A Belief-Preference Model of Choice for Experience and Credence Goods

Marco Costanigro*and Yuko Onozaka†

Abstract

We develop a methodology addressing the issue of confounded beliefs and preferences in models of discrete choice. First, we formalize the theoretical framework and logical underpinnings of a belief-preference model of choice for experience and credence goods, where subjective beliefs relate to uncertain product quality. Then, we present the experimental procedure within the context of an online choice experiment studying consumer food preferences. The empirical strategy leverages information from a quality sorting task to identify and estimate beliefs, while choice data are used to recover preferences. By conditioning product choices on predicted quality perceptions, the issue of endogenous beliefs is resolved.

This research was funded by Norwegian Seafood Research Fund (FHF) project number 7351031.

*Associate Professor, Department of Agricultural and Resource Economics, Colorado State University, U.S.A.

†Corresponding Author. Professor, UiS Business School, University of Stavanger, N-4036 Stavanger, Norway. Phone: +47 51833716. e-mail: *yuko.onozaka@uis.no*. Senior authorship is shared.

1 Introduction

Since the increased availability of non-aggregated data created the premise for the rise of the microeconometrics literature, empirical work in economics abandoned the idea of studying a "representative" or average consumer to focus on modeling heterogeneous preferences (Heckman, 2001). The literature on discrete choices has been especially prolific: after the introduction of the random utility framework and the conditional logit model (McFadden, 1974), an array of novel methods relaxed the assumption of homogeneous preferences to account for "unobserved" or "unexplained" (by the traditional observational covariates) heterogeneity. Examples include the mixed (McFadden and Train, 2000) and latent class (Arcidiacono and Jones, 2003) logit models, and more recently the generalized multinomial logit (Fiebig et al., 2010). While these methods have been shown to capture heterogeneity and fit the data quite well (e.g., Keane and Washi, 2013), a lingering issue is the lack of a formal explanation for the observed behavioral differences. The typical argument may mention varying consumer taste, but in practice heterogeneity—while recorded—is often left unexplained.

In this article we show how distingushing subjective beliefs from consumer preferences in discrete choice models can help resarchers in understanding heterogeneity. The premise is that economic agents are routinely forced to make decisions under partial information, and often have to rely on expectations and perceptions to fill the informational gap (Bertrand and Mullainathan, 2001; Manski, 2004). The fact that perceptions are generally not observed by the researcher is the rationale behind the fundamental issue of identification pointed out by Manski (2004): if choices can be explained by multiple combinations of preferences for alternative outcomes and beliefs about the likelihood of each outcome, the inference we can draw from analyzing choice data alone is, at best, tenuous. For example, people may not want to purchcase hybrid cars because they do not *believe* that such vehicles are effective in reducing CO_2 emissions, or it may be that they have weak *preferences* for environmental outcomes. Surely, there are many other complex trade-offs in the hybrid vs. conventional decision where both beliefs and preferences will play a determinant role. Whatever the specific case, a researcher observing car choice data alone will not be able to distinguish between competing explanations, unless one can be sure that decision-makers are rational, have access to similar sources of information, and interpret it in a consistent manner (Manski, 2004; Delavande, 2008; Zafar, 2011).

The novel methodology for discrete choice experiments (DCE) preseneted here leverages subjective beliefs to explain heterogeneous choice behavior. Our work draws inspiration from recent literature on the role of subjective expectations in financial and other consumer decisions involving uncertain outcomes (see Manski (2004) for a review), but is specifically designed to study markets where product quality is the principal source of uncertainty (i.e., experience (Nelson, 1970; 1974) and credence (Darby and Karni, 1973) goods). Following a long-established representation of the quality perception process (Brunswik, 1955; Dudycha and Naylor, 1966; Steenkamp, 1990), we conceptualize subjective beliefs as the filter through which observable product characteristics and other quality cues are mapped into quality perceptions; and define consumer preferences as the set of implicit weights describing systematic, subjective trade-offs between multiple quality dimensions and price. The application we study involves food choices (chicken vs. salmon), where we first exmine the role of extrinsic cues (expiration date and other point-of sale information) in shaping expectations about product qualities (taste, freshness, healthiness, safety and convenience), and then determine how trade-offs among qualities and price affect food choice.

The experimental procedure we propose entails two tasks: in the first *quality sorting* task, products are described by a set of characteristics typically observable at the time of purchase, and participants are asked to indicate the "best performing" product in a number of salient quality dimensions. The second *product choice* task is the typical DCE procedure eliciting the preferred option or product in a choice set. The empirical strategy is similarly comprised of two-steps: choices from the first task are used to estimate beliefs and to predict quality perceptions. Then, we estimate the preference parameters by conditioning product choices on predicted qulity perceptions (i.e. estimated beliefs and product characteristics). The resulting *belief-preference* model decomposes unexplained heterogeneity in Willingness-to-Pay (WTP) for product characteristics in two separate sources of variation: the subjective interpretation of the observable quality cues (*beliefs*), and an individual's system of trade-offs among quality dimensions and price (*preferences*).

While we are not aware of any previous applications of the methodological approach presented here, our work contributes to a growing body of economics work on subjective perceptions. The *Journal of Applied Econometrics* recently dedicated a full issue (Bellemare and Manski, 2011) to studying how appropriately measured subjective expectations can augment choice data and provide a solution to the identification problem. The consensus arising from this line of work is that beliefs can be measured via subjective probabilities elicited in *ad hoc* questionnaires, and then included in a model based on the Von Neumann-Morgenstern expected utility maximization framework (e.g., Delavande, 2008; Zafar, 2011). Indeed, this approach offers some important desirable features: in addition to laying its foundation on established economic theory, researchers have developed "proper" (i.e., incentive-compatible) scoring rules for the elicitation of subjective expectations in the lab (see Savage, 1971; Nyarko and Schotter, 2002; Karni, 2009). One object of debate is whether people are willing and able to express subjective beliefs in a probabilistic form,[1] an assumption disputed by some cognitive psychologist (e.g., Zimmer, 1983), but researchers have found that describing probabilities as frequencies ("x times out of 100 cases") (Gigerenzer, 1991), or visual aids (Delavande et al., 2011) can simplify the elicitation task.

While the direct application of the probabilistic approach to discrete choice

[1]For example, a heaping of probabilities at 50% often represents "epistemic uncertainty" (i.e. it's a fifty-fifty chance) rather than the perceived likelihood of an event (de Bruin et al., 2000)

experiment is possible (Lusk et al. (2014), for example, estimated a model of meat choice incorporating the subjective probabilities that a steak will be tender), its application to experience and/or credence goods suffers from some important limitations. For one, incentive-compatible belief elicitation involves fairly elaborate and complex procedures, so the applicability to several empirical settings (including online surveys) appears questionable.[2] At a more fundamental level, specifying a density function—even a subjective one—requires defining one or more points of support. Subjective probabilities are meaningful when attached to mutually exclusive events (e.g., the likelihood of a bull or bear market (Hurd et al., 2011), getting pregnant or not (Delavande, 2008), finding a job after graduation or not (Zafar, 2011), but not so much for product choice involving quality perceptions. In the context of credence and experience goods, subjective beliefs refer to how observable product characteristics (e.g., the information on a label, as in organic food) are used to infer qualities, and quality perceptions generally imply context-dependent product comparisons (Steenkamp, 1990). Familiar expressions such as "this car is more environmentally friendly than another" or "organic products are healthier than conventionally grown foods" imply relative judgements of superiority/inferiority that cannot be easily translated into probabilistic assessments.

So, if not probabilities, what method should be used to measure beliefs and quality perceptions? In the marketing literature, researchers often establish the most salient quality dimensions (e.g., Brucks et al., 2000 find that versatility, durability and performance are important when purchasing durable goods), and then measure perceptions by means of Likert-type ratings. These scales are succinct, easily understood by the consumer and, by portraying perceived quality on a linear spectrum, facilitate comparisons between differentiated products. The linear representation is

[2]Many studies on subjective beliefs have fairly small sample size (N<200), often because of the complexity and the length of the instrument. For example, Zafar (2011) surveyed students at Northwestern University, and reported that students needed about 45 minutes to one hour to complete the questionnaire. Implementing such experimental procedures in marketing or other studies of purchasing behavior, where a large number of participants is necessary to ensure sample representativeness, is impractical and often impossible.

also consistent with the extant economics theory on quality expectations (Akerlof, 1970; Mussa and Rosen, 1978; etc.), but economists have generally been resistant to adopt such scales in empirical work, confining research to instances where quality is "of an objective kind" [3] (Lancaster, 1966), or resort to "objective" expert ratings (e.g., wine tasting scores—see Landon and Smith, 1998; Costanigro et al., 2009; Costanigro et al., 2010) to measure unobserved product qualities. The pervasive reluctance to incorporate subjective data in economic analysis has been lamented by other authors (e.g., Bound, 1989; Benítez-Silva et al., 2004; Manski, 2004), but some concerns are well-grounded. Likert scales are inherently qualitative, hard to generalize, and notoriously prone to framing effects. Endogeneity is another substantive issue: elicited beliefs and other subjective data (including probabilities) may be measured with error or correlated with other unobserved choice factors, causing biased or inconsistent estimation (Bertrand and Mullainathan, 2001; Benítez-Silva et al., 2004; Teisl and Roe, 2010). In sum, even though attitudinal questions have been shown to increase explanatory power (Bertrand and Mullainathan, 2001) or provide a rationale for people's choices (e.g. Costanigro et al., 2014; Malone and Lusk, 2017; Malone and Lusk, 2018), this type of analysis is often met with skepticism.

An alternative approach is the use of standard DCE technique to elicit quality perceptions, rather than product choices, in order to estimate beliefs (as in Costanigro et al. (2015)). In this paper we show how eliciting both quality perceptions and product choices allows to separately identifying beliefs and preferences. The model depicts how a certain cue affects multiple dimensions of the quality perception constructs (e.g., "hybrid cars are environmentally friendly but tend to be slower"), allowing to decompose consumers' WTP for product characteristics through the different quality perception channels and, ultimately, providing an explanation as to *why* some consumers are willing to pay more for certain features. The experimental procedure is relatively simple to implement, and the two-step elicita-

[3] That is, engineering quality, such as the speed of a computer processor.

tion simulates the familiar process of making quality comparisons before deciding on a purchase, without requiring participants to quantify subjective probabilities, representing quality on a Likert scale, or articulating beliefs verbally. Another appealing feature is that the use of predicted quality perceptions as regressors in the second (preference) estimation step allows bypassing endogeneity concerns, as subjective data are replaced by estimated parameters (beliefs) and product characteristics, which are experimentally controlled and therefore exogenous. While the current application pertains to food choices, the method is fully general, and could be readily adapted to the numerous disciplines where DCE have been used to study consumer choice (e.g., marketing, environmental and natural resource economics, transportation and health economics). Conceptually, one could also adapt the belief-preference framework to a situation where non-hypothetical consumer choices are observed, and the quality sorting task of the DCE procedure is used to augment the data.

2 Conceptual Framework

We take the Lancasterian (1966) view that quality is a product's effectiveness in its intended use(s), and consumers derive utility from the intrinsic qualities of a good, rather than the good itself. This quality construct generating utility is also inherently multidimensional (Parasuraman et al., 1985; Zeithmal, 1988). For instance, reliability and gas mileage are important qualities for most car buyers, and similarly most shoppers will enjoy nutritious, tasty and convenient food. Assuming that product j possesses the Q-dimensional vector of quality levels $\mathbf{Q}_j' = (Q_j^1, Q_j^2, \cdots, Q_j^Q)$, an individual i who consumes a unit of it will realize a level of utility $U_{ij} = U_i(\mathbf{Q}_j, Price_j; \boldsymbol{\gamma}_i)$, where $\boldsymbol{\gamma_i}$ is a vector of consumer-specific preference weights and $Price_j$ is the price paid.

In many relevant cases, the true nature (or state) of the qualities is unobserved at the time of the purchase (experience qualities (Nelson, 1970), such as taste), or never verifiable (credence qualities Darby and Karni, 1973, such as environmental friendliness). In such instances, consumers will rely on perceived or expected qualities, which we indicate as $\mathbf{Q}_{ij}$ to emphasize the subjective nature of perceptions, yielding: $U_{ij} = U_{ij}(Q_{ij}^1, Q_{ij}^2, \cdots Q_{ij}^Q, \ Price_j; \boldsymbol{\gamma}_i)$. How are the perceived qualities formulated? Drawing from Steenkamp (1990)'s conceptualization of the lens model (Brunswik, 1955; Dudycha and Naylor, 1966), we postulate that consumers utilize observable quality cues[4] to evaluate the true quality state of a given product. These may include intrinsic cues (e.g., smell and color) or extrinsic ones (e.g., labels or brand names). Suppose the market environment provides a total of K quality cues for product j that individual i is exposed to, which we include in the vector $\mathbf{X}_{\mathrm{ij}}{}' = (\mathbf{X}_{i1}, \mathbf{X}_{i2}, \cdots \mathbf{X}_{iK})$. Then, for each quality dimension q the perceived quality $Q_{ij}^q = Q(\mathbf{X}_{ij}; \boldsymbol{\beta}_{\mathrm{i}}^q)$ is a function of market cues and $\boldsymbol{\beta}_{\mathrm{i}}^{q\prime} = (\beta_{i1}^q, \beta_{i2}^q, \cdots \beta_{iK}^q)$, a vector of subjective belief parameters mapping market cues into quality perceptions. It follows that heterogeneity in consumers' perceptions of quality can arise from individual exposure to varying cues, but also from different interpretations of what a given quality cue signifies.[5] For example, if a consumer is convinced that organic products are healthier, then the organic cue will be associated with a positive belief parameter for healthiness, and *ceteris paribus* organic products will be preferred to conventional ones. If a consumer is skeptical about the increased healthiness of organic food, then $\beta_{i,k=Organic}^{q=Health} = 0$, and the organic cue does not influence the perceived healthiness of the product. It should be noted that the belief-formation process is not modeled here: beliefs are taken as static or given at a point in time

[4] We refrain from using the term "product attribute," as it has been used in the vertical product differentiation literature quite ambiguously to refer to both cues and qualities. Likewise, we use the term "quality" rather than "characteristic," as Lancaster(1966)'s original conception portrays characteristics as objective and invariant among individuals. Instead, we follow the path of Maynes (1976), Zeithmal (1988), and Steenkamp (1990), arguing that perceived qualities, being the result of a subject-object interaction, are neither completely subjective, nor wholly objective.

[5] As in most applications, we abstract from limited-information situations where consumers may utilize prices as a quality cue (Gneezy et al., 2014), which may be especially important with prestige-seeking behavior (e.g., purchasing expensive wine or jewelry). We did not expect this phenomenon to be relevant for our application, packaged food choice, where low correlation between perceived quality and prices are reported (Riesz, 1979) and generally the use of price-quality heuristics diminishes as more information becomes available (Chang and Wildt, 1996).

(e.g., at the time of a survey). To the extent that the survey instrument itself does not alter beliefs, this assumption is reasonably benign.

We now turn to developing the theoretical framework that will guide our empirical work. Consider a simple model where consumer heterogeneity is relegated to the stochastic term η_{ij}. Assuming that consumers face identical prices for the same products, the typical specification of many choice models within this setting is

$$U_{ij} = U(\mathbf{X}_{ij}, Price_j; \boldsymbol{\delta}) = \mathbf{X}'_{ij}\boldsymbol{\delta}_1 + \delta_{price}Price_j + \eta_{ij} \tag{1}$$

which, in the case of a choice between product A and B ($j = A, B$), yields the familiar probabilistic statement

$$Pr(U_{iA} > U_{iB}) = Pr((\mathbf{X}_{iA} - \mathbf{X}_{iB})'\boldsymbol{\delta}_1 + \delta_{price}(Price_A - Price_B) > \eta_{\mathrm{iB}} - \eta_{\mathrm{iA}}); \tag{2}$$

that is, consumers compare products based on market cues and prices. Simple inspection reveals that the model in equation 1 is a *reduced form model* specification, as belief ($\boldsymbol{\beta}$) and preference ($\boldsymbol{\gamma}$) parameters are confounded in $\boldsymbol{\delta}$. Estimating $\boldsymbol{\delta}$ allows calculating willingness to pay for each quality cue, but it is not clear how the primitive parameters could be recovered. Random parameter specifications (see McFadden and Train, 2000) of the model in equation 1 could be used to introduce heterogeneity in WTP for observable characteristics, but they still won't address the fundamental issue of why some people are willing to pay more than others.

Again, our approach to recovering belief and preference parameters from choice data consists of a relatively simple two-step process, involving a *quality sorting* and a *product choice* task. In the quality sorting task participants are asked to examine a set of products and associated cues, and select the best one in a number of relevant quality dimensions (e.g., convenience, taste, healthiness, ...). Then, in the product choice task, product prices are added to the information set and participants indicate which product they would buy (if any) at the posted prices. The idea is to use the first set of choices to estimate beliefs, and use the second set to recover preferences. To illustrate, examine the case where consumers are asked which of the two products A and B is superior in the quality dimension q, and the following model of quality perceptions based on market cues and beliefs:

$$Q_{ij}^{q} = Q(\mathbf{X}_{ij}; \boldsymbol{\beta}^{q}) + \epsilon_{ij}^{q} = \mathbf{X}_{ij}' \boldsymbol{\beta}^{q} + \epsilon_{ij}^{q} \tag{3}$$

where $\boldsymbol{\beta}^{q}$ is an average or constant vector of beliefs in the consumer population and ε_{ij} is a random error term. The *beliefs model* in equation 3 implies that $Pr(Q_{iA}^{q} > Q_{iB}^{q}) = Pr((\mathbf{X}_{iA} - \mathbf{X}_{iB})' \boldsymbol{\beta}^{q} > \varepsilon_{iB}^{q} - \varepsilon_{iA}^{q})$, an expression detailing the probability that product A is perceived to perform better than product B in a given quality dimension as a function of the available market cues and and a vector of (average) beliefs. Once quality perceptions are established, a model of how consumer preferences determine product choices takes the form

$$U_{ij} = U(\mathbf{Q}_{ij},\, Price; \boldsymbol{\gamma}) + \nu_{ij} = \mathbf{Q}_{ij}^{'} \boldsymbol{\gamma}_1 + \gamma_{price} Price_j + \nu_{ij} \tag{4}$$

where $\boldsymbol{\gamma}_1$ is a $Q \times 1$ vector capturing the marginal utility of each quality, γ_{price} is the marginal utility of money, and ν_{ij} is a random error term. Thus, conditional on making a purchase, the probability of purchasing A vs. B is

$$Pr(U_{iA} > U_{iB}) = Pr((\mathbf{Q}_{iA} - \mathbf{Q}_{iB})' \boldsymbol{\gamma}_1 + \gamma_{price}(Price_A - Price_B) > \nu_{iB} - \nu_{iA}). \tag{5}$$

The *preference model* in equation 4 and 5 is intuitively quite appealing, as it portrays the familiar task of comparing two products on the grounds of price vs. quality tradeoffs. Figure 1 depicts a graphical representation of the reduced-form model in equation 1 (the top half), as well as the belief-preference model implied by equations

3 and 4 (the bottom half).

It is useful at this point to show how marginal effects and WTP can be calculated in the reduced form vs. belief-preference model. In the reduced-form model, the marginal utility garnered by a given cue is simply $\frac{\partial \mathbf{U_{ij}}}{\partial \mathbf{x_k}} = \delta_k$, and changes in WTP attributable to an increase in the k^{th}quality cue are calculated as $WTP_k = \delta_k / \left(-\delta_{price}\right)$. The beliefs-preference model permits a part-worth decomposition of this total change in WTP, via the channels of the Q-dimensional vector of perceived qualities. That is, $\frac{\partial U_{ij}}{\partial x_k} = \sum_q \frac{\partial U_{ij}}{\partial Q_j^q} \frac{\partial Q_j^q}{\partial x_k} = \sum_q \gamma_q \beta_k^q$ and the part-worth decomposition of the WTP for quality cue x_k is

$$WTP_k = \gamma_1 \beta_k^1 / \left(-\gamma_{price}\right) + \gamma_2 \beta_k^2 / \left(-\gamma_{price}\right) + ... + \gamma_Q \beta_k^Q / \left(-\gamma_{price}\right). \tag{6}$$

One immediate challenge posed by the belief-preference model is that the perceived qualities $\mathbf{Q}_{ij}$ are unobserved by the analyst, and will need to be either directly elicited in an experimental setting, or replaced by their estimates obtained from the belief model in 3, i.e. $\hat{\mathbf{Q}}_j' = (\mathbf{X}_{ij}'\hat{\boldsymbol{\beta}}^1, \mathbf{X}_{ij}'\hat{\boldsymbol{\beta}}^2, \cdots, \mathbf{X}_{ij}'\hat{\boldsymbol{\beta}}^Q)$, which is the approach we propose and present here. Thus, the empirical choice model we are set to estimate take the form:

$$Pr(U_{iA} > U_{iB}) = Pr((\hat{\mathbf{Q}}_{iA} - \hat{\mathbf{Q}}_{iB})'\boldsymbol{\gamma}_1 + \gamma_{price}(Price_A - Price_B) > \nu_{iB} - \nu_{iA}). \tag{7}$$

The advantage of leveraging the belief model to predict quality perceptions is that $\hat{\mathbf{Q}}_{ij}$ is a function of (estimated) parameters $\boldsymbol{\beta}^q$and the vector of cues $\mathbf{X}_{ij}$, which in a choice experiment is controlled by the researcher and therefore exogenous. The quality perception model in equation 3, however, will produce the same predicted quality for all consumers, unless the quality cues vary across participants, suggesting that modeling heterogeneity in beliefs is germane to the identification of preferences. This is not a real drawback: differences in beliefs across consumers are a fact of life, and learning about them may reveal important marketing or policy implications. The armamentarium of today's applied research include many instruments well-suited from the task, including random coefficients, finite mixtures, and other more recent model developments (see Keane and Washi, 2013). Defining $f\left(\boldsymbol{\beta^q}\right)$ to represent the distribution of beliefs for quality trait q in the consumer population, and $f\left(\boldsymbol{\gamma}\right)$ to be the distribution of preferences, we can cast the belief and preference models in their random coefficient counterparts:

$$Pr(Q^q_{iA} > Q^q_{iB}) = \int Pr(Q^q_{iA} > Q^q_{iB}) f\left(\boldsymbol{\beta^q}\right) d\boldsymbol{\beta^q}; \tag{8}$$

and

$$Pr(U_{iA} > U_{iB}) = \int Pr(U_{iA} > U_{iB}) f\left(\boldsymbol{\gamma}\right) d\boldsymbol{\gamma}. \tag{9}$$

3 Survey, Experimental Design and Data

The applied focus of the project, funded by Norwegian Seafood Research Fund (FHF), was to understand consumers' perceptions of salmon compared to chicken in the U.S. market.[6] The survey was administered in May, 2015 to a panel maintained by the Survey Sampling Inc., and is representative of the US population in terms of age, gender and geographic distribution. Given the marketing scope of our research, only participants who reported consuming both products were retained in the sample for the choice experiment. Before starting the choice experiment, participants completed a short psychometric questionnaire designed to measure the importance attributed to food-related decisions (i.e., the Food Involvement scale developed by Bell and Marshall, 2003, which we will further discuss in Section 4). Our sample consists of 1,202 complete responses,[7] and summary statistics are presented in Table

[6]Salmon producers consider chicken to be their main market competitor.

[7]The survey collected responses from 2,068 respondents, among which 1,419 (69%) reported consuming both products. Of those, 217 had missing information and were removed from the

1.

Among many possible cuts of chicken and salmon, we chose to compare boneless skinless chicken breasts and salmon fillets. The rationale is that these two cuts are commonly present in the US grocery stores, and they are reasonably similar in terms of occasions of consumption and cooking efforts. In addition to product type (chicken vs. salmon), three quality cues ($K = 3$) were selected as the most salient among those typically observed in a grocery store settings: display (shelf vs. counter), eat before date (3, 7, and 14 days), and Modified Atmosphere Packaging (MAP). The display cue relates to the choice between purchasing prepackaged products (which can be quickly grabbed from the shelves) or interacting with the fishmonger/butcher to select products and portion sizes. Expiration dates have been shown to have significant impact on consumers' purchase decisions (Shah and Hall-Phillips, 2017). MAP is a technology developed to delay food spoilage by filling the product packaging with a special mixture of gases instead of normal air. In the US, MAP packaging has been approved as "generally recognized as safe" by the U.S. Food and Drug Administration in 2002, and is therefore exempt form explicit labeling mandates (Grebitus et al., 2013). However, consumer research (Grebitus et al., 2013)has shown that MAP labeling may decrease consumer WTP for meat products.

Our interest lays in examining how different beliefs and preferences modulate the choice of chicken vs. salmon. Based on the major quality concerns generally reported by consumers while selecting food items (Grunert, 2005), we examine a total of five quality dimensions ($Q = 5$): freshness, taste, food safety, convenience, and healthiness. Each choice set included one chicken and one salmon product described by means of varying quality cues, and associated imagery. In the first step (*quality sorting task*), participants stated which of the products was superior (i.e. chicken, salmon, or "they are the same") in each of the five quality dimensions. In the second

sample.

phase (*product choice task*), price information was added to the quality cues, and respondents were asked to select the product they would purchase (chicken, salmon, or neither). The full experimental plan included 18 choice sets and was based on a labeled[8] fractional factorial design with three attributes: display (shelf vs. counter), eat before date (3,7,14 days) and prices (three levels).[9]

To limit the cognitive burden, we assigned a subset of six choice tasks to each participant, and devised a partitioning strategy accounting for the sequential release of information (quality cues first, and then prices). Namely, six blocks of three choice sets each were identified to maximize within-block similarity in the quality cue levels (price excluded), and participants received a randomly drawn question from each block, thereby generating 27 ($3 \times 3 \times 3$) unique surveys comprised of six choice tasks. In this way choice sets differing only in price levels would never be assigned to the same participant, thereby eliminating the possibility of repeating the same choice set during the quality sorting task.

Several additional constraints were imposed while translating the design into choice tasks in order to ensure that each choice set presented meaningful comparisons and plausible shopping scenarios. Table 2 summarizes all product descriptors (product type, quality cues and prices), corresponding levels, and constraints. Eat before dates were revealed only for the products purchased from the shelves, whereas counter products always appeared with a simple note stating "fresh from the counter". This reflects the current retail environment where prepackaged perishable products sold on the shelf have to display an expiration date, while such mandate is not imposed for products sold at the counter, with the expectation that they will be consumed in a relatively short amount of time.

The effect of MAP packaging on quality perceptions and product choices was identified by randomly assigning participants to a treatment or control group.

[8]A labeled choice experiment (see Bekker-Grob et al., 2010) was warranted as the product descriptors (quality cues) do not fully capture the fundamental differences between chicken and salmon.

[9]For this task, we used the software NGENE.

In the control group, no labeling cues or information on MAP were presented. In the treatment group participants received a basic science-based explanation of MAP packaging,[10] and all products sold on the shelf with a 14-day expiration date also included a MAP label. This choice reflects the reality that MAP requires specialized machinery, and only prepackaged products sold on the shelves usually employ MAP. Finally, price levels were made to be product-specific (to embed in the experiment existing market differences) with three levels: a medium base price ($5/lb for chicken and $10/lb for salmon based on market data from Norwegian Seafood Council and Kantar Worldpanel), low level (-25 percent) and high level (+25 percent). A representative screenshot is presented in Figure 2.

4 Empirical Analysis

Table 3 summarizes the basic results from the quality sorting and product choice tasks. Participants were able to discriminate the quality of the two products offered about 50% of the times, while in the remaining cases the products offered were deemed of equal quality. One exception is "freshness," for which the percentage of ties is lower (36%). This suggests that the quality cues we presented were more relevant to the determination of freshness than the other quality dimensions. Even though quality perceptions were similar in a number of cases, participants were able to signal purchasing choices once price information was made available: out of 8,081 choice occasions, chicken was selected 58% of the time, salmon in 28% of the occasions, and "no buy" only 15% of the times.

[10]The provided information was: When the "Eat Before Data" is very long, such as 14 days, it is because the product is packed with a special technology. One such technology is called Modified Atmosphere Packaging (MAP). In MAP, package is sealed with special mixture of gases instead of normal air. This packaging substantially slows down the processes of food spoilage so that products can stay fresh longer. A product labeled with MAP is also labeled with a statement "Packed with a protective atmosphere" below the eat before date.

4.1 Reduced Form Model

The first model we estimate is the reduced form specification in equation 1, where beliefs and preferences are confounded. The utility derived by individual i selecting product j (the choice occasion subscript is suppressed for simplicity) is defined as:

$$\begin{aligned} U_{ij} &= U(\mathbf{X}_{ij}, P_j; \delta) \\ &= \delta_0 Chicken_{ij} + \delta_1 Shelf_{ij} + \delta_2 Dates_{ij} + \delta_3 MAP_{ij} + \delta_4 Price_j + \eta_{ij} \quad (10) \end{aligned}$$

if chicken or salmon is purchased, and $U_{ij} = \delta_5 + \eta_{ij}$ if the "no purchase" option is chosen. In equation 10, $Chicken$ is an indicator variable for the product types (Chicken vs. Salmon) with associated alternative-specific constant δ_0, $Shelf$ is the indicator for shelf vs. counter display, $Dates$ are the number of days before expiration, MAP is the indicator for Modified Atmosphere Packaging (14 days eat-before-date × MAP information treatment), and $Price$ is the unit price in dollars. If there are J choices and under the assumption that the error term is Type I Extreme Value (McFadden, 1974) , the probability of product j being purchased by participant i takes the MN Logit[11] form: $Pr_{ij} = \frac{exp(\mathbf{X}'_{ij}\boldsymbol{\delta})}{\sum_j exp(\mathbf{X}'_{ji}\boldsymbol{\delta})}$.

The estimation results are shown in Table 4. As one would expect, the coefficient for Price is negative and significant. The average consumer is willing to pay \$1.06 more for products sold at the counter, and longer expiration dates are valued at 8 cents per day. However, MAP preservation lowers WTP by an average of -0.245/0.208=\$1.18. This is in contrast with Grebitus et al. (2013), who find positive valuation of MAP in American consumers, but their estimates did not separately control for increase shelf life, which has a positive effect here. While these empirical findings do have some obvious marketing implications, the model is silent about why counter display is better than shelf, or the reasons behind consumers' rejection of MAP. That is the topic of the following sections.

[11] We present the multinomial specification for the sake of generizability, but in our case we only have two alternatives.

4.2 Belief-Preference Model: Aggregated Specification

Next, we estimate the belief-preference model in equations 3 and 4. For all dimensions, perceived quality was made to be a function of product cues,[12] according to:

$$\begin{aligned} Q_{ij}^{q} &= \boldsymbol{X}'_{ij}\boldsymbol{\beta}^{q} + \varepsilon_{ij}^{q} = \beta_0^q Chicken_{ij} + \beta_1^q Shelf_{ij} \\ &\quad + \beta_2^q Date_{ij} + \beta_3^q MAP_{ij} + \varepsilon_{ij}^{q}; \text{ for } q = 1, \ldots, 5. \end{aligned} \tag{11}$$

For each participant, the data relative to the five quality dimensions (freshness, taste, safety, convenience and healthiness) were stacked and the five parameter vectors were estimated jointly via MNLogit in a fully interacted model (as if conducting a Chow test), so that one vector of estimates is obtained for each quality dimension.

The estimated coefficients (see Table 5) show that products sold at the counter are perceived to be fresher, tastier, safer and healthier, but prepackaged food on the shelf is more convenient. Longer expiration dates improve the perception of product freshness and safety, but not when a longer shelf life is obtained via the MAP technology. Holding expiration dates constant, all quality dimensions (except for convenience, which has a non-significant coefficient) are diminished with MAP-labeling, especially freshness and healthiness. There are also some product-specific effects—on average, chicken is perceived to be fresher, safer and more convenient than salmon, while the latter is more likely to be seen as healthier.

Parameter estimates are then used to predict perceived quality, $\hat{Q}_{ij}^{q}$, to be used as regressors in the ensuing product choice model.[13] Even though the model in equation 3 does not allow for belief heterogeneity, the experimental design ran-

[12] The reader will note that the alternative-specific constant for product type (Chicken vs. Salmon) is also included in equation 11. Even though product type is not exactly a quality cue, there might be differences in quality perceptions between the two products that are not fully explained by the quality cues we included in the experimental design. The alternative-specific constant is also included when we estimate the preference model for similar reasons.

[13] The "No Buy" option received a predicted quality of zero.

domized participants across 27 unique surveys of six choice tasks, thereby ensuring some exogenous variation in $\hat{Q}^q_{ij}$ within and across participants. However, while the experimental design ensures low correlations between quality cues, perceived qualities exhibit high degrees of correlation (see Table 6). For example, cues increasing perceived taste also increase freshness and safety, while convenience is negatively correlated to healthiness. While these findings are in a way informative, severe collinearity posed a challenge to estimating the preference parameters.

After attempting several specifications, including a principal component decomposition of the predicted qualities, we opt for a model favoring simplicity in interpretation and robust identification of the model parameters.[14] We construct a composite quality index, $\bar{\hat{\mathbf{Q}}}_{ij}$, by averaging the four positively correlated quality dimensions (freshness, taste, safety and healthiness), and left convenience as a separate independent variable, yielding the following model parameterization:

$$U_{ij} = \gamma_0 Chicken + \gamma_1 \bar{\hat{Q}}_{ij} + \gamma_2 \widehat{Convenience}_{ij} + \gamma_3 Price_j + \nu_{ij}. \tag{12}$$

Admittedly, this simplified parameterization loses some detail in determining the fine trade-offs between all quality dimensions. However, equation 12 is still quite useful in empirically investigating the Quality vs. Convenience vs. Price heuristics that are known to drive consumers' perceptions of value (Zeithmal, 1988).

Results of the MNLogit estimation are shown in Table 7. Both composite quality and convenience are positive and significant, while the price coefficient is negative and significant. These results conform to the idea that consumers attribute importance to both quality and convenience, but, as suggested by the negative correlations in Table 6, people face exogenous constrains in finding both within the same product. It is hard to interpret these estimates quantitatively, since the

[14]The principal component decomposition identifies similar contrasts between convenience vs. other quality dimensions, but using principal component scores would complicate model interpretation.

independent variables are the abstract, latent quality constructs in equation 11. However, the part-worth WTP decomposition presented in equation 8 can be used to understand relative magnitudes. For example, *ceteris paribus*, the quality cue "shelf" (vs. counter) awards a convenience premium of $\frac{(0.284\times0.614)}{0.207} = \$0.84$, but also a penalty of $\frac{\{[1/4(-1.282-0.708-0.662-0.743)]\times0.449\}}{0.207} = -\$1.84$ owed to a reduction in perceived freshness, taste, safety and healthiness.[15] Overall, the sum of the two effects is negative, (-$1.00, which is comparable to the estimate obtained from the reduced form model) because, on average, the gains in convenience fail to compensate the losses in the other quality dimensions. For expiration dates, the part-worth decomposition confirms that the consumer positive valuation of a longer shelf-life arises more from considerations about freshness and overall product quality than from convenience.

One reasonable objection is that, in terms of in-sample fit, the log-likelihood values reported in Table 7 are similar to what we obtained for the reduced form model in Table 4. This might seem disappointing at first: after all, we would like the model to improve our ability to predict choices. However, our interest here lies in recovering the primitive belief and preference parameters, which we argue have behavioral implications generalizable outside of the sample, just like recovering demand and supply elasticities from reduced form parameters is the objective of structural equation modeling.

4.3 Belief-Preference Model: A Latent Class Approach

A model of the quality perception process with heterogeneous beliefs (equation 8) was estimated via latent class modeling (LCM, also known as finite mixture, see Kamakura and Russell, 1989; Chintagunta, 1996), where it is assumed that a finite number of consumer "types" (or classes) exists. As LCM methods are now well-established, we maintain the compact presentation and focus on illustrating how

[15] The 1/4 scaling arises from the averaging in the composite quality index

standard LCM models can be used to fit the two-step belief-preference model, generating a wealth of information. To limit clutter in the notation, we also continue presenting the case of a single choice. Extension to the case of multiple choices and the resulting likelihood function is available from several sources (Greene and Hensher, 2003;Train, 2008; Pacifico and Yoo, 2013). Assuming that there are C classes of belief parameters, and defining $Pr(Q_{ij}^q)$ as the logistic probability that consumer i perceives product j to be superior among all the J available products in quality dimension q, the LCM specification is:

$$Pr(Q_{ij}^q) = \sum_{c=1}^{C} \pi_{ci} \left[\frac{exp\left(\boldsymbol{X}'_{ij}\boldsymbol{\beta}_c^q\right)}{\sum_{j=1}^{J} exp\left(\mathbf{X}'_{ij}\boldsymbol{\beta}_c^q\right)} \right] \; for \; q = 1, \ldots, 5; \tag{13}$$

where π_{ci} represents the probability that individual i belongs to class c.

The LCA framework also allows parameterizing π_{ci}, the stochastic process governing class membership. Typically, economists have used socioeconomic characteristics to identify different behavioral types, but the belief-preference model we propose allows differentiating between the determinants of one's beliefs from the drivers of preferences. Without diving too deeply into this matter, socioeconomic variables such as income and family size seem more suitable to identify preference types rather than beliefs about food. For example, wealthy families with children may be more likely to attribute more importance to healthiness than other consumers, while the link with food beliefs seems more tenuous. Thus, we hypothesized that different belief types could be identified by assessing the level of a consumer's involvement with food decisions, as measured by the food involvement (FI) psychometric scale developed by Bell and Marshall (2003) (see Table 1). The FI psychometric scale assigns a score to 12 items assessing a person's involvement through all the stages of food consumption (acquisition, preparation, cooking, eating, and disposal) on a 7-point Likert scale. In the nutrition and food choice literature the concept of food involvement has been successfully used to characterize different attitudes towards food consumption, including organic products (Chen, 2007), loyalty to specific brands (Mittal and Lee, 1989), and many other applications (e.g., Marshall and Bell, 2004; Eertmans et al., 2005; Barker et al., 2008).[16]

After calculating cumulative FI-scores for each participant (higher scores indicate greater food involvement), the probability that individual i belongs to class c in equation 13 was parameterized as:

$$\pi_{ci} = \frac{exp\left(\mathbf{F}_i'\boldsymbol{\theta}^c\right)}{\sum_{c=1}^{C} exp\left(\mathbf{F}_i'\boldsymbol{\theta}^c\right)} \tag{14}$$

where the vector $\mathbf{F}_i$ includes a constant and the food involvement score. $\boldsymbol{\theta}^c$ is a vector of class-specific segmentation parameter estimated by imposing the identification restriction $\boldsymbol{\theta}^C = \mathbf{0}$ for class C. As one can note from equations 13 and 14, π_{ci} does not vary across quality dimensions. That is, a belief class presents an overall characterization of how a type-c consumer interpret the cues in $\boldsymbol{X}_{ij}$, as described by the five vectors of belief parameters $\boldsymbol{\beta}_c^q$; $q = 1, \ldots, 5$. Empirically, the model can be estimated by stacking all the data from the five quality sorting tasks into a single vector, and then organizing the regressors in a block-diagonal matrix, as one would do when testing $\boldsymbol{\beta}_c^1 = \boldsymbol{\beta}_c^2 = \boldsymbol{\beta}_c^3 = \boldsymbol{\beta}_c^4 = \boldsymbol{\beta}_c^5$ in a Chow test.

Parameter estimates for the model in equations 13 and 14 are presented in Table 8 for the case of three classes ($C = 3$)[17]. Model parameters were estimated via the EM algorithm, following Bhat (1997) and Train (2008).[18] Three classes of consumer types with different beliefs are identifiable: class 3 (23% of the sample), the base category, is the least interested in the preparation and consumption of food, while Class 1 (30% of sample) and 2 (47%) are more involved. Low involvement

[16]We refer the reader to Bell and Marshall (2003) for a review of this literature

[17]LCM estimation requires to empirically determine the total number of classes C, which is exogenous in equations 13, but this determination is largely inductive and based on the joint consideration of various criteria (Geiser, 2013). As the purpose of our application is largely illustrative, we maintained a strong prior for parsimonious specifications. After examinining class sizes, degree of separation between classes and interpretability, we identified the three class case as the best to present. Larger models would be preferable based on fit criteria (AIC, BIC). However, these statistics do not provide a holistic assessment of the two-step belief-preference model, so they may be prone to overfitting.

[18]Estimation of latent class model was carried out in STATA, using an EM procedure developed by Pacifico and Yoo (2013)

consumers (Class 3) display a clear preference for chicken, which they find to be better than salmon in almost all quality dimensions. Other than finding products on the shelf to be more convenient, they made little use of the quality cues we presented. In a way, this is fitting: low food involvement participants may not pay attention to quality cues, because they simply don't care about food choices.

The higher involvement classes share a dislike for prepackaged shelved food, which they believe to be of lower quality. The most noticeable difference between Class 1 and 2 relates to the comparison between chicken and salmon: Class 1 thinks that salmon is fresher, tastier, safer and healthier than chicken, but somewhat less convenient. Class 2 thinks that chicken is much better than salmon in all quality dimensions but healthiness. Class 2 also displays a very strong aversion to any type of food processing: they think that prepackaging (shelf) and longer expiration dates decrease all dimensions of product quality without any real gain in convenience, and categorically reject the use of MAP, particularly for its perceived effect on taste and food safety. On the other hand, Class 1 appears more balanced in the evaluation of pros and cons: they interpret longer expiration dates as a signal of higher quality (especially freshness and healthiness), and concede that prepackaged products on the shelf are more convenient, even though the other quality dimensions suffer. They also display a negative attitude towards MAP packaging, but not quite as categorical as Class 2. From a policy/marketing perspective, our results indicate that decreased perceived freshness, healthiness and, for a subset oc consumers, taste are the main reason for the negative WTP for MAP observed in the reduced form model. The implication is that investments in improving the technology and/or consumer messaging should be directed to address these perceptions.

Analogously to the beliefs model, a latent class specification of the utility model with heterogeneous preferences in equation 9 takes the form

$$Pr_{ij} = \sum_{d=1}^{D} \pi_{di} \left[\frac{exp\left(\boldsymbol{Q}'_{ij}\boldsymbol{\gamma}_1^d + Price_j \gamma_2^d\right)}{\sum_{j=1}^{J} exp\left(\boldsymbol{Q}'_{ij}\boldsymbol{\gamma}_1^d + Price_j \gamma_2^d\right)} \right]; \tag{15}$$

where Pr_{ij} is the probability that product j yields the highest utility among all the J available products, and π_{di} is the probability that consumer i displays d-type preferences among D possible classes. The probability of belonging to different preference types, was represented as

$$\pi_{di} = \frac{exp\left(\mathbf{Z}_i'\boldsymbol{\xi}^d\right)}{\sum_{d=1}^{D} exp\left(\mathbf{Z}_i'\boldsymbol{\xi}^d\right)} \tag{16}$$

where $\mathbf{Z}_i$ includes a constant, household income, respondent's age, gender, and the educational attainment. Again, $\boldsymbol{\xi}^d$ represents the segmentation parameters to be estimated for each $d = 1, ..., D-1$, with $\boldsymbol{\xi}^D = \mathbf{0}$ for identification purposes. The quality profiles $\mathbf{Q}_{ij}$ in equation 15 is again replaced by the predicted quality perceptions $\hat{\boldsymbol{Q}}_{ij}$ from the belief model in 13 and 14, but given the latent class framework the q^{th} entry was calculated as $\hat{Q}_{ij}^q = \sum_{c=1}^{C} \hat{\pi}_{c|i}\left(\boldsymbol{X}_{ijt}'\hat{\boldsymbol{\beta}}_c^q\right)$, where $\hat{\pi}_{c|i}$ is the posterior probability of class membership (see Greene and Hensher, 2003), and then aggregated to the two dimension of convenience and overall quality, as show in equation 12.

Parameter estimates are presented in Table 9. Again, three classes were identified based on consumer demographics: compared to the base class (Class 3) the representative consumer in the first class has higher income, and tend to be younger males with college education and children in the household. The second class was not very different from the third class, but consumers in Class 2 are younger. Not surprisingly, consumers in Class 1 are the least price sensitive (they also tend to be richer), and have the highest WTP for quality and convenience, while class 3 participants are the most price sensitive and least attentive to product quality. Consumers in Class 2 are situated in-between these two more extreme classes. Separation in classes in the preference model is not as stark as in belief models: if we exclude the alternative-specific constants, all significant coefficients have the same sign. Jointly, the two latent class models of beliefs and preferences identify nine ($C = 3 * D = 3$) beliefs-preference types that could be described in a more in-depth marketing analysis[19].

5 Discussion and Conclusions

The importance of incorporating subjective probabilities in models of choice to identify consumer preferences has been pointed out in recent literature. In this paper we formalized the theoretical framework and logical underpinnings of a belief-preference model suited to study consumer choices in markets for experience and credence goods. We first showed how the typical specification of consumer choice as a function of product characteristics is in fact a reduced-form model where the primitive parameters—beliefs and preferences—cannot be recovered. Then, we presented a novel DCE procedure comprised of two steps (quality sorting and product choice) and documented its many advantages.

Unlike other methods requiring participants to articulate beliefs in a probabilistic or other form, the elicitation process is easily understood by participants, as choosing products after comparing quality are familiar tasks. For researchers, survey design implies minor modifications of standard DCE techniques, and belief/preference parameters can be estimated within well-established modeling frameworks, even when heterogeneous beliefs are modeled *vis-à-vis* to heterogeneous preferences. Another substantive advantage of our empirical approach, which leverages predicted quality perceptions to estimate preferences, is that the intricate issue of endogeneity arising from using subjective data in regression models is bypassed altogether. In the preference model, choices are conditioned on exogenous product characteristics and estimated belief parameters, so it is not necessary to overlay experimental information treatments to induce exogenous variation in beliefs (as for example in Teisl and Roe, 2010), or attempting to find legitimate variables for instrumentation, either observational (as in Lusk et al., 2014) or from the experiment

[19] To keep the focus on broader methodological considerations, we leave such discussion for a more specialized outlet.

(as in Malone and Lusk, 2018).

The application we presented utilizes data from a choice experiment on food alternatives, where we examined the effect of extrinsic cues on product purchase. Estimates of consumer beliefs show that certain cues can increase some dimensions of perceived quality, while diminishing others. For example, prepackaged food on shelf display is considered convenient, but less fresh and healthy than the food offered on the counter. Estimated preference parameters on the other hand allow quantifying the implicit trade-offs between quality, convenience and price, which ultimately govern purchasing decisions. Jointly, the two set of estimates can be used to obtain part-worth premium/discounts decomposing WTP for a given cue through its effects on the different dimensions of product quality. Even in the simplest specification with homogeneous consumers, the belief-preference approach has the fundamental advantage of explaining *why*, on average, people are willing to pay more (or less) for a given product characteristic, generating a number of implications relevant for food labeling policies and/or marketing strategies.

Introducing consumer heterogeneity in beliefs and preferences via latent class modeling added resolution and detail to the decision process schematized in the belief-preference model. Two findings are most revealing. First, the latent-class belief estimates document how consumers can interpret identical information signals in fundamentally different ways. For example, one class of participants interpreted longer expiration dates as a signal of freshness and healthiness, while another drew opposite inference. This goes to prove, once again, that assuming a uniform and rational interpretation of all available information cues in order to ferret policy implications out of choice data will inevitably lead to misleading conclusions.

For preferences on the other hand, some differences in WTP for quality and prices are detectable, but the sign of our estimates never change across consumer classes. This finding is both expected and reassuring: after all, quality is, by definition, “what everyone wants more of.” Contrasting the two set of results is instructive however, because it contains the seed of a deeper consideration: a substantive amount of what economists call "unexplained heterogeneity in preferences" might be owed more to differences in beliefs, rather than preferences.

While the choice experiment presented here is a useful proof of concept and first application of the belief-preference model, we are well aware that much remains unresolved. From a strictly methodological point of view, the most limiting constraint we faced was imposed by the correlation between quality perceptions, which can arise regardless of orthogonality in the descriptive cues/levels. This suggests that, to attain a finer identification of the preference parameters, innovative experimental design techniques will need to be devised to minimize correlations between regressors in both estimation steps. The two-step estimation process itself, while simple to implement and effective, could perhaps be ameliorated by jointly estimating beliefs and preferences in a unified parametric approach. In addition to plausible efficiency gains, joint estimation would provide a more straightforward way of evaluating model fit (e.g., AIC, BIC, etc.) and selecting the most appropriate specification.

In terms of other applications, the possibilities for applying the belief-preference model are countless, but a few seem particularly enticing. In this paper, the use of the food involvement construct to identify belief classes and socio-demographic variables to distinguish between preference groups was a maintained rather than tested hypothesis. A more detailed investigation into the determinants of beliefs and preferences is a fascinating area for future research. The belief-preference framework would also be useful to study how information and knowledge affect beliefs and choices, and the associated welfare implications (see Foster and Just, 1989; Leggett, 2002). Lastly, augmenting revealed preference purchasing data (e.g. scanner data) with belief information elicited in the way we have shown would take us one step further in the direction advocated by Manski (2004).

References

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Arcidiacono, P. and Jones, J. B. (2003). Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm. *Econometrica*, 71(3):933–946.

Barker, M., Lawrence, W., Woadden, J., Crozier, S. R., and Skinner, T. C. (2008). Women of Lower Educational Attainment Have Lower Food Involvement and Eat Less Fruit and Vegetables. *Appetite*, 50(2-3):464–468.

Bekker-Grob, E. W. D., Hol, L., Donkers, B., and Dam, L. V. (2010). Labeled versus Unlabeled Discrete Choice Experiments in Health Economics: An Applications to Colorectal Cancer Screening. *Value in Heatlh*, 13(2):315–323.

Bell, R. and Marshall, D. W. (2003). The Construct of Food Involvement in Behavioral Research: Scale Development and Validation. *Appetite*, 40(3):235–244.

Bellemare, C. and Manski, C. F. (2011). Introduction: Measurement and Analysis of Subjective Expectations. *Journal of Applied Econometrics*, 26(3):10–11.

Benítez-Silva, H., Buchinsky, M., Chan, H. M., Cheidvasser, S., and Rust, J. (2004). How Large is the Bias in Self-Reported Disability? *Journal of Applied Econometrics*, 19(6):649–670.

Bertrand, B. M. and Mullainathan, S. (2001). Do People Mean What They Say? Implications for Subjective Survey Data. *American Economic Review*, 91(2):67–72.

Bhat, C. R. (1997). An Endogenous Segmentation Mode Choice Model with an Apphcation to Intercity Travel. *Transportation Science*, 31(1):34–49.

Bound, J. (1989). Self-Reported vs. Objective Measures of Health in Retirement Models. *NBER Working Paper*.

Brucks, M., Zeithaml, V. A., and Naylor, G. (2000). Price and Brand Name as Indicators of Quality Dimensions for Consumer Durables. *Journal of the Academy of Marketing Science*, 28(3):359–374.

Brunswik, E. (1955). Representative Design and Probabilistic Theory in a Functional Psychology. *Psychological Review*, 62(3):193–217.

Chang, T.-z. and Wildt, A. R. (1996). Impact of Product Information on the Use of Price as a Quality Cue. *Psychology and Marketing*, 13(1):55–75.

Chen, M. F. (2007). Consumer Attitudes and Purchase Intentions in Relation to Organic Foods in Taiwan: Moderating Effects of Food-Related Personality Traits. *Food Quality and Preference*, 18(7):1008–1021.

Chintagunta, P. K. (1996). Investigating the Effects of A Line Extension or New Brand Introduction on Market Structure. *Marketing Letters*, 7(4):319–328.

Costanigro, M., Deselnicu, O., and Kroll, S. (2015). Food Beliefs: Elicitation, Estimation and Implications for Labeling Policy. *Journal of Agricultural Economics*, 66(1):108–128.

Costanigro, M., Kroll, S., Thilmany, D., and Bunning, M. (2014). Is It Love for Local/Organic or Hate for Conventional ? Asymmetric Effects of Information and Taste on Label Preferences in an Experimental Auction. *Food Quality and Preference*, 31:94–105.

Costanigro, M., McCluskey, J. J., and Goemans, C. (2010). The Economics of Nested Names: Name Specificity, Reputations, and Price Premia. *American Journal of Agricultural Economics*, 92(5):1339–1350.

Costanigro, M., Mittelhammer, R. C., and McCluskey, J. J. (2009). Estimating Class-Specific Parametric Models under Class Uncertainty: Local Polynomial Regression Clustering in an Hedonic Analysis of Wine markets. *Journal of Applied Econometrics*, 24(7):1117–1135.

Darby, M. R. and Karni, E. (1973). Free Competition and the Optimal Amount of Fraud. *Journal of Law and Economics*, 16(1):67–88.

de Bruin, W. B., Fischhoff, B., Millstein, S. G., and Halpern-Felsher, B. L. (2000). Verbal and Numerical Expressions of Probability: "It's a Fifty-Fifty Chance". *Organizational Behavior and Human Decision Processes*, 81(1):115–131.

Delavande, A. (2008). Pill Patch or Shot? Subjective Expectations and Birth Control Choice. *International Economic Review*, 49(3):999–1042.

Delavande, A., Gine, X., and McKenzie, D. (2011). Eliciting Probabilistic Expectations with Visual Aids in Developing Countries: How Sensitive Are Answers to Variations in Elicitation Design? *Journal of Applied Econometrics*, 26(3):479–497.

Dudycha, L. W. and Naylor, J. C. (1966). Characteristics of the Human Inference Process in Complex Choice Behavior Situations. *Organizational Behavior and Human Performance*, 1(1):110–128.

Eertmans, A., Victoir, A., Vansant, G., and Van den Bergh, O. (2005). Food-Related Personality Traits, Food Choice Motives and Food Intake: Mediator and Moderator Relationships. *Food Quality and Preference*, 16(8):714–726.

Fiebig, D. G., Keane, M. P., Louviere, J., and Washi, N. (2010). The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. *Marketing Science*, 29(3):393–421.

Foster, W. and Just, R. E. (1989). Measuring Welfare Effects of Product Contamination with Consumer Uncertainty. *Journal of Environmental Economics and Management*, 17(3):266–283.

Geiser, C. (2013). *Data Analysis with MPlus*. Guiford Press, New York.

Gigerenzer, G. (1991). How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases". *European Review of Social Psychology*, 2(1):83–115.

Gneezy, A., Gneezy, U., and Lauga, D. O. (2014). A Reference-Dependent Model of the Price-Quality Heuristic. *Journal of Marketing Research*, 51(2):153–164.

Grebitus, C., Jensen, H. H., and Roosen, J. (2013). US and German Consumer Preferences for Ground Beef Packaged under a Modified Atmosphere - Different Regulations, Different Behaviour? *Food Policy*, 40:109–118.

Greene, W. H. and Hensher, D. A. (2003). A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit. *Transportation Research Part B: Methodological*, 37(8):681–698.

Grunert, K. G. (2005). Food quality and Safety : Consumer Perception and Demand. *European Journal of Agricultural Economics*, 32(3):369–391.

Heckman, J. J. (2001). Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture. *Journal of Political Economy*, 109(4):673–748.

Hurd, M., van Rooij, M., and Winter, J. (2011). Stock Market Expectations of Dutch Households. *Journal of Applied Econometrics*, 26(3):416–436.

Kamakura, W. and Russell, G. (1989). A Probabilistic Choice Model for Market Segmentation and Elasticity Structure. *Journal of Marketing Research*, 26(4):379–390.

Karni, E. (2009). A Mechanism for Eliciting Probabilities. *Econometrica*, 77(2):603–606.

Keane, M. and Washi, N. (2013). Comparing Alternative Models of Heterogeneity in Consumer Choice Behavior. *Journal of Applied Econometrics*, 28:1018–1045.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2):132–157.

Landon, S. and Smith, C. E. (1998). Quality Expectations, Reputation, and Price. *Southern Economic Journal*, 64(3):628–647.

Leggett, C. G. (2002). Environmental Valuation with Imperfect Information. *Environmental and Resource Economics*, 23(3):343–355.

Lusk, J. L., Schroeder, T. C., and Tonsor, G. T. (2014). Distinguishing Beliefs from Preferences in Food Choice. *European Review of Agricultural Economics*, 41(4):627–655.

Malone, T. and Lusk, J. L. (2017). Taste Trumps Health and Safety: Incorporating Consumer Perceptions into a Discrete Choice Experiment for Meat. *Journal of Agricultural and Applied Economics*, 49(1):139–157.

Malone, T. and Lusk, J. L. (2018). An Instrumental Variable Approach to Distinguishing Perceptions from Preferences for Beer Brands. *Managerial and Ecision Economics*, 39(4):403–417.

Manski, C. F. (2004). Measuring Expectations. *Econometrica*, 72(5):1329–1376.

Marshall, D. and Bell, R. (2004). Relating the Food Involvement Scale to Demographic Variables, Food choice and Other Constructs. *Food Quality and Preference*, 15(7-8 SPEC.ISS.):871–879.

Maynes, E. S. (1976). The Concept and Measurement of Product Quality. *Household Production and Consumption*, 40(5):529–559.

McFadden, D. (1974). *Conditional Logit Analysis of Qualitative Choice Behavior*. New York: Academic Press.

McFadden, D. and Train, K. (2000). Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics*, 15(5):447–470.

Mittal, B. and Lee, M.-S. (1989). A Causal Model of Consumer Involvement. *Journal of Economic Psychology*, 10(3):363–389.

Mussa, M. and Rosen, S. (1978). Monopoly and Product Quality. *Journal of Economic Theory*, 18(2):301–317.

Nelson, P. (1970). Information and Consumer Behavior. *Journal of Political Economy*, 78(2):311–329.

Nelson, P. (1974). Advertising as Information. *Journal of Political Economy*, 82(4):729–754.

Nyarko, Y. and Schotter, A. (2002). An Experimental Study of Belief Learning Using Elicited Beliefs. *Econometrica*, 70(3):971–1005.

Pacifico, D. and Yoo, H. I. (2013). lclogit: A Stata Command for Fitting Latent-Class Conditional Logit Models via the Expectation-Maximization Algorithm. *Stata Journal*, 13(3):625–639.

Parasuraman, A., Zeithaml, V. A., and Berry, L. L. (1985). A Conceptual Model of Service Quality and Its Implications for Future Research. *Journal of Marketing*, 49:41–50.

Riesz, P. C. (1979). Price-Quality Correlation for Packaged Food Products. *Journal of Consumer Affairs*, 13(2):236–247.

Savage, L. J. (1971). Elicitation of Personal Probabilities and Expectations. *Journal of the American Statistical Association*, 66(336):783–801.

Shah, P. and Hall-Phillips, A. (2017). Antecedents and Implications of Expiration Date Search Effort. *Journal of Consumer Affairs*, 52(2):229–251.

Steenkamp, J.-B. E. M. (1990). Conceptual Model of the Quality Perception Process. *Journal of Business Research*, 21:309–333.

Teisl, M. F. and Roe, B. E. (2010). Consumer Willingness-to-Pay to Reduce the Probability of Retail Foodborne Pathogen Contamination. *Food Policy*, 35(6):521–530.

Train, K. E. (2008). EM Algorithms for Nonparametric Estimation of Mixing Distributions. *Journal of Choice Modelling*, 1(1):40–69.

Zafar, B. (2011). Can Subjective Expectations Data Be Used in Choice Models? Evidence on Cognitive Biases. *Journal of Applied Econometrics*, 26:520–544.

Zeithmal, V. (1988). Consumer Perceptions of Price , Quality, and Value: A Means-End Model and Synthesis of Evidence. *Journal of Marketing*, 52(3):2–22.

Zimmer, A. C. (1983). Verbal vs. Numerical Processing of Subjective Probabilities. *Advances in Psychology*, 16:159–182.

Table 1: Sample Descriptive Statistics

		Mean	S.D.
Demographic Characteristics	Age	41.211	(12.83)
	Household Income	70,628	(47,679)
	Female (0/1)	0.45	(0.50)
	Married (0/1)	0.60	(0.49)
	Attended College	0.831	(0.375)
	Kids in HH (0/1)	0.365	(0.482)
Food Involvement Scores (1=strongly disagree, 7=strongly agree)	I don't think much about food each day. (R)	5.00	1.76
	Cooking or babequeing is not much fun. (R)	5.37	1.87
	Talking about what I ate or am going to eat is something I like to do.	4.78	1.63
	Compared with other daily decisions, my food choices are not very important. (R)	4.87	1.94
	When I travel, one of the things I anticipate most is eating the food there.	5.39	1.51
	I do most or all of the clean up after eating.	5.46	1.66
	I enjoy cooking for others and myself.	5.40	1.66
	When I eat out, I don't think or talk much about how the food tastes. (R)	5.25	1.82
	I do not like to mix or chop food. (R)	4.98	1.86
	I do most or all of my own food shopping.	5.88	1.50
	I do not wash dishes or clean the table. (R)	5.79	1.83
	I care whether or not a table is nicely set. Note: (R) indicates reversed scale.	3.55	1.82
			N=1,202

Based on the analytic sample of N=1,202. (R) in food involvement scores indicates a reversed scale.

Table 2: Experimental Design

Descriptors	Constraints	Levels
Product Type (labels)	-	Chicken, Salmon
Display	-	Shelf, Counter
Eat Before	If display=shelf	3, 7, 14 days from today
	If display=counter	“fresh from the counter”
MAP	If Eat Before= “fresh from counter”	never
	If Eat Before=3	never
	If Eat Before= 7	never
	If Eat Before= 14	MAP or no MAP (random blocking)
Prices	If Chicken	\$3.75, \$5.00, \$6.25 /lbs
	If Salmon	\$7.50, \$10.00, \$12.50 /lbs

Table 3: Choice Statistics

		Chicken	Salmon	No Buy/Tie
Quality Sorting Tasks	Freshness	34%	29%	36%
	Taste	27%	30%	44%
	Food Safety	25%	23%	52%
	Convenience	29%	20%	51%
	Healthiness	19%	36%	46%
Product Choice Tasks		58%	28%	15%

Note: Preferred options in the quality sorting and product choice task.

Table 4: Reduced Form Model Estimates

	Coefficients	Implied WTP
Shelf	-0.221***	-$1.06
	(0.037)	
Dates	0.017***	$0.08
	(0.006)	
Map	-0.245***	-$1.18
	(0.087)	
Price	-0.208***	
	(0.017)	
Nobuy	-2.613***	
	(0.159)	
Chicken ASC	0.037	
	(0.067)	

Note: Number of observations=20,634; Number of cases=6,878; Number of individuals=1,202; Log-likelihood=-6434.66. Robust standard errors clustered by individual are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Estimated Beliefs, Aggregated Model

	Freshness	Taste	Safety	Convenience	Healthiness
Shelf	-1.282***	-0.708***	-0.662***	0.284***	-0.743***
	(0.056)	(0.048)	(0.055)	(0.052)	(0.055)
Dates	0.026***	0.009	0.021***	0.010	0.015**
	(0.008)	(0.007)	(0.008)	(0.007)	(0.007)
MAP	-0.473***	-0.332***	-0.256**	-0.126	-0.583***
	(0.108)	(0.097)	(0.112)	(0.104)	(0.106)
Chicken ASC	0.277***	-0.055	0.114**	0.367***	-0.637***
	(0.043)	(0.052)	(0.045)	(0.052)	(0.056)

Note: Number of observations=40,756; Number of cases=20,378; Number of individuals=1,202; Log-likelihood=-12,948.21. Robust standard errors clustered by individual in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Correlation Coefficients among Perceived Qualities

	Freshness	Taste	Food Safety	Convenience
Taste	0.9544			
Food Safety	0.9964	0.9529		
Convenience	-0.6022	-0.7995	-0.6196	
Healthiness	0.7189	0.8912	0.7298	-0.9544

Table 7: Estimated Preference Parameters, Aggregated Model

	Coefficients	Part-Worth WTP Decomposition		
		Shelf	Dates	Map
Composite Quality	0.449***	-\$1.84	\$0.09	-\$0.89
	(0.112)			
Convenience	0.614**	\$0.84	\$0.03	-\$0.36
	(0.313)			
Price	-0.207***			
	(0.017)			
No Buy	-0.697			
	(0.794)			
Chicken ASC	-0.153			
	(0.118)			

Note: Conditional logit estimates of preference parameters. Perceived qualities are obtained using the coefficients in Table 5. Number of observations=20,634; Number of cases=6,878; Number of individuals=1,202; Log-likelihood=-6434.88. Robust standard errors clustered by individual in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Latent Class Quality Logit

	Freshness	Taste	Safety	Convenience	Healthiness
			—Class 1 (High)—		
Shelf	-1.853***	-0.980***	-0.797***	0.499***	-2.031***
	(0.104)	(0.095)	(0.096)	(0.087)	(0.186)
Dates	0.109***	0.069***	0.086***	0.046***	0.144***
	(0.014)	(0.013)	(0.014)	(0.014)	(0.023)
MAP	-0.713***	-0.330*	0.0241	0.161	-1.437***
	(0.177)	(0.177)	(0.192)	(0.194)	(0.319)
Chicken ASC	-0.566***	-0.761***	-0.548***	0.247***	-2.61***
	(0.069)	(0.080)	(0.063)	(0.062)	(0.143)
			—Class 2 (Mid)—		
Shelf	-3.316***	-2.147***	-2.450***	0.022	-2.013***
	(0.234)	(0.188)	(0.240)	(0.136)	(0.185)
Dates	-0.148***	-0.090***	-0.138***	-0.061***	-0.103 ***
	(0.029)	(0.021)	(0.028)	(0.020)	(0.025)
MAP	-1.372***	-1.984***	-1.618***	-0.868***	-1.382***
	(0.499)	(0.473)	(0.495)	(0.277)	(0.440)
Chicken ASC	1.488***	1.600***	1.776***	1.391***	0.009
	(0.167)	(0.147)	(0.194)	(0.101)	(0.132)
			—Class 3 (Low)—		
Shelf	-0.023	0.021	0.121	0.257**	0.173
	(0.123)	(0.096)	(0.107)	(0.103)	(0.111)
Dates	0.027	0.002	0.0369**	-0.000	0.012
	(0.017)	(0.013)	(0.016)	(0.015)	(0.015)
MAP	-0.380	-0.022	-0.304	-0.014	-0.151
	(0.234)	(0.194)	(0.221)	(0.202)	(0.217)
Chicken ASC	1.093***	0.122*	0.380***	-0.070	0.514***
	(0.082)	(0.072)	(0.071)	(0.072)	(0.092)
FIS Class1	0.089***				
	(0.010)				
Constant	-5.075***				
	(0.593)				
FIS Class2	0.082***				
	(0.010)				
Constant	-4.149***				
	(0.563)				

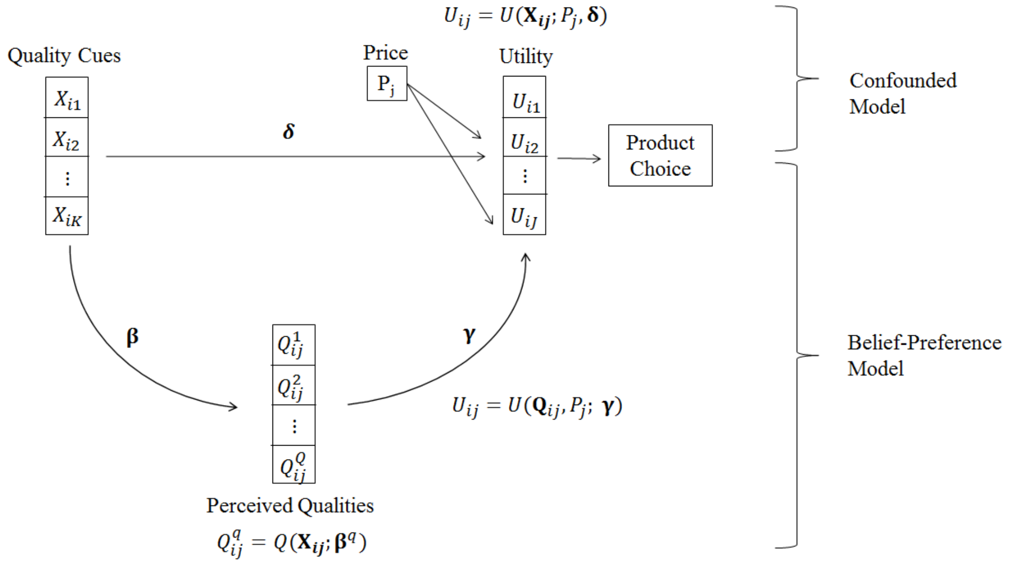
Note: The estimation results from the logit model for quality sorting tasks, where latent class (C=3) are simultaneously estimated. The results are class-specific belief parameters, from Class 1 (high food involvement class) to Class 3 (low food involvement class). The bottom part of the table shows the segmentation parameters (constant and Food Involvement Scale) for Class1 and Class 2, whereas Class3 was the base. Robust standard errors clustered by individual are in parentheses. Number of observations=44,138; Number of cases=22,069, Number of individuals=1,202; log-likelihood=-11228.79. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Product Choice Logit with Class-Weighted Qualities

		Class 1	Class 2	Class 3
Product Choice Parameters	Composite Quality	0.304***	0.432***	0.119**
		(0.038)	(0.051)	(0.051)
	Convenience	0.370***	0.467***	0.353***
		(0.084)	(0.107)	(0.131)
	Price	-0.109***	-0.368***	-0.560***
		(0.036)	(0.044)	(0.060)
	No Buy	1.159*	0.582	-0.181
		(0.662)	(0.923)	(1.072)
	Chicken ASC	-0.854***	0.539***	-1.204***
		(0.163)	(0.137)	(0.186)
Segmentation Parameters	HHincome	0.006***	0.003	–
		(0.002)	(0.002)	–
	Age	-0.049***	-0.034***	–
		(0.009)	(0.007)	–
	Kids	0.769***	-0.082	–
		(0.230)	(0.191)	–
	Female	-0.926***	-0.080	–
		(0.217)	(0.172)	–
	College	0.807**	-0.094	–
		(0.327)	(0.21)	–
	Constant	1.406***	2.266***	–
		(0.558)	(0.434)	–

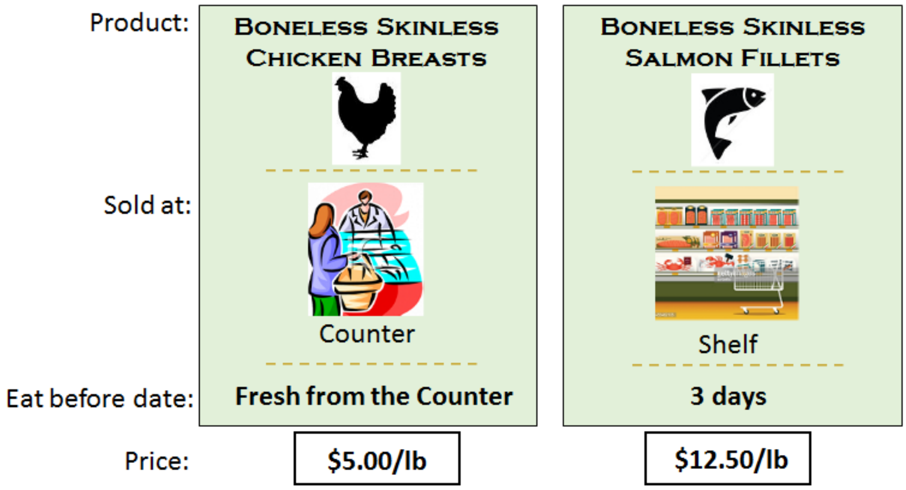
Note: The top half of the table shows the estimated parameters for the product choice logit model with class-weighted perceived qualities estimated in Table 8. The bottom half shows the estimated segmentation parameters. Class 3 is the base so the segmentation parameters are in comparison to the base class. Number of observations=20,634; Number of cases=6,878; Number of individuals=1,202; log-likelihood=-5276.54. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: Conceptual Model



Note: The diagram represents the processes through which observed cues affect the product choices. The top part of the diagram represents the reduced form model (equation 1) where beliefs and preferences are not separated. The bottom part of the diagram depicts the belief-preference models (equations 3 and 4) where observed cues are projected to perceived qualities via belief parameters ($\boldsymbol{\beta}$), and perceived qualities and preference parameters ($\boldsymbol{\gamma}$) appear as argument of the utility function.

Figure 2: Representative choice set



Note: Product information except for price was shown in the quality sorting task, in which respondents were asked to select a product that they think has a higher quality in each of five criteria; freshness, (good) taste, food safety, convenience and healthiness (ties allowed). In the product choice task, the full figure including price was shown, and respondents were asked to select a product they would purchase (opting-out allowed).