select profiles queried in each subsequent question.

**3.4.1. Algorithm to Estimate Individual-Specific Part-worths on the Fly.** Let $\mathbf{x}_j^i$ ($i$ = 1, 2 ,…, $I$; $j$ =1, 2, …, $J$) denote the aspect type coded product profile $i$ labeled by respondent $j$ with the label $y_j^i = 1$ when the respondent indicates that he would consider the profile and $y_j^i = -1$ if he would not consider it. Following the tradition in the conjoint literature, we employ a main-effects only model where $\mathbf{w}_j^I * \mathbf{x}_j^i$ denotes the utility estimate of product profile $i$ ($i$ = 1, 2 ,…, $I$) for respondent $j$ and $\alpha_j^I$ being the utility estimate of his "no-choice" option after $I$ profiles are labeled. The utility of the "no-choice" option represents the decision boundary where the consumer would only consider a profile if its utility is no less than the baseline utility associated with the "no-choice" option (Haaijer, Kamakura, and Wedel 2001). Namely, consumer $i$ will only consider profile $j$, i.e., $y_j^i = 1$, if $\mathbf{w}_j^I * \mathbf{x}_j^i - \alpha_j^I \geq 0$; $y_j^i = -1$ otherwise.

Within this context, the primary purpose of the SVM estimation algorithm is to find an individual-specific part-worths (i.e., $\tilde{\mathbf{w}}_j^I = (\mathbf{w}_j^I, \alpha_j^I)$ ) that can correctly classify labeled profiles into the two classes of "would consider" and "would not consider". Finding such a part-worths vector can be challenging in practice, since 1) there may be response errors in the data and 2) the true decision process may not be representable by a linear utility function as specified above. With regards to the first issue, we employ a fuzzy SVM algorithm that assigns different weight to each labeled profile along with a regularization parameter that enables classification violation (we will provide more details on this when we present the algorithm). And the second issue can

be alleviated by employing aspect type coded product utility and/or by introducing a non-linear kernel to the SVM algorithm. As compared to the alternative continuous/attribute level order coding, the aspect type coded product utility functions enable greater flexibility to accommodate nonlinear preference within each attribute (e.g., the utility function does not require monotone preferences for screen size, such as smaller/bigger size is strictly better). Additionally, nonlinear kernels could be used if there are nonlinear preferences across attributes. For example, if prior knowledge suggests that interaction effects exist among two or more product attributes levels, the SVM estimation algorithm can be readily adapted to accommodate such a nonlinear utility function. Vapnik (1998) and Evgeniou et al. (2005) provide detailed discussions on the generalization of the SVM estimation algorithm to such nonlinear models, which also maintains its computational efficiency even with highly nonlinear utility functions.

For simplicity, we demonstrate our estimation algorithm below using the example of the main-effects only model. Formally, upon the labeling of *I* profiles (*i* = 1, 2 ,…, *I*), the following algorithm is used to estimate respondent *j*'s part-worths vector (Vapnik 1998; Tong and Koller 2001; Lin and Wang 2002):

$$\text{(2)} \qquad \begin{aligned} &\min_{\mathbf{w}_j^I, \alpha_j^I, \xi_j^i} \quad \left[ \frac{1}{2} \left\| \mathbf{w}_j^I \right\| + C \sum_{i=1}^{I} u_j^i \xi_j^i \right] \\ &s.t. \quad y_j^i \left( \mathbf{w}_j^I * \mathbf{x}_j^i - \alpha_j^I \right) \geq 1 - \xi_j^i \\ &\qquad \xi_j^i \geq 0 \end{aligned}$$

where *C* is an aggregate-level regularization parameter that allows a certain degree of prior misclassification at the aggregate level, $\xi_j^i$ is a slack variable that can be thought of as a measure of the amount of misclassification associated with profile *i*, and $u_j^i$ assigns different weight to each labeled profile. When $u_j^i = 1$, Equation (2) corresponds to the soft margin SVM algorithm.

The regularization parameter $C$ in Equation (2) needs to be determined outside of the SVM optimization.[6] While this parameter may partially absorb the negative impact of noises in the data by allowing a certain degree of prior misclassification at the aggregate level, we discuss below how the fuzzy membership method provides additional flexibility to gauge potential response errors on the fly.

Instead of assuming that all labeled data belong to one of the two classes with 100% accuracy, the fuzzy SVM method assigns a fuzzy membership to each labeled profile so that data points with high probability of being corrupted by noises will be given lower values of fuzzy memberships (Lin and Wang 2002). Therefore, rather than giving each labeled data point equal weight in the optimization, profiles with higher probabilities of being meaningless will be given less weight in the estimation under the fuzzy SVM method.

In practice, researchers often do not have complete knowledge about the causes and/or nature of noises in the data. Therefore, in the machine learning literature, researchers have explored various approaches to discern noises/outliers in the data (e.g., Lin and Wang 2002, 2004; Wang, Wang, and Lai 2005; Shilton and Lai 2007; Heo and Gader 2009). We adopt a method similar to Lin and Wang (2002) to assign each labeled profile with a fuzzy membership $u_j^i$ ($0 < u_j^i \leq 1$) as follows:

(3)
$$u_j^i = \begin{cases} 1 - \dfrac{\left\| \mathbf{x}_j^i - \bar{\mathbf{x}}_{j+}^I \right\|}{r_{j+}^I + \eta} & if\ y_j^i = 1 \\ 1 - \dfrac{\left\| \mathbf{x}_j^i - \bar{\mathbf{x}}_{j-}^I \right\|}{r_{j-}^I + \eta} & if\ y_j^i = -1 \end{cases}$$

[6] Following approaches described in Evgeniou et al. (2005) and Toubia et al. (2007), we employ a cross-validation method based on pretest data from the same population as the main study to determine the values of $C$ in our two empirical applications.

where $r_{j\pm}^{I} = \max_{i \in N_{j\pm};} (\left\| \mathbf{x}_j^i - \bar{\mathbf{x}}_{j\pm}^{I} \right\|)$ represents the radius of each class (with the two classes being consider vs. not consider), $\bar{\mathbf{x}}_{j\pm}^{I} = \frac{1}{N_{j\pm}^{I}} \sum_{i \in N_{j\pm}} \mathbf{x}_j^i$ denotes each class's group center, $N_{j+}^{I} = \{i \mid y_j^i = 1\}$ and $N_{j-}^{I} = \{i \mid y_j^i = -1\}$ indicates the number of profiles in each class upon labeling of $I$ profiles, and $\eta$ is a small positive number to ensure that all $u_j^i > 0$.

Because the respondent's true part-worths are unknown to researchers, given the two classes of labeled profiles, we use the center of each class to approximate our current best guess about a profile that is representative of its class. We then define the fuzzy membership as a function of the Euclidean distance of each labeled profile to its current class center. Namely, given our current knowledge about the respondent's product preferences, we assess the possibility of response errors by examining to what extent the labeled profile differs from other labeled profiles in its class.

Therefore, depending on its distance to the current class center, the labeled profile may be assigned a 90% probability belonging to one class and a 10% probability of being meaningless, or a 20% probability belonging to one class and an 80% probability of being meaningless. In Equation (2), profiles with higher probabilities of being meaningless (i.e., profiles with smaller $u_j^i$ estimates) are given less weight in the fuzzy SVM estimation algorithm.

We repeat the process outlined in Equations (2) and (3) iteratively upon the labeling of each additional profile. Specifically, each time an additional profile is labeled, we assign it with a fuzzy membership given the class center of prior labeled profiles in its class, based on which an updated part-worths is estimated. Next, we update our class center estimates and the $u_j^i$ ($i = 1,2,...,I$) estimate for each labeled profile to date. As a result, as we gain additional

information about respondent *j*'s product preferences, the algorithm can be used to iteratively refine both the part-worths estimate and the fuzzy membership estimates. Because the intermediate part-worths estimate is used to adaptively select the subsequent question, the proposed fuzzy SVM method can be used not only to gauge possible response errors in the data but also to reduce the effects of such noises in the adaptive question design.

We have also used synthetic data experiments to explore several alternative approaches to define a profile's class membership probability, such as defining a profile's class membership as a function of its distances to both its own class center and the center of the opposite class, or imposing an underlying error distribution assumption similar to the Logit, Probit, or the Gaussian Mixture models. To the extent that the estimation is feasible, we do not observe improvement in model performance by implementing such alternative weighing schemes (details in Web Appendix C).

It is also worth noting that the class of fuzzy SVM methods discussed above faces both potential gains and losses in adaptive question selection. On the positive side, this method alleviates the negative effect of response errors if such errors exist (see more investigation on this matter in Section 5.3). On the negative side, if the respondent does not incur an error, our fuzzy membership estimates may render the estimation less efficient. Given that the slack variable $\xi_j^i$ in Equation (2) equals to 0 for all non-support vectors in the solution, the efficiency loss only occurs when the solutions to the optimization in Equation (2) are affected by the $u_j^i$ estimates associated with correctly classified support vectors. Synthetic data experiments reveal that, when employed to data with no response errors, the fuzzy SVM active learning indeed incurs a minor efficiency loss when compared to the soft margin SVM active learning (Section 5.3). Therefore, if researchers are uncertain about the degree of response errors in adaptive

question design, the fuzzy SVM method described above may be used to alleviate the negative effect of possible response errors at the expense of a potentially minor efficiency loss. On the other hand, the soft margin SVM can be used to maximize the efficiency in active learning if prior experiences indicate that consumer's responses are highly deterministic (i.e., response errors play a negligible role).

**3.4.2. Algorithm to Adaptively Select Profiles to Be Queried in Subsequent Question.** In this section we discuss how we adaptively select the next set of profiles shown to the respondent based on the latest estimate of his individual-specific part-worths. We employ an approach adapted from Tong and Koller (2001). Our primary goal is to query each subsequent profile in such a way that we can reduce the feasible region of the part-worths as rapidly as possible. Intuitively, one good way of achieving this goal is to choose a query that halves such a region.

Let $\tilde{\mathbf{w}}_j^I = (\mathbf{w}_j^I, \alpha_j^I)$ denote the part-worths vector obtained from the optimization in Equation (2) after *I* profiles are labeled ($\tilde{\mathbf{w}}_j^0$ in Equation (1) is used if no profiles have been labeled). In its simplest form, Tong and Koller (2001) suggest that the next profile to be queried can be the one with the smallest distance (margin) to the current hyperplane estimate represented by $\tilde{\mathbf{w}}_j^I$. Within our context, the margin of an unlabeled profile is computed as $m_j^g = \left|\mathbf{w}_j^I \mathbf{x}_j^g - \alpha_j^I\right|$, with *g* being the index of unlabeled profiles. Tong and Koller (2001) show that, when the training data are symmetrically distributed and the feasible region of part-worths is nearly sphere shaped, active learning via this simple margin approach is able to reduce the current version space into half.

In the context of adaptive question design, the training data are very likely to be asymmetrically distributed and/or the feasible region of the current part-worths can be elongated.

To overcome such restrictions in the simple margin approach, Tong and Koller (2001) propose the ratio margin approach as an augmentation. While conceptually appealing, the ratio margin approach is considerably more computationally burdensome than the simple margin approach. In this research, we employ the following hybrid method to combine the two approaches.

Let us assume that, with the simple margin approach, we have narrowed down to *S* trial profiles that are closest to the current hyperplane estimate represented by $\tilde{\mathbf{w}}_j^I$. We then take each trial profile *s* (*s* = 1, 2, …, *S*) from this set, give it a hypothetical label of 1, calculate a new part-worths by combining this new trial profile to the labeled profiles, and obtain a hypothetical margin $m'^{s+}_j$. Next, we perform a similar calculation by relabeling this profile as -1, calculating the resulting part-worths vector, and obtaining a hypothetical margin $m'^{s-}_j$. And the ratio margin of this profile is defined as $\max\left(\frac{m'^{s+}_j}{m'^{s-}_j}, \frac{m'^{s-}_j}{m'^{s+}_j}\right)$. After repeating these for all the *S* trial profiles, we pick the profile with the smallest ratio margin as the next profile shown to the consumer (i.e., $\min_{s=1,\ldots,S}\left[\max\left(\frac{m'^{s+}_j}{m'^{s-}_j}, \frac{m'^{s-}_j}{m'^{s+}_j}\right)\right]$). By asking the consumer to reveal his preference for such a profile iteratively, we will be able to most rapidly reduce the current feasible region of the part-worths (Tong and Koller 2001).

In our empirical application, because more than one profile is shown to the respondent at one time (Figure B3 in Web Appendix), the set of profiles (e.g., five) with the smallest ratio margins under the most recent part-worths estimate are selected to query the respondent. Additionally, upon satisfying the smallest ratio margin criterion, if two or more profiles have the same ratio margins, the profiles with the shortest overall distances to both class centers will be chosen. We impose this modification to the original approach by Tong and Koller (2001) so that,

within the context of our fuzzy SVM estimation, if such profiles turn out to be correctly labeled support vectors, the efficiency loss will be minimized.

This process is repeated iteratively until $Q1$ questions are asked. As discussed in Dzyabura and Hauser (2011), the number of questions to be included may be chosen based on prior experience or managerial judgment. Within our context, we carried out synthetic data experiments with the same dimensions as the two studies in our empirical investigation. We discovered that the proposed method could correctly classify the vast majority of profiles in various contexts after eight screens of profiles are queried, with each screen comprising of five profiles. Consequently, we adopted this stopping rule for the data collection in our empirical application. Similar approaches were used to determine the number of questions to be asked in the choice stage.

### 3.5. Choice-Based Fuzzy SVM Active Learning

Upon completion of the consideration stage, we use the latest part-worths estimate to compute the utilities of all candidate profiles for respondent $j$, from which a set of $M_j$ profiles is selected to be considered in the choice stage. Similar in spirit to the "uncertainty sampling" rule adopted in Dzyabura and Hauser (2011), the profiles are selected such that their utility estimates are the closest to the decision boundary determined by the most recent part-worths estimate.

In choice tasks, when an individual frequently opts for the no-choice option, we cannot efficiently learn about his favorability towards various product features. In contrast, we obtain substantially more information about how an individual makes tradeoffs among different product features when he chooses one profile over the competing profile(s) in a choice set. Therefore, in addition to selecting profiles whose utility estimates are closest to the baseline utility estimate (i.e., the no-choice option), we employ a selection rule in which the majority (e.g., 90%) of

profiles in $M_j$ has utility estimates above the threshold defined by the "no-choice" option. And the remaining profiles in $M_j$ have utility estimates less than the threshold to allow for potential estimation error from our consideration stage.

**3.5.1. Algorithm to Estimate Individual-Specific Part-Worths on the Fly.** For simplicity, we illustrate our approach using the example of a choice question with two product profiles and a "no-choice" option. The general principles apply to choice questions consisting of more than two profiles. Let us denote the two profiles in the $k^{\text{th}}$ choice question as $\mathbf{x}_j^{kA}$ and $\mathbf{x}_j^{kB}$. After a total of $T$ responses from the respondent (including both at the consideration and the choice stages), depending on respondent $j$'s choice among {profile A, profile B, none of the two}, we obtain the following information correspondingly:

$$\begin{aligned} \text{Chose A}: &\quad \mathbf{w}_j^T\left(\mathbf{x}_j^{kA}-\mathbf{x}_j^{kB}\right)\geq 0 \quad \text{and} \quad \mathbf{w}_j^T\mathbf{x}_j^{kA}\geq \alpha_j^T \\ \text{Chose B}: &\quad \mathbf{w}_j^T\left(\mathbf{x}_j^{kA}-\mathbf{x}_j^{kB}\right)< 0 \quad \text{and} \quad \mathbf{w}_j^T\mathbf{x}_j^{kB}\geq \alpha_j^T \\ \text{Chose None}: &\quad \mathbf{w}_j^T\mathbf{x}_j^{kA}< \alpha_j^T \quad \text{and} \quad \mathbf{w}_j^T\mathbf{x}_j^{kB}< \alpha_j^T \end{aligned} \tag{4}$$

As shown in Equation (4), we obtain two data points each time the respondent makes a choice. For all inequalities containing $\alpha_j^T$, the fuzzy membership of the labeled response can be obtained directly using Equation (3), with class center and radius estimates calculated from pooled responses from both the consideration and the choice stages. When the inequalities in Equation (4) entail utility comparison between the two profiles, we denote $\mathbf{x}_j^{kAB}=\left(\mathbf{x}_j^{kA}-\mathbf{x}_j^{kB}\right)$ and rewrite such inequalities as $\mathbf{w}_j^T * \mathbf{x}_j^{kAB}\geq or < 0$. Next, we assign a fuzzy class membership to each data point obtained, with $\mathbf{x}_j^{kAB}$ replacing $\mathbf{x}_j^i$ in Equation (3). Under such scenarios, the class center of each class captures the mean differences between the two profiles when one profile is favored over the other. Conceptually, if the position of $\mathbf{x}_j^{kAB}$ considerably deviates from its class center, it implies that the labeled response does not align with our current knowledge about the

respondent's product preferences. Therefore, we will assign a low class membership to such a response. Formally, the optimization we solve at this stage of the adaptive question design can be expressed as:

$$
\begin{aligned}
&\min_{\mathbf{w}_j^T,\alpha_j^T,\xi_j^i,\xi_j^{kAB}} \quad \left[\frac{1}{2}\left\|\mathbf{w}_j^T\right\| + C\left(\sum_{i=1}^{I'} u_j^i \xi_j^i + \sum_{k=1}^{K'} u_j^k \xi_j^{kAB}\right)\right] \\
(5) \qquad &s.t. \quad y_j^i\left(\mathbf{w}_j^T * \mathbf{x}_j^i - \alpha_j^T\right) \geq 1 - \xi_j^i \\
&\qquad\quad y_j^{kAB}\left(\mathbf{w}_j^T * \mathbf{x}_j^{kAB}\right) \geq 1 - \xi_j^{kAB} \\
&\qquad\quad \xi_j^i, \xi_j^{kAB} \geq 0
\end{aligned}
$$

with the first constraint denoting all labeled responses related to $\alpha_j^T$ (hence including responses from both consideration and choice stages) and the second constraint containing all responses related to utility comparison between the two profiles. It is evident that this optimization remains to be convex.

Similar to Section 3.4.1, we update our fuzzy membership estimates for all prior labeled responses (including the ones obtained in the consideration stage) each time after the respondent makes a choice among the two product profiles and the no-choice option. Consequently, our fuzzy membership estimates are refined over time as we accumulate additional knowledge about the focal respondent's product preferences.

**3.5.2. Algorithm to Adaptively Select Profiles in the Next Choice Question.** We discuss below how we identify the next set of profile pair (i.e., profile $g_1$ and profile $g_2$) to be shown to the respondent. In the choice stage, $m_j^{g1,g2} = \left|\mathbf{w}_j^T * (\mathbf{x}_{g1} - \mathbf{x}_{g2})\right|$, $m_j^{g1} = \left|\mathbf{w}_j^T * \mathbf{x}_{g1} - \alpha_j^T\right|$, and $m_j^{g2} = \left|\mathbf{w}_j^T * \mathbf{x}_{g2} - \alpha_j^T\right|$ capture our most recent estimates of respondent $j$'s relative preferences across the two profiles and the no-choice option. Based on these utility estimates, we first use the criterion $\min\left\{\max(m_j^{g1,g2}, m_j^{g1}, m_j^{g2})\right\}$ for all $\left\{(g_1, g_2) \in M_j\right\}$ to select a trial set of choice sets that

we would consider for the next choice question. The utilities of all options in this choice set are a priori near equal. This corresponds to the utility balance criterion commonly adopted in the conjoint literature. As suggested by various prior studies (e.g., Huber and Hansen 1986; Haaijer, Kamakura, and Wedel 2001; Toubia et al. 2004), utility balanced questions are the most informative in further refining the part-worths estimates. This criterion is also consistent with the simple margin approach proposed by Tong and Koller (2001), which aims to cut the feasible region of part-worths approximately in half.

Given that labeled responses are likely to be asymmetrically distributed and/or the feasible region of part-worths may be elongated, we employ a hybrid of simple margin and ratio margin approaches similar to the one described in Section 3.4.2. After using simple margin approach to obtain a trial set of choice sets, we calculate a ratio margin $\max\left(\max(\frac{m'^{g1,g2+}_j}{m'^{g1,g2-}_j},\frac{m'^{g1,g2-}_j}{m'^{g1,g2+}_j}),\max(\frac{m'^{g1+}_j}{m'^{g1-}_j},\frac{m'^{g1-}_j}{m'^{g1+}_j}),\max(\frac{m'^{g2+}_j}{m'^{g2-}_j},\frac{m'^{g2-}_j}{m'^{g2+}_j})\right)$ for each choice set in this trial set. We first give this choice set a hypothetical response of choosing profile $g_1$, calculating a new part-worths by combining this response with the prior obtained responses, and obtaining hypothetical margins $m'^{g1,g2+}_j$ and $m'^{g1+}_j$. We then give this choice set of a hypothetical response of choosing profile $g_2$ and obtain hypothetical margins of $m'^{g1,g2-}_j$ and $m'^{g2+}_j$. Similar approach is used to obtain $m'^{g1-}_j$ and $m'^{g2-}_j$ when the choice set is given a hypothetical response of no-choice. Then the pair of profiles with the smallest ratio margin (i.e., $\min_{s=1,\ldots,S}\left[\max\left(\max(\frac{m'^{g1,g2+}_j}{m'^{g1,g2-}_j},\frac{m'^{g1,g2-}_j}{m'^{g1,g2+}_j}),\max(\frac{m'^{g1+}_j}{m'^{g1-}_j},\frac{m'^{g1-}_j}{m'^{g1+}_j}),\max(\frac{m'^{g2+}_j}{m'^{g2-}_j},\frac{m'^{g2-}_j}{m'^{g2+}_j})\right)\right]$) is selected as the next choice set shown to the consumer. Similar to Section 3.4.2, conditional on satisfying the smallest ratio margin criterion, if two or more choice sets have the same ratio margins, the choice

set with the shortest overall distance to the corresponding class centers will be chosen to minimize efficiency loss. We repeat this process iteratively to shrink the feasible region of part-worths as rapidly as possible, until $Q_2$ questions are asked.

## 4. Empirical Investigation

In this section we describe two empirical studies involving digital cameras and computer tablets. Our pretest indicates that both product categories are of interest to the respondents' population (undergraduate students). The overall complexity of the digital camera category (with 30+ attribute levels) is parallel to product categories studied in prior research that elicits consumers' preferences for complex products (e.g., Park et al. 2008; Netzer and Srinivasan 2011; Scholz et al. 2010). The computer tablet category (with 70+ attribute levels) is considerably more complex than the ones used in extant methods, particularly in the context of decompositional preference elicitation methods. To keep our empirical applications meaningful and realistic, we conducted pretests to choose a set of attributes that the respondents typically consider. We then used retail websites such as Bestbuy.com and Amazon.com to identify the ranges and values of attribute levels used in both empirical applications.

### 4.1. Digital Camera Study

**4.1.1. Research Design.** A total of 425 participants are randomly assigned to one of the six preference measurement conditions. We included two conditions of the fuzzy SVM method: Condition 1: fuzzy SVM with collaborative filtering; Condition 2: fuzzy SVM without collaborative filtering. The overall flow of the two fuzzy SVM conditions follows Figure 1, with the exception that in Condition 2 the initial part-worths is attained solely based on the focal respondent's self-configured product profile. We further compare the predictive validity of the proposed method with the following four benchmark methods: Condition 3: the self-explicated

method; Condition 4: an upgrading method similar to Park et al. 2008 with no incentive alignment; Condition 5: the adaptive Choice-Based Conjoint (ACBC); and Condition 6: the traditional Choice-Based Conjoint (CBC). The list of attributes and attribute levels included in our digital camera study is provided in Table A1 in Web Appendix D.

Similar to prior studies in the literature (e.g., Scholz, et al. 2010; Netzer and Srinivasan 2011), participants in all conditions first complete the preference measurement task, followed by an external validation task and a post-survey feedback task. Identical across the six experimental conditions, the external validation task comprises two choice questions, each including two camera profiles. Generated using fractional factorial design, the profiles are carefully chosen so that one profile does not clearly dominate the other in each choice set.

**4.1.2. Results.** We obtain the individual-specific part-worths estimates from the two fuzzy SVM conditions using the fuzzy SVM estimation algorithm. The self-explicated estimates are obtained by multiplying the attribute importance weights with the corresponding desirability ratings (Srinivasan 1988). And hierarchical Bayesian estimation was used to obtain individual-level part-worths estimates from the upgrading method, the ACBC method, and the CBC method.

Following the tradition in this literature (e.g., Evgeniou et al. 2005; Netzer and Srinivasan 2011; Park et al. 2008), we use the hit rate of the external validity tasks to gauge the predictive validity of the six preference measurement methods (Table 1).[7] We find that the two fuzzy SVM

[7] We also tried to incorporate the Kullback-Leibler (KL) measure used by Dzyabura and Hauser (2011), Ding et al. (2011), and Hauser et al. (2014) in our digital camera application. We discovered that this measure is only applicable with three or more validation tasks. Our digital camera application includes two choice validation tasks. In such cases, when the consumer chooses the profile on the left in one task and the profile on the right in the other task, the KL measure equals to zero regardless whether both predictions are correct, wrong, or one being correct and one being wrong. Essentially, the KL measure does not discriminate whether observed and predicted choices are aligned when the total number of validation tasks is two. In our computer tablet study, we included six choice validation tasks and were able to use the KL measure to gauge predictive validity.

conditions perform significantly better than all benchmark methods in correctly predicting respondents' choices in hold-out tasks ($p < .05$).

Comparing hit rates across the two fuzzy SVM conditions, we find that collaborative filtering improved predictions but not significantly (.801 vs. .779, $p > .05$). This finding is consistent with empirical results from Dzyabura and Hauser (2011). It is possible that, after all the consideration and choice questions are queried in our adaptive survey, incremental benefits from the collaborative filtered initial part-worths have diminished. This matter is explored further in synthetic data experiments in Section 5.4.

We also compared participants' responses to the post survey feedback questions across the six experimental conditions. Overall, participants provided more favorable feedback to the format and questions generated under the fuzzy SVM conditions than those from the benchmark methods (Table A2 in Web Appendix D).[8]

<Insert Table 1 about here>

### 4.2. Computer Tablet Study

**4.2.1. Research Design.** In a second empirical study, we examine the performance of the proposed method for the more complex product category of computer tablets (with 70+ attribute levels). The complete list of attributes and attribute levels included in this study can be found in Figure A1 in Web Appendix B. In addition to the different product category and the increased number of attribute levels, we also made the following modifications in this empirical application. *First*, given the high complexity of the product category, we included a warm-up task so that the participants can get familiar with the different attribute levels before the preference measurement task. Such task has been shown to improve the accuracy of preference

[8]In both empirical applications, we also recorded the amount of time it takes for each participant to complete the survey. Such information is provided in Tables A3 and A4 in Web Appendix D.

elicitation (Huber et al. 1993). Specifically, we provided the list of 14 attributes used to describe the computer tablets, followed by a brief verbal and graphic description for each attribute level, displayed for one attribute at a time.

*Second*, as a number of prior studies (e.g., Ding et al. 2005; Ding 2007; Ding et al. 2009; Dong et al. 2010; Ding et al. 2011; Hauser et al. 2014) suggest that incentive alignment offers benefits such as greater respondent involvement, less boredom, and higher data quality, we incorporated incentive alignment in this application. At the beginning of the experiment, we told the participants that we would award a computer tablet device to one randomly selected participant from this study, plus cash representing difference between the price of the tablet device and $900. We set the $900 as the maximum prize value because the vast majority of computer tablets cost less than $900 at the time of our study. The participants were told that the total number of participants for this study would be approximately 150 individuals (i.e., the chance of winning is about 1 in 150). Because we wanted both the preference elicitation tasks and the validation tasks to be incentive aligned, the participants were told that we would randomly decide which of the two tasks to use when determining the final prize. We also told each participant that, if chosen as a winner, he would receive a computer tablet based on: 1) either his choice from one of the validation questions; or 2) his most preferred tablet among a list of 25 tablets, inferred from his answers to the preference elicitation questions. Following Ding et al. (2011) and Hauser et al. (2014), participants were told that this list was pre-determined by the researchers and it would be made public after the study. Therefore, the respondents have incentives to answer the questions carefully and truthfully.

*Lastly*, we modified our validation procedure relative to the digital camera study. We included both initial and delayed validation tasks as in Ding et al. (2011) and Hauser et al. (2014).

Following Hauser et al. (2014), the delayed validation questions were sent to the respondents by email one week after the preference measurement task. Additionally, we included both pre- and post- preference measurement validation questions as suggested by Netzer and Srinivasan (2011). Each respondent answered six validation choice questions, with three before the preference measurement task and three in the delayed validation task. As pointed out by Netzer and Srinivasan (2011), the standard validation task procedure as the one employed in our digital camera study might be susceptible to idiosyncrasies of the chosen validation questions. In the computer tablet study, we followed Netzer and Srinivasan (2011) by including a broader set of validation choice questions in the validation task. We first used fractional factorial design to generate a set of orthogonal balanced choice questions. Next, we scanned through the generated questions and only retained those comprising profiles available in the marketplace (as any one of the computer tablets in the validation questions can be potentially awarded to a participant). To allow for appropriate comparison across conditions, we also followed Netzer and Srinivasan (2011) by using the same sets of randomly drawn validation questions in all conditions.

In this study, a total of 151 participants are randomly assigned to one of the four preference measurement conditions. To better understand the incremental benefit from including consideration questions in our adaptive question design, we compared the proposed method (Condition 1) to an alternative fuzzy SVM method in which the choice questions are followed immediately after the self-configuration task and the unacceptable and must-have questions (Condition 2). Furthermore, we included the self-explicated method (Condition 3) and the ACBC method (Condition 4) to replicate results from the first empirical application.

The flow of each experimental condition is described below. *First*, the participant was introduced to the study along with a basic description of the incentive alignment mechanism.

*Second*, the participant was presented with the warm-up task. *Third*, three validation questions are shown to the participant, with each consisting of two tablet profiles. *Fourth*, the participant was presented with the preference measurement task (which varies by experimental condition). *Lastly*, a week later, the participant was followed up by an email with a survey link to a second set of three validation questions.

**4.2.2. Results.** Table 2 reports the predictive validity of the four experimental conditions. In addition to the traditional hit rate measure, we used the Kullback-Leibler (KL) divergence measure to gauge the degree of divergence from predicted choices to those that are observed in the validation data. The KL divergence measure is an information-theory-based measure of divergence (Kullback and Leibler 1951; Chaloner and Verdinelli 1995). Dzyabura and Hauser (2011), Ding et al. (2011), and Hauser et al. (2014) demonstrate that, for consideration data, the KL divergence measure provides an evaluation of predictive ability that is rigorous and discriminates well. We followed the formulae provided in Dzyabura and Hauser (2011) to calculate the KL divergence measures. In our context, we conceptualize a false-positive prediction as the case in which the respondent is predicted to choose a profile but did not actually choose it in a validation question; and a false-negative prediction as the case that the respondent is predicted not to choose a profile but actually chose it in a validation question. Since this measure evaluates divergence from perfect prediction, a smaller KL divergence measure indicates better model prediction.

<Insert Table 2 about here>

Consistent with findings from our digital camera application, the proposed method exhibits superior predictive ability when compared to the self-explicated and ACBC methods in both the initial and delayed validation tasks. We also find that the proposed method has smaller

KL divergence than the fuzzy SVM condition without consideration questions (marginally significant at $p < .1$). It is evident that the inclusion of consideration questions is helpful in improving preference elicitation in our context. It is possible that, when consideration questions are absent, respondents may encounter greater cognitive difficulty in making accurate trade-offs of choice questions. Respondents may also be less likely to incur early response errors (which is known to be detrimental in adaptive question design), when they are presented with the less demanding consideration questions prior to the more cognitively challenging choice tasks.

We also compared the KL divergence measure from the initial validation task with that from the delayed validation task within each experimental condition. No significant differences are observed. Therefore, when pooling results across the initial and the delayed validation tasks for each individual, the KL divergence measures follow the same pattern as the one discussed above.[9] The hit rate comparisons are also provided in Table 4. The proposed method also outperforms the three benchmark methods in terms of hit rate.

One of the main advantages of adaptive question design is the opportunity to reduce respondents' cognitive burden by asking fewer questions. We further examine the out-of-sample performance of the proposed method when only the first $k$ questions are used for each respondent (Table 3). We find that both the KL divergence and the hit rate measures gradually improve with the inclusion of self-configurator ($k = 1$), consideration questions ($k = 2$ to 9 with 8 screens of consideration questions with 5 profiles on each screen), and choice questions ($k = 10$ to 34), indicating that all these questions positively contribute to our preference elicitation task. Table 3 also reveals that the proposed method performs well with much fewer choice questions. In particular, the predictive validity after only 6-8 choice questions (i.e., $k = 15$ or 17) is already

[9] Note that the values of the KL measure depend on the number of the validation tasks (Hauser et al. 2014). Therefore, the pooled KL measures differ in magnitude from those based on initial or delayed validation tasks only.

similar to the predictive validity with all the 25 choice questions (i.e., $k$ =34). Indeed, after only the first 8 choice questions (i.e., $k$ = 17), the proposed method already exhibits significantly better predictive validity than both the self-explicated method and the ACBC method. This finding suggests that respondent burden in this application may be substantially reduced with a much shorter survey. This is consistent with Netzer and Srinivasan (2011) who also report negligible improvement in predictive validity after 5-7 adaptive paired comparison questions.

<Insert Table 3 about here>

Overall, in comparing the predictive ability of the proposed method with that of the ACBC and self-explicated methods, our computer tablet application replicated results from the digital camera application. The comparison between the two fuzzy SVM conditions also reveals that the inclusion of consideration questions is useful in facilitating preference elicitation in our context. In addition, this application illustrates that the proposed method scales well when the focal product category is considerably more complex than the ones used in prior studies.

## 5. Synthetic Data Experiments

In this section we describe a series of synthetic data experiments we conduct to complement our empirical investigation. In Section 5.1, we compare the performance of the proposed question selection algorithm with that of Dzyabura and Hauser (2011) for consideration questions when the true consideration decisions are conjunctive. In Section 5.2, we examine the performance of the proposed method with that of benchmark methods when the true consideration and choice decisions are both based on a part-worths model. Under this comparison, we use the question selection algorithm in Dzyabura and Hauser (2011) as the benchmark question selection method for consideration questions and that in Abernethy et al. (2008) as the benchmark method for choice questions. To investigate the applicability of these question selection methods to high-

dimensional problems, in both Sections 5.1 and 5.2, we examine the scalability of each algorithm when the focal product category is equipped with up to 100+ attribute levels. To ensure fair comparisons across methods, only the focal respondent's responses are used in the adaptive question selection in all comparisons described above. We also test the upper bound of parameter recovery and predictive validity by assuming no response errors in such comparisons. In Section 5.3, we compare the performances of the fuzzy SVM vs. soft margin SVM active learning with and without response errors. In Section 5.4, we examine the improvements of collaborative-filtered initial part-worths over non-informative initial part-worths on parameter recovery and predictive validity. We report our main findings below. Additional implementation details can be found in Web Appendix E.

### 5.1. Using Proposed vs. Conjunctive Question Selection When the True Consideration Decisions are Conjunctive

In Section 3 we describe a framework where a set of part-worths is used to characterize consumer's responses to both consideration and choice questions. In the literature, conjunctive-like criteria are often used to examine answers to consideration questions (e.g., Bettman 1970; Hauser et al. 2010). Therefore, we carry out synthetic data experiments to examine the performance of the proposed question selection algorithm for consideration questions when the true consideration model is conjunctive. Specifically, the adaptive question selection algorithm in Dzyabura and Hauser (2011) is used as the benchmark method in this comparison.

Given our emphasis on complex products, we examine three scenarios in which the focal product category comprises 15, 25, and 35 attributes with 3 levels each. Under each scenario, we simulate 300 synthetic respondents (i.e., 300 conjunctive decision rules) as described in Dzyabura and Hauser (2011). We then perform active learning for each individual where 40 consideration questions are adaptively selected, with each synthetic respondent labeling the

profile as "would consider" or "would not consider" based on the underlying true conjunctive decision rule. In order to compare the relative performance of the two question selection methods, we generate 3,000 validation profiles for each synthetic respondent, with the respondent considering 1,500 profiles and not considering the remaining under the individual's true underlying conjunctive decision rule. Because each synthetic respondent considers 50% of the validation profiles, the null model that predicts "randomly considers profiles" would achieve a hit rate of 50%. Therefore, while the hit rate measure can be misleading for consideration data empirically, it provides a valid measure of predictive validity in our synthetic setting.

Table 4 provides comparison results. To compare the question selection methods, we keep the estimation method constant. Specifically, we present the comparison results using both the conjunctive estimation method as in Dzyabura and Hauser (2011) and those using the fuzzy SVM estimation method proposed in this paper. In terms of performance metrics, hit rate, KL divergence, and $U^2$ are used to measure predictive validity and parameter recovery (Table 4a). $U^2$ is an information-theoretic measure of parameter recovery (Hauser 1978). It measures the percentage of uncertainties explained by the model, with $U^2$ = 100% indicates perfect parameter recovery.[10] To examine scalability of these two question selection methods, we also report the average time it takes to generate the next question in seconds (Table 4b). The reported

[10]While the original $U^2$ measure in Hauser (1978) is based on choice probabilities, the fuzzy SVM estimation gives rise to dichotomous (e.g., consider vs. not consider; choose product A vs. product B) rather than probabilistic predictions. Therefore, when the fuzzy SVM algorithm is used for model estimation, we calculated the $U^2$ measure based on a logit transformation, with the deterministic component of the product utility calculated from the estimated part-worths given by the fuzzy SVM algorithm. Because the scale of utility estimates matters in the magnitude of $U^2$ (if we multiply the part-worths estimates by a constant, larger part-worths results in more extreme choice probabilities, hence more extreme $U^2$ estimate), we normalize the estimated part-worths to the scale of the true part-worths and use the relative $U^2$ (the $U^2$ calculated from the fuzzy SVM part-worth estimates divided by the $U^2$ calculated from the true part-worths) to remove the effect of scaling. This measure was not used in our empirical studies because the true part-worths are unknown empirically and it is ambiguous regarding how to determine the baseline scale in our empirical comparisons.

computing times are all based on Matlab code run on an Intel 3.2GHz personal computer with Windows 7 Operating System.

<Insert Table 4 about here>

When the estimation method in Dzyabura and Hauser (2011) is used (i.e., for each attribute level, the model estimates a probability for which the respondent finds it acceptable), the question selection algorithm by Dzyabura and Hauser (2011) exhibits superior hit rate, KL divergence, and $U^2$ than the focal question selection algorithm in almost all problem instances. Indeed, when the product dimension is relatively moderate (15 attributes and 3 levels), Dzyabura and Hauser (2011)'s method yields perfect parameter recovery and holdout prediction after 40 questions. Such results reveal that, when the true consideration model is conjunctive, the algorithm proposed by Dzyabura and Hauser (2011) works exceptionally well in estimating the probability for which the respondent finds each attribute level acceptable.

Because our approach aims to obtain a set of part-worths estimates for each individual, we also compare the performance of the two question selection methods when the fuzzy SVM method is used for estimation. In such cases, an individual-specific part-worths vector is estimated after all 40 questions are queried under each question selection algorithm. The estimated part-worths are then used to predict the validation profiles. Interestingly, we discover that, when the fuzzy SVM estimation is used, the proposed method outperforms the method by Dzyabura and Hauser (2011) in terms of the hit rate, KL divergence, and $U^2$ measures. Such results are likely to be driven by the fact that the algorithm by Dzyabura and Hauser (2011) is specifically developed for uncovering the probability for which the respondent considers each attribute level, rather than a utility estimate associated with the attribute level. Therefore, when

the primary focus is to estimate part-worths, this method does not perform as well, even when the true consideration model is conjunctive.

We further study scalability of the two methods by examining the average time it takes to generate the next question under each question selection algorithm. When the proposed question selection algorithm is used, on average it takes less than 0.25 second to generate the next question in all scenarios. In contrast, under the question selection algorithm by Dzyabura and Hauser (2011), the average time it takes to generate the next question are considerably longer (ranging from 2.856 seconds to 11.861 seconds in the three scenarios). Note that such discrepancies may diminish considerably if we were to optimize and/or code both algorithms in a more computationally efficient language such as C or C++.

Overall, our synthetic data experiments reveal the following. First, even when the true consideration model is conjunctive, the part-worths estimate from the proposed method exhibit a reasonably good ability to predict whether the respondent would consider a profile. Given that both the part-worths model of consideration and the conjunctive model of consideration as in Dzyabura and Hauser (2011) aim to capture a set of linear decision rules in the consideration process, we believe that such findings are quite reasonable. Second, the proposed framework is not restricted to the question selection algorithm discussed in Section 3.4. Indeed, the algorithm by Dzyabura and Hauser (2011) could be used in conjunction with the proposed algorithm in a consider-then-choice framework to uncover both consideration heuristics and conjoint part-worths. In such cases, different utility functions may also be employed to model consumer's consideration and choice decisions separately.

### 5.2. Performance Comparisons When True Consideration and Choice Decisions are Both Based On a Part-Worths Model

We now compare the performance of the proposed method with that of benchmark methods when the true consideration and choice decisions are both based on a part-worths model. For simplicity, we examine the case where the same utility function is used for both consideration and choice decisions. If different part-worths models were specified for consideration and choice questions, we expect that our key findings would not change qualitatively.

In each scenario under study, we perform active learning where 40 consideration questions are adaptively designed for each individual, followed by 25 adaptive choice questions with 2 alternatives each. In the first benchmark condition, the question selection algorithm by Dzyabura and Hauser (2011) is used for the adaptive design of consideration questions. In the second benchmark condition, the question selection algorithm by Abernethy et al. (2008) is used for the adaptive design of choice questions. Because the algorithm by Dzyabura and Hauser (2011) is designed for consideration questions only and that by Abernethy et al. (2008) is for choice questions only, the fuzzy SVM active learning is used as the question selection method for choice questions in the first benchmark condition and for consideration questions in the second benchmark condition to ensure fair comparison. For similar reasons, the fuzzy SVM method is used as the estimation method after all questions are queried in all three conditions.

Similar as before, we examine three scenarios in which the focal product category comprises 15, 25, and 35 attributes with 3 levels each. 300 synthetic respondents are simulated in each scenario. The part-worths for each respondent are randomly generated with ($-\beta$, 0, $\beta$) for each attribute, with $\beta \sim N$ (0.5, 1.5). To compare performances across conditions, 3,000 validation questions consisting of two profiles are generated for each respondent, with the respondent choosing the profile on the left 50% of the time and choosing the profile on the right

in the remaining questions. Under this setup, the null model for the hit rate and $U^2$ is the one that predicts "randomly choose among the two profiles". In addition to the hit rate, KL divergence, and $U^2$ measures, we also use MAE (mean-absolute-error) and RMSE (root-mean-square-error) to measure the ability of each question selection method in recovering the true part-worths. For comparability across methods, we normalize the estimated part-worths to the scale of the true part-worths in each problem instance.

Our comparison results are shown in Table 5. When the true consideration and choice decisions are both based a part-worths model, the proposed question selection method outperforms the two benchmark methods in both parameter recovery and predictive validity (Table 5a). It is not surprising that the question selection algorithm by Dzyabura and Hauser (2011) does not work as well in this setting, as the algorithm is specifically developed to uncover consideration heuristics when the true consideration decisions follow a set of conjunctive rules, rather than a part-worths model. Meanwhile, the lack of performance from the question selection algorithm by Abernethy et al. (2008), particularly in high dimensional problems, is likely to be related to the discrete transformation required by its gradient-based algorithm when used for a large number of discrete attributes in our setting.[11]

<Insert Table 5 about here>

[11] While we obtain significantly better performance from the proposed method, the hit rate differences between the proposed method and Dzyabura and Hauser (2011) are rather small in the case of 35 attributes with 3 levels each (.901 vs. 0.893 as in Table 5a). We further examine the percentages of times that the estimated part-worths from the two methods could correctly predict the most preferred attribute level in each attribute. We do not find significant differences between the two methods in this case (both methods are able to correctly predict the most preferred attribute levels about 86% of the time). We also compare the MAPEs (mean absolute percentage error) between the predicted and actual part-worths from the two methods. Consistent with findings from the MAE and RMSE measures in Table 5, the proposed method provides more accurate part-worths estimates than Dzyabura and Hauser (2011) in all cases. Managerially, if the primary focus of the firm is to identify the most preferred attribute level in each attribute, we think that such differences in hit rates do not entail additional insights. Nevertheless, if the firm's central goal is to obtain precise forecast of market share or product profit, we believe that the improved accuracy in our part-worths estimates would be beneficial.

We also report the average time it takes to generate the next question under each question selection method in Table 5b. Because the algorithm by Dzyabura and Hauser (2011) is used for consideration questions only and that by Abernethy et al. (2008) is used for choice questions only, the corresponding average time is reported in this table. When it comes to consideration questions, the algorithm by Dzyabura and Hauser (2011) takes approximately 3 or more seconds on average to generate the next question (again, the computing speed may improve considerably if the code were optimized and/or written in a more computationally efficient language). Regarding choice questions, the algorithm by Abernethy et al. (2008) is very fast computationally because the solution used to generate the next question can be derived in closed form. Therefore, when the focal product consists of a large number of continuous attributes, this method is a good alternative for adaptive design of choice questions.

### 5.3. Performance Comparisons between Fuzzy SVM and Soft Margin SVM Active Learning With and Without Response Errors

A key advantage of the proposed method is its ability to gauge response errors on the fly. In this section, we investigate the use of the fuzzy SVM active learning vs. that of the soft margin SVM active learning without fuzzy membership probabilities. Because both methods scale well in high dimensional problems, we examine the case where the focal product category consists of 35 attributes with 3 levels each. We use similar methods as in Section 5.2 to simulate the true part-worths and the holdout profiles for the 300 synthetic respondents used in each problem instance.

We first investigate performance comparisons of the two methods when there are response errors. Given that the positions of response errors play an integral role in the performance of adaptive question design, we examine the two question selection methods' abilities to recover true part-worths and to predict holdout profiles under the scenarios of 1) early vs. 2) middle vs. 3) late response errors. To ensure fair comparisons, we set the instances of

response errors to be 15% across the three scenarios. In the early/middle/late response errors scenario, all response errors occur during the first/middle/late 1/3 of adaptive questions. For each synthetic respondent, we use random draws from a uniform distribution to determine the positions of the errors. Within each scenario, we hold error positions constant across the two question selection methods so that our results are comparable.

Our performance comparisons are reported in Table 6a.[12] When response errors take place during early or middle portions of the adaptive study, the fuzzy SVM active learning method exhibits superior ability to recover the true part-worths and to predict holdout questions than the soft-margin SVM active learning. Nevertheless, if response errors occur towards the end of the adaptive survey, both question selection methods perform similarly. This pattern is quite reasonable because early response errors are in general more detrimental than errors taking place later in adaptive question design. Given its primary goal of alleviating negative effect from response errors on the fly, the fuzzy SVM active learning method provides the most improvement over the soft-margin SVM method when the impact of response errors is salient. In contrast, because the negative effect from response errors is limited when errors take place towards the end of the adaptive question selection, the advantage from using fuzzy SVM over soft margin SVM diminishes correspondingly.

<Insert Table 6 about here>

We also examine performances of the two methods when respondents do not make any response error. As expected, the use of the fuzzy SVM active learning incurred an efficiency loss,

[12] For simplicity, we report the performance metrics with 40 adaptive questions where each synthetic respondent labels the profile as “would consider” or “would not consider”. Similar patterns are found when choice questions are added after the consideration questions. In line with Table 6a, the advantage of the fuzzy SVM active learning is the most salient when errors take place towards the beginning and middle portions of the consideration/choice questions. We also conducted similar comparisons under varying levels of error instances and product dimensions. We find that the general results hold qualitatively as long as the error instance is not excessive (if the respondents incur too many errors, neither method can effectively recover the true part-worths).

originated from the less than perfect fuzzy membership probabilities assigned to the correctly labeled support vectors (Table 6b). Nevertheless, such an efficiency loss is relatively minor, possibly due to the fact that that active learning is inherently less challenging in the absence of response errors.

**5.4. Tests of Improvements over Non-informative Initial Part-worths**

In the synthetic experiments above, we only utilize the focal respondent's information in the design of adaptive questions. In this section, we examine the use of collaborative-filtered initial part-worths vs. that of non-informative initial part-worths in parameter recovery and predictive validity. Because the initial individual-specific part-worths in our proposed framework is obtained via collaborative filtering between the focal and previous respondents' self-configured product profiles, we also investigate whether the degree of heterogeneity in synthetic respondents' configurators play a role in the usefulness of incorporating data from other respondents. Intuitively, if all respondents have the same product preferences and self-configure the same profile, past respondents' data should be quite informative in determining the focal respondent's initial part-worths. In contrast, when respondents differ vastly in their most favorite product profiles, collaborative filtering may not be very helpful.

We consider the following two scenarios of homogenous vs. heterogeneous configurators accordingly. In the homogenous case, we simulate 300 synthetic respondents with identical part-worths (hence identical self-configured profile). In the heterogeneous case, we consider 300 synthetic respondents with all different self-configurators. By excluding perfect over-lap in the focal and previous respondents' self-configured profiles, the average cosine similarity ( $s(j, j')$ in Equation 1) in the heterogeneous case is 0.331.

In each case described above, we consider a baseline model where non-informative initial part-worths are used at the outset of the adaptive question design. Instead of utilizing information from both the focal and previous respondents' configurators as in the two collaborative filtering conditions, the non-informative initial part-worths are obtained by randomly querying one profile that the synthetic respondent would consider and one profile the respondent would not consider from the training data. We then compare improvements obtained from collaborative-filtered initial part-worths over non-informative initial part-worths in the respective cases of homogenous vs. heterogeneous configurators. In both comparisons, we examine the scenario where the focal product category consists of 35 attributes with 3 levels each. The adaptive question design is based on 40 consideration questions and 25 choice questions as in Section 5.2. We follow the same procedure as described in Section 5.2 to construct the 3,000 validation questions for each synthetic respondent.

Table 7 provides comparison results from these two cases as a function of the number of questions queried. Consistent with our conjecture, at the outset of the adaptive question design, incremental benefits from the collaborative-filtered vs. the non-informative initial part-worths are considerably more salient in the homogenous configurator case. This finding is rather intuitive because previous respondents' estimated part-worths are quite information rich if all respondents have the same product preferences. Interestingly, Table 7 also reveals that, in both cases, the advantages from collaborative-filtered initial part-worths diminish during the course of the adaptive question survey. Indeed, after the respondents answer about 40 consideration questions, the benefits from the collaborative-filtered initial part-worths become more or less negligible. Such finding implies that, analogous to the role of priors in the Bayesian literature, benefits from informative initial part-worths can lessen considerably as more data become

available. Similar results hold when we employ an alternative baseline model where the initial part-worths vector is calculated based on the focal respondent's configurator alone. These synthetic experiments suggest that, collaborative-filtered initial part-worths in the proposed method may only be beneficial when consumers exhibit similar product preferences and practical concerns preclude longer surveys.

<Insert Table 7 about here>

## 6. Conclusions

In this paper, we propose an adaptive decompositional framework to elicit consumers' preferences for complex products. Our research suggests that the proposed method has the potential to provide the following new capabilities to complement existing preference elicitation methods. *First*, compared to extant methods, the proposed algorithm is particularly suitable for high-dimensional problems. Our empirical and synthetic studies demonstrate that the proposed framework is able to rapidly and effectively elicit individual-level preference estimates for product categories equipped with 70-100 attribute levels. This is typically considered prohibitive for decompositional preference elicitation methods. *Second*, we demonstrate that the fuzzy SVM active learning method provides a natural remedy for a long-standing challenge in adaptive question design by gauging the possibility of response errors on the fly and incorporating it into the survey design. We illustrate via synthetic data experiments that, the proposed algorithm is particularly effective when response errors take place towards the beginning/middle portions of adaptive questions. *Lastly*, while most adaptive question selection methods only utilize information from the focal respondent, our research explores in a live setting how previous-respondent data may be used to assist active learning of the focal respondent's product preferences. Overall, our research suggests that the proposed approach is a promising new

method that can be used to complement extant preference elicitation methods, particularly in the context of complex products.

Our research is also subject to limitations and suggests promising avenues for future research. First, while we use part-worths models to characterize consumers' responses to both consideration and choice questions, our algorithm does not explicitly uncover decision heuristics used by consumers. Future research may adapt the proposed algorithm to directly capture such heuristics. In such cases, the SVM classification would be performed at the product feature level, rather than at the product level as in the proposed method. Separate utility functions may also be used in the consideration and choice stages if the underlying decision rules for the consideration phase versus the choice phase are known to be different.

Second, while the current research is among the first efforts to explore the use of previous respondent's data in complex product preference elicitation in a live setting, we only observe incremental benefits of collaborative filtering at the outset of the adaptive question survey. Future research may consider alternative approaches to take better advantages of this technique. For example, richer covariates such as demographic, social-economic, and/or product usage information may be incorporated into collaborative filtering. Furthermore, advantages from using other respondents' data might become more salient if researchers were to incorporate collaborative filtering into the entire course of adaptive question design rather than only in the selection of initial questions.

Lastly, while the proposed algorithm is flexible enough to accommodate nonlinear utility functions, the specifications of such nonlinear utility functions are ad hoc by nature. Managerial insights and/or pretests are needed to determine the exact form of the kernel function. If an inappropriate kernel is used, response errors may be undistinguishable from the incorrectly

specified utility form. Therefore, prior to the adaptive question design, managerial consultation and/or pretests are needed to carefully specify the exact utility model to be estimated.

With regards to extensions, future research may further explore the use of semi-supervised active learning in marketing context, particularly in the area of adaptive question design. For example, a common challenge faced by adaptive question design is the lack of labeled data points, particularly at the beginning of the survey. The basic idea of semi-supervised active learning is to iteratively identify unlabeled data points that are similar to labeled data, and to assign pseudo labels to such points so that the training data set can be enlarged. Recent research has shown that such efforts can effectively alleviate the problem of small-sized training data (e.g., Wu and Yap 2006; Hoi et al. 2009; Leng et al. 2013). Future research may further explore how to utilize such methods to improve extant adaptive question design, or in any marketing context where individual-level consumer data are relatively sparse. Additionally, if consumer responses are classified in multiple categories (e.g., not preferred; neutral; preferred), researchers can leverage recent methods that use support vector machine classifier for active learning of multiclass classification (e.g., Patra and Bruzzone 2012). We leave such endeavors as fruitful areas for future research.

## References

Abernethy J, Evgeniou T, Toubia O, Vert JP (2008) Eliciting Consumer Preferences Using Robust Adaptive Choice Questionnaires. *IEEE Trans. Knowledge and Data Engineering* 20(2):145-155.

Bettman J (1979) *An Information Processing Theory of Consumer Choice* (Addison-Wesley, Reading, MA).

Breese JS, Heckerman D, Kadie C (1998) Empirical Analysis of Predictive Algorithms for Collaborative Filtering. *In Proc. 14th Conf. Uncertainty in Artificial Intelligence* 43-52.

Chaloner K, Verdinelli I (1995) Bayesian Experimental Design: A Review. *Statistical Sci.* 10(3):273-304.

Corts C, Vapnik V (1995) Support-Vector Networks. *Machine Learning*, 20(3):273-297.

Cui D, Curry D (2005) Prediction in Marketing Using the Support Vector Machine. *Marketing Sci.* 24(4), 595-615.

Dong JX, Krzyzak A, Suen CY (2005) Fast SVM Training Algorithm with Decomposition on Very Large Data Sets. *IEEE Trans. Pattern Analysis and Machine Intelligence* 27(4):603-618.

Dong S, Ding M, Huber J (2010) A Simple Mechanism to Incentive-Align Conjoint Experiments. *Internat. J. Res. Marketing*, 27(1):25-32.

Ding M (2007) An Incentive-Aligned Mechanism for Conjoint Analysis. *J. Marketing Res.* 44 (2):214-223.

Ding M, Grewal R, Liechty J (2005) Incentive-Aligned Conjoint Analysis. *J. Marketing Res.* 42(1):67-82.

Ding M, Hauser JR, Dong S, Dzyabura D, Yang Z, Su C, Gaskin SP (2011) Unstructured Direct Elicitation of Decision Rules. *J. Marketing Res.* 48(1):116-127.

Ding M, Park YH, Bradlow ET (2009) Barter Markets for Conjoint Analysis. *Management Sci.*55(6):1003-1017.

Dzyabura D, Hauser JR (2011) Active Machine Learning for Consideration Heuristics. *Marketing Sci.* 30(5):801-819.

Evgeniou T, Boussios C, Zacharia G (2005) Generalized Robust Conjoint Estimation. *Marketing Sci.* 24(3):415-429.

Evgeniou T, Pontil M, Toubia O (2007) A Convex Optimization Approach to Modeling Consumer Heterogeneity in Conjoint Estimation. *Marketing Sci.* 26(6):805-818.

Green PE, Srinivasan VS (1990) Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice. *J. Marketing* 54(4):3-19.

Guyon I, Matić N, Vapnik VN (1996) *Discovering Information Patterns and Data Cleaning* (MIT Press, Cambridge, MA).

Haaijer R, Kamakura WA, Wedel M (2001) The no-choice alternative in conjoint choice experiments. *Internat. J. Market Res.* 43(1):93-106.

Hauser JR (1978) Testing the Accuracy, Usefulness, and Significance of Probabilistic Choice Models: an Information-theoretic Approach. *Oper. Res.* 26(3):406-421.

Hauser JR, Dong S, Ding M (2014) Self-Reflection and Articulated Consumer Preferences. *J. Product Innovation Management* 31(1):17-32.

Hauser JR, Rao VR (2003) Conjoint Analysis, Related Modeling, and Applications. Wind Y Green PE eds. *Marketing Research and Modeling: Progress and Prospects* (Springer, New York).

Hauser JR, Toubia O (2005) The Impact of Utility Balance and Endogeneity in Conjoint Analysis. *Marketing Sci.* 24(3):498-507.

Hauser JR, Toubia O, Evgeniou T, Befurt R, Dzyabura D (2010) Disjunctions of Conjunctions, Cognitive Simplicity, and Consideration Sets. *J. Marketing Res.* 47(3):485-496.

Heo G, Gader P (2009) Fuzzy SVM for Noisy Data: A Robust Membership Calculation Method. *2009 Fuzzy Systems Conf. Proc.* 431-436.

Herbrich R, Graepel T, Campbell C (2001) Bayes Point Machines. *J. Machine Learning Res.* 1: 245-279.

Hoi, SC, Jin R, Zhu J, Lyu MR (2009) Semisupervised SVM Batch Mode Active Learning with Applications to Image Retrieval. *ACM Transactions on Information Systems (TOIS)* 27(3) 16.

Huber J, Hansen D (1986) Testing the Impact of Dimensional Complexity and Affective Differences in Adaptive Conjoint Analysis. *Adv. Consumer Res.* 14(1):l59-l63.

Huber J, Wittink DR, Fiedler JA, Miller R (1993) The Effectiveness of Alternative Preference Elicitation Procedures in Predicting Choice. *J. Marketing Res.* 30(1):105-114.

Johnson RM, Orme BK (2007) A New Approach to Adaptive CBC. *Sawtooth Software Research Paper Series*.

Leng Y, Xu X, and Qi G (2013) Combining Active Learning and Semi-supervised Learning to Construct SVM Classifier. *Knowledge-Based Systems* 44: 121-131.

Lin CF, Wang SD (2002) Fuzzy Support Vector Machines. *IEEE Trans. Neural Networks* 13(2):464-471.

Lin CF, Wang SD (2004) Training Algorithms for Fuzzy Support Vector Machines with Noisy Data. *Pattern Recognition Lett.* 25(14):1647-1656.

Kullback S, Leibler RA (1951) On Information and Sufficiency. *The Annals of Mathematical Statistics* 22(1):79-86.

Netzer O, Srinivasan VS (2011) Adaptive Self-Explication of Multiattribute Preferences. *J. Marketing Res.* 48(1):140-156.

Orme B (2007) Three Ways to Treat Overall Price in Conjoint Analysis. *Sawtooth Software Research Paper Series.*

Orme B (2010) *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*, 2nd ed. (Research Publishers LLC, Madison, WI).

Park YH, Ding M, Rao VR (2008) Eliciting Preference for Complex Products: A Web-Based Upgrading Method. *J. Marketing Res.* 45(5):562-574.

Patra S, Bruzzone L (2012) A Batch-mode Active Learning Technique Based on Multiple Uncertainty for SVM Classifier. *Geoscience and Remote Sensing Let.* 9(3): 497-501.

Salton G, McGill MJ (1986) *Introduction to Modern Information Retrieval* (McGraw-Hill, New York).

Sattler H, Hensel-Börner S (2000) A Comparison of Conjoint Measurement with Self-Explicated Approaches. Gustafsson A, Huber F, eds. *Conjoint Measurement: Methods and Applications* (Springer, Berlin), 147-159.

Scholz SW, Meissner M, Decker R (2010) Measuring Consumer Preferences for Complex Products: A Compositional Approach Based on Paired Comparisons *J. Marketing Res.* 47(4):685-698.

Shilton A, Lai DT (2007) Iterative Fuzzy Support Vector Machine Classification. *2007 Fuzzy Systems Conf. Proc.* 1391-1396.

Srinivasan VS (1988) A Conjunctive-Compensatory Approach to the Self-Explication of Multiattributed Preferences. *Decision Sci.* 19(2):295-305.

Tong S, Koller D (2001) Support Vector Machine Active Learning with Applications to Text Classification. *J. Machine Learning Res.* 2:45-66.

Toubia O, Evgeniou T, Hauser JR (2007) Optimization-Based and Machine-Learning Methods for Conjoint Analysis: Estimation and Question Design. Gustafsson A, Herrmann A, eds. *Conjoint Measurement: Methods and Applications*, 4th ed. (Springer, New York), 231-258.

Toubia O, Hauser JR, Simester DI (2004) Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis. *J. Marketing Res.* 41(1):116-131.

Toubia O, Hauser JR, Rosanna Garcia R (2007) Probabilistic Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis: Theory and Application. *Marketing Sci.* 26(5):596-610.

Toubia O, Johnson E, Evgeniou T, Delquié P (2013) Dynamic Experiments for Estimating Preferences: An Adaptive Method of Eliciting Time and Risk Parameters. *Management Sci.* 59(3): 613-640.

Toubia O, Simester DI, Hauser JR, Dahan E (2003) Fast Polyhedral Adaptive Conjoint Estimation. *Marketing Sci.* 22(3):273-303.

Vapnik VN (1998) *Statistical Learning Theory* (John Wiley & Sons, New York).

Wang Y, Wang S, Lai KK (2005) A Fuzzy Support Vector Machine to Evaluate Credit Risk. *IEEE Trans. Fuzzy Sys.* 13(6):820-831.

**Figure 1: Overall Flow of the Proposed Adaptive Question Design Method**

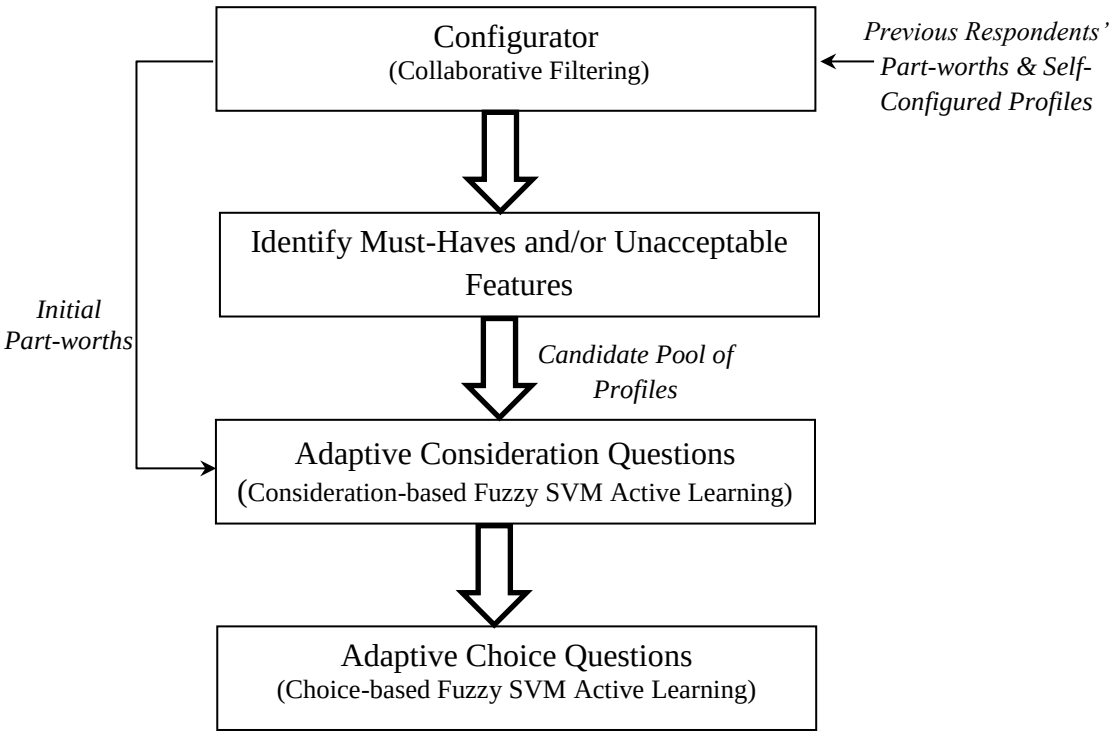

**Table 1: Comparison of Predictive Validity: Digital Camera Study**

| *Method (Sample Size)* | *Hit Rate* | |
|---|---|---|
| | *Ave.* | *SE* |
| *Cond 1: Proposed Method (N=73)* | .801* | .030 |
| *Cond 2: Fuzzy SVM without Collaborative Filtering (N=69)* | .779* | .032 |
| *Cond 3: Self-Explicated (N=66)* | .689 | .041 |
| *Cond 4: Upgrading (N=80)* | .563 | .038 |
| *Cond 5: ACBC (N=75)* | .660 | .038 |
| *Cond 6: CBC (N=62)* | .605 | .045 |

*best in column or not significantly different from best in column at the .05 level

**Table 2: Comparison of Predictive Validity: Computer Tablet Study**

| *Method* | *KL Divergence* | | | | | | | | *Hit Rate (Pooled)* | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *Initial Validation* | | | *Delayed Validation* | | | *Pooled* | | | |
| | *N* | *Ave.* | *SE* | *N* | *Ave.* | *SE* | *Ave.* | *SE* | *Ave.* | *SE* |
| *Cond 1: Proposed Method* | 35 | .288*[a] | .062 | 26 | .260*[a] | .073 | .471*[a] | .054 | .671* | .041 |
| *Cond 2: Fuzzy SVM without Consideration Questions* | 36 | .426 | .066 | 26 | .407 | .077 | .577 | .052 | .565 | .049 |
| *Cond 3: Self-Explicated* | 39 | .443 | .058 | 27 | .479 | .080 | .662 | .043 | .585 | .031 |
| *Cond 4: ACBC* | 41 | .472 | .059 | 30 | .531 | .062 | .652 | .048 | .520 | .033 |

*best in column or not significantly different from best in column at the .05 level

[a]marginally better than the fuzzy SVM without consideration questions condition ($p < .1$)

**Table 3: Predictive Validity by Number of Questions Asked: Computer Tablet Study**

| # of Questions (*k*) | KL-Divergence (smaller is better) | Hit Rate (larger is better) | |
|---|---|---|---|
| 1 | .557 | .538 | ← Self-Configurator |
| 3 | .600 | .586 | Consideration Questions |
| 5 | .602 | .619 | |
| 7 | .611 | .648 | |
| 9 | .522 | .614 | |
| 11 | .529 | .638 | Choice Questions |
| 13 | .504 | .648 | |
| 15 | .484 | .638 | |
| 17 | .494 | .671 | |
| 19 | .552 | .657 | |
| 21 | .498 | .667 | |
| 23 | .496 | .676 | |
| 25 | .500 | .667 | |
| 27 | .502 | .662 | |
| 29 | .502 | .676 | |
| 31 | .480 | .681 | |
| 33 | .480 | .652 | |
| 34 | .471 | .671 | |

**Table 4: Using Proposed vs. Conjunctive Question Selection When the True Consideration Decisions are Conjunctive: Synthetic Data Experiments**

Table 4a: Parameter Recovery and Predictive Validity Comparisons

| Product Dimension | Performance Measure | Question Selection Method | | Question Selection Method | |
|---|---|---|---|---|---|
| | | Proposed Method | Dzyabura and Hauser (2011) | Proposed Method | Dzyabura and Hauser (2011) |
| | | Conjunctive Estimation | | Fuzzy SVM Estimation | |
| 3x15 | Hit Rate | .986 | 1.000* | .966* | .919 |
| | KL Divergence | .072 | .000* | .091* | .264 |
| | $U^2$ | .928 | 1.000* | .728* | .515 |
| | | | | | |
| 3x25 | Hit Rate | .948 | .992* | .948* | .910 |
| | KL Divergence | .252 | .013* | .117* | .246 |
| | $U^2$ | .818 | .850 | .693* | .588 |
| | | | | | |
| 3x35 | Hit Rate | .945 | .994* | .936* | .892 |
| | KL Divergence | .232 | .018* | .131* | .267 |
| | $U^2$ | .685 | .743* | .614* | .577 |

*significantly better than the alternative method at the .05 level

Table 4b: Average Time to Generate Next Question Comparisons (In Seconds)

| Product Dimension | Question Selection Method | |
|---|---|---|
| | Proposed Method | Dzyabura and Hauser (2011) |
| 3x15 | .218* | 2.856 |
| 3x25 | .214* | 6.920 |
| 3x35 | .215* | 11.861 |

*significantly better than the alternative method at the .05 level

**Table 5: Performance Comparisons When True Consideration and Choice Decisions Are Both Based on A Part-Worths Model: Synthetic Data Experiments**

Table 5a: Parameter Recovery and Predictive Validity Comparisons

| Product Dimension | Performance Measure | Proposed Method | Dzyabura and Hauser (2011) | Abernethy et al. (2008) |
|---|---|---|---|---|
| 3x15 | Hit Rate | .948* | .929 | .937 |
| | KL Divergence | .279* | .365 | .327 |
| | $U^2$ | .736* | .701 | .701 |
| | MAE | .745* | .858 | .834 |
| | RMSE | .920* | 1.057 | 1.037 |
| 3x25 | Hit Rate | .933* | .925 | .870 |
| | KL Divergence | .343* | .373 | .553 |
| | $U^2$ | .679* | .623 | .596 |
| | MAE | .839* | .877 | 1.109 |
| | RMSE | 1.040* | 1.086 | 1.374 |
| 3x35 | Hit Rate | .901* | .893 | .812 |
| | KL Divergence | .460* | .480 | .694 |
| | $U^2$ | .625* | .583 | .487 |
| | MAE | 1.015* | 1.042 | 1.347 |
| | RMSE | 1.254* | 1.291 | 1.668 |

*significantly better than the alternative methods at the .05 level

Table 5b: Average Time to Generate Next Question Comparisons (In Seconds)

| Product Dimension | Question Type | Proposed Method | Dzyabura and Hauser (2011) | Abernethy et al. (2008) |
|---|---|---|---|---|
| 3x15 | Consideration | .269* | 5.779 | - |
| | Choice | .612 | - | .002* |
| 3x25 | Consideration | .206* | 3.075 | - |
| | Choice | .609 | - | .002* |
| 3x35 | Consideration | .217* | 4.702 | - |
| | Choice | .690 | - | .004* |

*significantly better than the alternative method at the .05 level

**Table 6: Performance Comparisons between Fuzzy SVM and Soft Margin SVM Active Learning: Synthetic Data Experiments**

Table 6a: Parameter Recovery and Predictive Validity Comparisons with Response Errors

| Error Positions | Performance Measure | Fuzzy SVM | Soft Margin SVM |
|---|---|---|---|
| Early | Hit Rate | .760* | .720 |
| | KL Divergence | .769* | .827 |
| | $U^2$ | .280* | .201 |
| | MAE | 1.694* | 1.834 |
| | RMSE | 2.103* | 2.275 |
| | | | |
| Middle | Hit Rate | .834* | .819 |
| | KL Divergence | .634* | .669 |
| | $U^2$ | .417* | .394 |
| | MAE | 1.519* | 1.596 |
| | RMSE | 1.889* | 1.978 |
| | | | |
| Late | Hit Rate | .893 | .890 |
| | KL Divergence | .483 | .492 |
| | $U^2$ | .511 | .507 |
| | MAE | 1.319 | 1.338 |
| | RMSE | 1.636 | 1.649 |

*significantly better than the alternative method at the .05 level

Table 6b: Parameter Recovery and Predictive Validity Comparisons with No Response Error

| Performance Measure | Fuzzy SVM | Soft Margin SVM |
|---|---|---|
| Hit Rate | .901 | .904[a] |
| KL Divergence | .460 | .448 |
| $U^2$ | .625 | .648* |
| MAE | 1.015 | .992* |
| RMSE | 1.254 | 1.226* |

*significantly better than the alternative method at the .05 level

[a]marginally better than the alternative method at the .1 level

**Table 7: Performance Improvements (in Percentage) Over Non-Informative Initial Part-worths**

| # of Questions | Heterogeneous Configurators | | | | | Homogenous Configurator | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Hit Rate | KL | $U^2$ | MAE | RMSE | Hit Rate | KL | $U^2$ | MAE | RMSE |
| 1 | 40.89% | 27.61% | 443.33% | 21.23% | 22.28% | 54.33% | 45.78% | 652.14% | 30.64% | 32.20% |
| 5 | 21.12% | 14.03% | 186.49% | 12.26% | 12.24% | 23.93% | 16.47% | 209.10% | 15.07% | 15.21% |
| 10 | 13.81% | 10.21% | 104.38% | 9.00% | 8.70% | 13.79% | 9.97% | 99.54% | 9.57% | 9.59% |
| 15 | 9.53% | 8.22% | 60.92% | 6.95% | 6.43% | 8.37% | 6.93% | 51.38% | 6.95% | 6.70% |
| 20 | 6.57% | 6.85% | 36.56% | 5.51% | 5.03% | 4.97% | 4.87% | 26.21% | 4.85% | 4.53% |
| 25 | 5.03% | 6.03% | 25.50% | 4.75% | 4.22% | 3.47% | 4.01% | 16.47% | 4.19% | 3.55% |
| 30 | 4.25% | 5.85% | 20.36% | 3.45% | 3.54% | 2.42% | 3.40% | 10.32% | 3.13% | 2.63% |
| 35 | 3.26% | 5.25% | 13.87% | 2.93% | 2.91% | 2.17% | 3.64% | 8.26% | 2.43% | 2.39% |
| 40 | 2.61% | 4.79% | 10.28% | 2.73% | 2.60% | 1.45% | 2.87% | 5.40% | 2.03% | 2.05% |
| 45 | .90% | 2.04% | 3.24% | 1.20% | 1.25% | -.42% | -1.05% | -1.58% | -.55% | -.51% |
| 50 | 1.18% | 3.19% | 4.46% | 1.53% | 1.64% | -.21% | -.65% | -.086% | -.85% | -.51% |
| 55 | .88% | 2.80% | 2.97% | 1.32% | 1.49% | -.16% | -.66% | -.57% | -.41% | -.39% |
| 60 | .65% | 2.32% | 2.22% | 1.21% | 1.28% | .04% | .27% | .04% | .23% | -.17% |
| 65 | .44% | 1.76% | 1.54% | 1.06% | 1.09% | -.10% | -.81% | -.36% | -.35% | -.42% |