Consumer Perceptions of Sustainable Farming Practices: A Best-Worst Scenario

Selected Paper prepared for presentation at the Agricultural and Applied Economics Association's 2011 AAEA and NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania, July 24-26, 2011.

Hillary Sackett*
Department of Agriculture, Food, and Resource Economics
Michigan State University
sacket14@msu.edu

Robert Shupp
Department of Agriculture, Food, and Resource Economics
Michigan State University
shupprob@msu.edu

Glynn Tonsor
Department of Agricultural Economics
Kansas State University
gtt@agecon.ksu.edu

Abstract This paper uses data collected in the summer and fall of 2010 from a national, web-based survey of 1002 households to initiate the process of examining consumer inferences and valuations of food products making "sustainably produced" claims. A Best-Worst scaling framework was implemented to identify what consumers believe "sustainably produced" labels mean and their preferences for each of the individual sustainable farming practices. The best-worst survey method forces respondents to make trade-offs by simultaneously choosing the most and least preferred attributes. The measured level of concern is then applied to a ratio scale. The ability of a firm to differentiate their product hinges critically on an accurate understanding of the perceptions consumers hold regarding what a credence labelling claim implies. Building upon existing work evaluating other food attribute labels (e.g., genetically-modified products, region of origin, use of growth hormones) and the impact of consumer inferences (e.g., implicit associations made from explicitly provided information), this work begins to address gaps in the literature regarding food products with "sustainably produced" claims.

Keywords: Sustainably Produced Food, Best-Worst, Consumer Perceptions

JEL classification: Q01; Q13; Q11

Financial support from the United Stated Department of Agriculture is gratefully acknowledged.

**Corresponding author: Department of Agricultural, Food, and Resource Economics, Michigan State University,East Lansing, MI 48824,*

Email: sacket14@msu.edu, Phone: (517) 505-5677, Fax: (517) 432-1800

1 Introduction

Food produced using "sustainable" production practices is receiving increasing attention in both public and private arenas. More food products are being marketed using "sustainable" or "sustainably produced" certification claims for differentiation. As interest in sustainably produced food rises, questions arise regarding what consumers perceive when faced with "sustainably produced" labels. Moreover, what is the corresponding demand for products making such claims? The viability and contribution of farms to food system sustainability has begun to rely on the exploitation of high-value niche markets for their products, as previously done with "organic". Before investing heavily in "sustainably produced" labeling schemes, consumer perceptions about sustainably produced foods, their willingness-to-pay for such products and the degree to which perceptions and price premiums can be altered by information about sustainable production practices need to be better understood.

Relatively few economic studies have focused on sustainability attributes in the context of agricultural production practices. The extent to which production practices can be considered sustainable in a variety of technical fields has been addressed in a survey conducted by Calker et al. (2005). The suggestions of Callens and Tyteca (1999), on a framework for evaluating the productive efficiency of agricultural firms in the context of overall sustainability using economic, social, and environmental metrics has been expanded upon by Rigby and Caceres (2001) and Rigby et al. (2001) in their study on the relationships between sustainability and organic farming practices. Similarly, Clonan et. al. (2010) have used seven guiding sustainability principles to assess UK consumer's priorities toward sustainable foods. Most recently, Santimanon and Weatherspoon (2010) used hedonic analysis to determine price premiums of sustainable attributes for fresh eggs. Additionally, Saunders et. al. (2010) have investigated consumer purchasing decisions toward sustainability claims on food in the context of carbon emission reduction.

Within this limited literature on sustainability, very little of the research has focused on consumer perceptions and corresponding demand for sustainable production practices and resulting food products. Tonsor and Shupp (2009) conducted preliminary analysis on consumer perceptions of "sustainably produced" food labels which has formed the basis of this study. Previously Lusk and Parker (2009) identified consumer preferences for amount and type of fat in ground beef utilizing a similar survey instrument structure to this work. Consumer preferences and willingness-to-pay for sustainable labeling in the context of farm production practices has been considered by Umberger et al. (2002) in the context of corn versus grass-fed beef, and Onozaka, Nurse, and McFadden (2010) and Onozaka and McFadden (2011) in the context of fresh produce claims.

The motivation for using a Best-Worst framework to assess consumer preferences has been largely inspired by the work of Mueller and Rungie (2009) on utility components that drive distinct consumer segments in the wine market, Umberger, Stringer and Mueller (2010) on market channel choice by small farmers in Indonesia, Flynn et al. (2007) on application of best-worst scaling for health care research, and Lusk and Briggeman(2008) on identifying consumer groups with similar food values. Casini and Corsi (2009) and Cohen (2009) utilized best-worst scaling methodology in response to the tourist management industry in wine marketing. Also, Magidson and Vermunt (2001) advanced the strength of best-worst data using latent class factor and cluster models in sociological contexts. Also of note, the models developed by Marley and Louviere (2005) on probabilistic models of best-worst choices provide the theoretical foundations for this analysis.

The core objective of this research is to initiate the process of examining consumer inferences and valuations of food products making "sustainably produced" claims. This analysis aims to first identify what consumers believe "sustainably produced" labels mean and to determine the relative importance of such attributes to the labeling scheme. Secondly, it aims to determine which attributes drive consumer segments for application in targeted marketing of sustainable food production practices.

Since "sustainably produced" is an attribute that, as of now, has no absolute definition and is thus much more open to consumer perceptions and inferences, we believe this topic could generate significant discourse. In the context of our study, the fact that there is no absolute definition raises the question, what do consumers infer from claims of "sustainably produced?" As noted by Darby et al. (2008) in their evaluation of "locally produced" foods, the ability of a firm to differentiate their product hinges critically on an accurate understanding of the perceptions consumers hold regarding what a credence labeling claim implies. Building upon existing work evaluating other food attribute labels (e.g., genetically-modified products, region of origin, use of growth hormones) and the impact of consumer inferences (e.g., implicit associations made from explicitly provided information), we seek to begin addressing these gaps in the literature regarding food products with "sustainably produced" claims.

2 Research Methodology

A national, web-based survey of 1002 households was developed to collect the data used for this analysis in the summer and fall of 2010. Five hundred of these households are from Michigan while the remaining 500 households are drawn from the other 49 states. The survey instrument was designed to address the research objectives such that respondents could easily identify their preferences. Marketing surveys that aim to measure the level of importance of given attributes often employ a scaled rating system approach. However, this method has several weaknesses. First, scaled rating systems do not force the respondents to make trade-offs between attributes as it is common for people to rate all attributes as "very important". An additional criticism of this method has been a confusion over a natural interpretation of the results since a number scale has meaning only inside the survey context. To help address these issues, this survey uses a best-worst analysis (See Lusk 2009, Mueller and Rungie 2009, and Flynn and Louviere 2006) to investigate preferences for variable sustainable farming practices.

Best-worst analysis requires the survey respondent to choose the most and least preferred attributes out of a subset of competing options, simultaneously. This method is also often referred to as “maximum difference”. The measured level of importance from best-worst analysis is applied to a ratio scale that allows the reader to determine with more certainty that one attribute is more important than another attribute by X percent. Respondents of this survey were asked to answer a series of six questions to determine preferences for sustainable farming methods. The ten farming practices included were devised using information from the USDA and third party sustainable certification companies as follows, for produce (apples in our survey): ground cover management, fertilizer use, pesticide use, herbicide use, pollinator management, pest control, farm size, geographic level of production, consumer food prices, and financial stability of farmers. Similarly the ten farming practices devised for meat (rib-eye steaks) were: use of antibiotics, use of growth hormones, use of genetically modified stock, animal health and safety, feed and waste management, pest control, farm size, geographic level of production, consumer food prices, and financial stability of farmers.

In designing the best-worst scenarios, the choice sets were created using a main effects fractional factorial design, implemented in SPSS. Six of these choice sets were chosen from the full factorial design such that the presence or absence of a particular farming practice was independent of the presence or absence of another. The choice made by the respondent can be conceptualized as choosing the two attributes that maximize the difference between them on an underlying scale of preference. As noted, this model is useful because it can identify the relative preferences of consumers on a ratio scale. The stated preference methods outlined here are intended to approximate observed consumer behavior in real markets. The systematic variation of attribute choice sets is used to estimate the utility gained by the consumer across attribute levels.

Traditional discrete choice questions fail to address relative utility impacts across attributes. Best-worst scaling as originally devised by Flynn and Louviere in 1992 is capable of analyzing the efficiency of choice tasks as presented to a respondent. The specification of attributes from a choice set of competing scenarios, repeated over a number of variable choice sets allows observation of trade-off behavior. Best-worst tasks provide more information than single choice designs while forcing respondents to consider the extremes of their utility space. The exclusion of an ”opt-out” option infers that the decisions made by the consumer are conditional on the respondent participating in the market. The additional utility or dis-utility from moving between attribute levels can be estimated by a probit or logit model. This model captures the systematic propensity to choose one attribute over another across all choice sets and respondents.

The choice experiment determined for this analysis is developed around $K = 10$ attributes with L_k levels of attribute k. For our purposes,$L_k = 2\ \forall k \in [1, K]$. The two levels simply indicate existence or in this case whether or not the good in question was farmed with or without adherence to a particular sustainable farming practice. Scenarios are presented one at a time to the respondents, in which they indicate which attribute exhibits the highest and lowest utility impacts for them.

Best-worst scaling has foundations in Random Utility Theory as hypothesized by Thurstone (1927) and generalized by

McFadden (1974). The formal statistical properties of this method were only recently proven by Marley and Louviere in 2005. The statistical assumptions underlying best-worst analysis theory propose that the proportional distance between two attribute levels on a latent utility scale represents the relative choice probability of a given pair of attributes. Cognitively, respondents are undertaking the task of identifying every pairing of attributes possible, calculating the difference in utility between every attribute pair, and choosing the pair that maximizes the utility difference between them. The distances between attributes are then modeled as a difference model with the result that the pair-wise utilities are estimated in relation to a single attribute level rather than to an entire scenario.

The number of unique best-worst pairings to be chosen in each scenario as well as the reverse worst-best pairings is given by,

$$P(K) = 2[(K-1) + (K-2) + (K-3) + ... + 2 + 1] = P(10) = 2[9+8+7+6+5+4+3+2+1] = 90 \tag{1}$$

In other words, algebraically there are

$$\sum_{j=1}^{K-1} j = (2(K-1)K)/2 = (K-1)K = 10(9) = 90 \tag{2}$$

pairs. The total number of possible scenarios to be presented to the respondent is given by,

$$levels^{attributes} = 2^{10} \tag{3}$$

where the two levels can be represented by a dummy variable

$$level = 1 \rightarrow included \tag{4}$$

$$level = 0 \rightarrow excluded \tag{5}$$

3 Data Collection

If, in each situation, each attribute has the same number of levels then the design is said to be balanced. The maximum number of possible scenarios, given this design, is given by the product of the number of levels across all attributes, in this case, 2^{10}, as derived above. For brevity, we chose to administer two versions of the orthogonal main effects plan indicated by "Block 1" or "Block 2". Respondents were randomly assigned to respond to only one block of six best-worst scenarios. Figure 1 shows an example scenario from the apple survey, as seen by respondents.

3.1 Example

Which one of the following aspects of apple farming do you believe are the most and least important in a sustainable apple production system? Please check only one in each column.

Least Important		Most Important
	Ground Cover and Area Management Practices are Employed	
	Little to No Chemical Pesticides are Used for Pest Management	
	Pollinator Management is Employed	
	Other Pests are controlled using preventative measures, and habitat controls	
	Production, distribution, and sale is done locally	
	Consumer food prices are affordable	

The orthogonal main effects pairing design ensured that each potential best-worst pair appeared exactly twice in each Block of six scenarios. The four economic attributes were left to be interpreted by the respondent. The definitions of all non-economic attributes were provided prior to the best-worst block questions as follows:

In the next section you will be asked to choose which aspects of sustainable apple farming are most and least important to you. Please take the time to read the following definitions as related to sustainable production practices to better help you in your responses.

Ground Cover and Area Management Practices are Employed: Adjacent areas are planted with hedgerows, windbreaks, or other low-maintenance plantings to encourage specific beneficial organisms. Within tree rows, ground cover or mulch are selected and maintained to improve soil microbial activity, organic matter levels and nutrient cycling.

Fertilizer and Nutrient Materials are used minimally: Soil quality, including organic matter content, is established at planting and maintained at an optimum level to minimize commercial fertilizer needs.

Little to No Chemical Pesticides are Used for Pest Management: Chemical pesticides are not used. Alternative strategies are employed, including biopesticides, mating disruption, trap out and/or augmentation with beneficial organisms.

Little to No Chemical Herbicides are Used for Weed Management: Soil quality and ground cover in the orchard and adjoining areas are planned and managed to prevent weeds and weed seed immigration into the orchard. Cultural, mechanical or biological methods are used to control weeds.

Pollinator Management is Employed: Bees are not placed in the orchard until blossoms are open. Pesticides hazardous to bees are not used, or if needed in an emergency, are applied such that they are not hazardous to bees.

Other Pests are controlled using preventative measures, and habitat controls: Habitat is modified around orchards to reduce nesting and perching sites for pest birds.

Additionally, for the beef survey:

In the next section you will be asked to choose which aspects of sustainable cattle farming are most and least important to you. Please take the time to read the following definitions as related to sustainable production practices to better help you in your responses.

Prohibited use of sub-therapeutic antibiotics : Animals may only be treated with antibiotics when necessary for treatment of illnesses, provided they are not slaughtered within 45 days of last treatment.

Prohibited use of growth hormones: The use of hormone treatments, including implants, to enhance growth is not permitted.

Prohibited use of genetically modified livestock: Animals produced through embryo transfer and those whose genetic material has been altered are not permitted.

Animal Health and Safety are Protected: Animal nutrition on the farm results in superior health as related to breeding success, weight gain, and freedom from illness. Policies are in effect for low-stress handling, preventative health measures, and regular maintenance and repair of facilities so as to prevent injury.

Feed is Pasture Based and Waste Management Systems Employed: Cattle receive majority of nutritional intake through grazing activity and animal movement is directed based on cattle's natural action and reaction to the situation. Manure resources are used to close the nutrient cycle on the farm, but only to the extent that overall nutrient levels are adequate and not excessive. Excess manure, if any, is put to good use off farm.

Pests are controlled using preventative measures, cultural and nutritional controls: Preventative measures and/or cultural controls such as movement of cattle, sanitation, and composting are used to reduce or eliminate the need for insecticides and miticides. Animals are free to choose and move to habitats that are most comfortable such as shady areas, windy spots, or wallows.

4 Analysis

Analyzing choice data from a best-worst model is less straight forward than in other more traditional discrete choice methods. Common statistical packages like Stata currently do not have standardized commands for analyzing best-worst data. Therefore, this data was manipulated manually using Microsoft Excel and a program coded into SAS for further exposition. There are two primary ways of approaching best-worst data analysis. "Paired" models are used to make inferences about the latent utility scale, while "marginal" models aggregate over all pairs that include a given attribute level to model choice frequencies. This work concentrates on a marginal analysis approach. Sackett, Shupp, and Tonsor intend to further this analysis using paired models in the continuation of this work. The paired model methods are examined here for comparison to the marginal analysis. Both methods have the same measurement properties and can be analyzed at the respondent or sample level.

4.1 Paired Model Analysis

Paired analysis models the possible best-worst pairs that a respondent may choose. The number of observations is equal to the number of unique best-worst pairs (ninety). In a balanced design, such as the one used here, every attribute has the same number of levels (two) and each possible pair will be available to be chosen the same number of times (four). Each survey version (apple and beef) is analyzed separately, each with ninety observations representing the unique best-worst pairs, each of which appears twice in each Block, for a total of four times in each survey. The impact weight for attribute k takes a value of one for all pairs in which attribute k was chosen as the most desirable and a value of minus one for all pairs in which attribute k was chosen as least desirable. These impact weights form the explanatory variables in the final regression used to estimate partial utility gain or loss. The equation to be estimated is given by

$$ln(c_1) = \alpha + \beta_1 L_1 + \beta_2 L_2 + \beta_3 L_3 + \beta_4 L_4 + ... + \beta_{10} L_{10} + \beta_{1,0} L_1^0 + \beta_{1,1} L_1^1 + ... + \beta_{10,0} L_{10}^0 + \beta_{10,1} L_{10}^1 \quad (6)$$

where c_1 is the total number of times a particular best-worst pair was chosen across all scenarios and all respondents. L_k^0 indicates a scenario in which attribute k is excluded. Similarly, L_k^1 indicates a scenario in which attribute k is included. Sampling zeros were adjusted by adding the reciprocal of the sample size, as suggested by Flynn and Louviere, to enable logs to be taken. The natural log of c, the total number of times a particular pair was chosen, is a linear function of the difference in utility acquired from each attribute. The parameter values estimated represent the average utility across the entire sample gained (or lost) by the particular sustainable farming production practice. The data can be used further to discover the extent of differences among subgroups of attribute impacts where subgroups are defined by respondent-level demographic characteristics. This kind of limited dependent variable model requires the difference in probabilities of pair-wise choice for various scenarios in a choice set to be associated with the differences in the explanatory variables. An important note about interpretation of the results needs to be made here. The main effect of a sociodemographic variable, such as age, on utility has no meaning in this context. However, the effect that age has on the utility gained for a specific attribute does have a meaningful interpretation. For our purposes each of the $2^{10} = 1024$ scenarios has $(K-1)K = 90$ observations yielding a data set of $92,160$ possible pair-wise observations. The most flexible method of analysis for this format is to run a logit model using common statistical software on the expanded data set. For example, each sustainable farming practices scenario for each person has 90 observations coded as the independent variables while the dependent variable is an indicator variable taking a value of one for the pair chosen and a value of zero if it is not chosen.

4.2 Marginal Model Analysis

The marginal method of analysis models the potential attribute levels that can be chosen. For the purposes of this work, each attribute is either strictly employed by the farmer for the good in question or is not. This method of analysis aggregates the data over best-worst pairs across all respondents to determine the utility gained by inclusion of a particular attribute. There are a total of $2\sum_{k=1}^{K} L_k = 2(20) = 40$ observations. Each of the attribute levels contribute two observations, a best and a worst total. There are $K-1 = 9$ impact variables and $L_k - 1 = 1$ effect coded scale level variables for each of the $K = 10$ attributes. The best-worst indicator variables take a value of one for all observations where the particular best-worst pair is chosen and negative one when the pair is chosen in the reverse order. The equation to be estimated is given by

$$ln(c_2) = \alpha + bw_{indicator} + \beta_1 L_1 + \beta_2 L_2 + \beta_3 L_3 + \beta_4 L_4 + ... + \beta_{10} L_{10} + \beta_{1,0} L_1^0 + \beta_{1,1} L_1^1 + ... + \beta_{10,0} L_{10}^0 + \beta_{10,1} L_{10}^1 \quad (7)$$

where c_2 is the total number of times a particular attribute was chosen across all scenarios and all respondents, with a similar adjustment for observations of zero. Therefore, there are $\sum_{k=1}^{K} L_k = 20$ best totals and similarly $\sum_{k=1}^{K} L_k = 20$ worst totals. The regression results should be consistent with those of the paired method analysis. To analyses this marginal model in a multinomial framework, a logit model can again be estimated using common statistical software where each respondent contributes twenty observations; ten attributes that can be picked as best and ten attributes that can be picked as worst. If the most and least important farming production practices differ among consumers, we can classify consumers into unique segments of buyers allowing producers and certifying agencies to better understand the targeted consumer base for applications in marketing and policy.

Individual best-worst scores were determined for each attribute by the summation of the number of times each respondent indicated the attribute was most important less the summation of the number of times each respondent indicated the attribute was least important. The larger the B-W score, the more important the specific attribute is to the individual. The individual attribute sums were aggregated across the sample to obtain the aggregate measure of B-W for each attribute. Using a standardized interval scale, the relative importance of each attribute can be more easily interpreted. First the square root was taken of the aggregate frequency of best divided by the aggregate frequency of worst for each attribute. The highest $\sqrt{\frac{BEST}{WORST}}$ is scaled to 100 and all other attributes scaled relative to this attribute.

4.3 Latent Class Analysis

Latent Class cluster analysis examines the heterogeneity of consumers in their ratings of farming production attributes and whether unique segments of consumers exist that can be explained by household and targeted marketing characteristics. This clustering technique assumes that individuals belong to one of L latent classes of a size pre-determined. The most common clustering methods involve minimizing within cluster variance and maximizing across cluster variance. Latent class clustering techniques estimate the probability of membership using the model parameters and observes impact measures of individual respondents. The covariation across individual observed preference scores measure utility to predict each respondent's unique membership within a particular latent class. Unobserved utility is heterogeneous across classes and homogeneous within a class. Using best-worst scaling we measure the individual importance of sustainable farming attributes to consumers.

The latent class cluster analysis uses the 1002 individual B-W scores as dependent variables to explore the heterogeneity across consumers in their perceptions of sustainable farming practices. Once the clusters are separated, ANOVA is used to determine significant differences in respondent characteristics across clusters. The relative importance of each attribute to the sample is determined by evaluating the standardized interval scale and explained above. Additionally, an objective of this analysis is to determine which consumers are more or less likely to respond to certain marketing channel attributes.

In addition to the Latent Class clustering model, a principle component analysis is used to derive distinct utility components that drive consumer behavior in each segment. The distinction of the principle components allows greater comparison of the segments across multiple utility dimensions. Principle component analysis is very useful in linking heterogeneity with the underlying drivers of consumer behavior.

5 Results

By first calculating the variance-covariance matrix from the individual B-W scores, the attribute importance heterogeneity and co-relations of attributes is made more clear. Higher variance indicates a greater degree of heterogeneity across respondents. High covariance metrics indicate a strong relationship between attributes that may be preferred by the same consumer segment. This method of analysis is expanded on in Mueller and Rungie (2009) as applied to wine markets.

Table 1: Apple Production Attribute Variance-Covariance Matrix

	GC	FN	CP	CH	PM	OP	FS	PL	CFP	FF
GC	0.83									
FN	-0.05	0.63								
CP	-0.25	0.04	1.28							
CH	-0.12	0.07	0.46	0.86						
PM	0.12	-0.01	-0.22	-0.12	0.75					
OP	0.09	-0.03	0.08	-0.02	0.02	0.59				
FS	-0.16	-0.26	-0.54	-0.32	-0.18	-0.27	2.58			
PL	-0.03	-0.13	-0.34	-0.25	-0.08	-0.16	-0.03	1.15		
CFP	-0.23	-0.13	-0.22	-0.22	-0.33	-0.11	-0.51	-0.09	1.76	
FF	-0.15	-0.09	-0.24	-0.28	-0.07	-0.15	-0.3	-0.06	0.14	1.19

Table 2: Beef Production Attribute Variance-Covariance Matrix

	SA	GH	GM	AS	PF	OP	FS	PL	CFP	FF
SA	0.71									
GH	0.18	1.31								
GM	-0.09	0.28	1.16							
AS	-0.07	-0.1	0.02	1.19						
PF	-0.01	-0.04	-0.08	0.04	0.54					
OP	-0.05	-0.14	-0.09	-0.06	0.04	0.9				
FS	-0.27	-0.65	-0.35	-0.37	-0.13	-0.22	2.33			
PL	-0.05	-0.26	-0.27	-0.24	-0.15	-0.07	0.05	1.12		
CFP	-0.15	-0.29	-0.36	-0.25	-0.14	-0.11	-0.31	-0.12	1.6	
FF	-0.19	-0.27	-0.22	-0.12	-0.14	-0.18	-0.08	-0.01	0.13	1.1

5.1 Attribute Importance

The attribute that is most important across the sample is the attribute with the highest B-W score. In this example it may not necessarily be the attribute chosen most important with the highest frequency, since it is dependent on the maximum difference score. The mean individual B-W score is the average B-W score for each respondent and is determined by dividing the aggregate B-W score by the sample size. The relative importance of each attribute is then standardized to a ratio scale that has consistent interpretation outside of the survey context. The ratio scale is interpreted as the probability that a respondent prefers a particular sustainable farming practice attribute over the other nine. The ratio scale is calculated by transforming the square root of the best divided by worst scores to a scale of $[0, 100]$. All measures of attribute importance result in a ranking of the attributes in the same order.

The mean B-W score can be visualized with the aid of Figure 2. The bars represent the net average of the frequency each attribute was chosen as most or least important. Each attribute can be chosen a maximum of four times in this experimental design. For attributes chosen as most important with more frequency than least important the B-W score is greater than zero and is indicated by greater area to the right of center. Attributes in the middle were chosen as most important and least important a similar number of times.

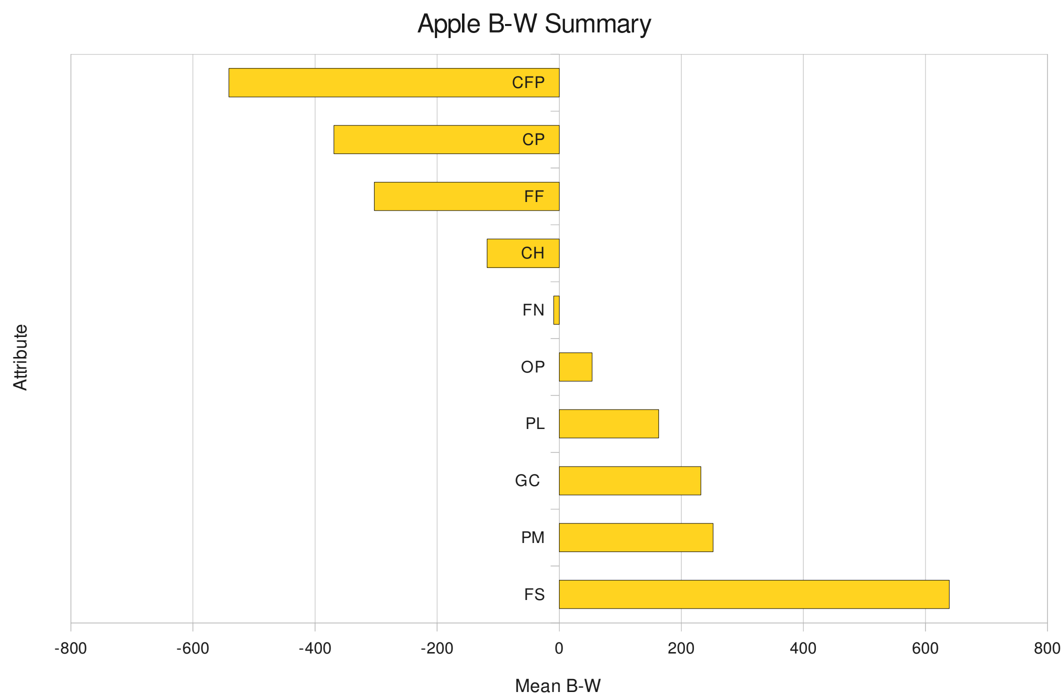
5.1.1 Apple Survey

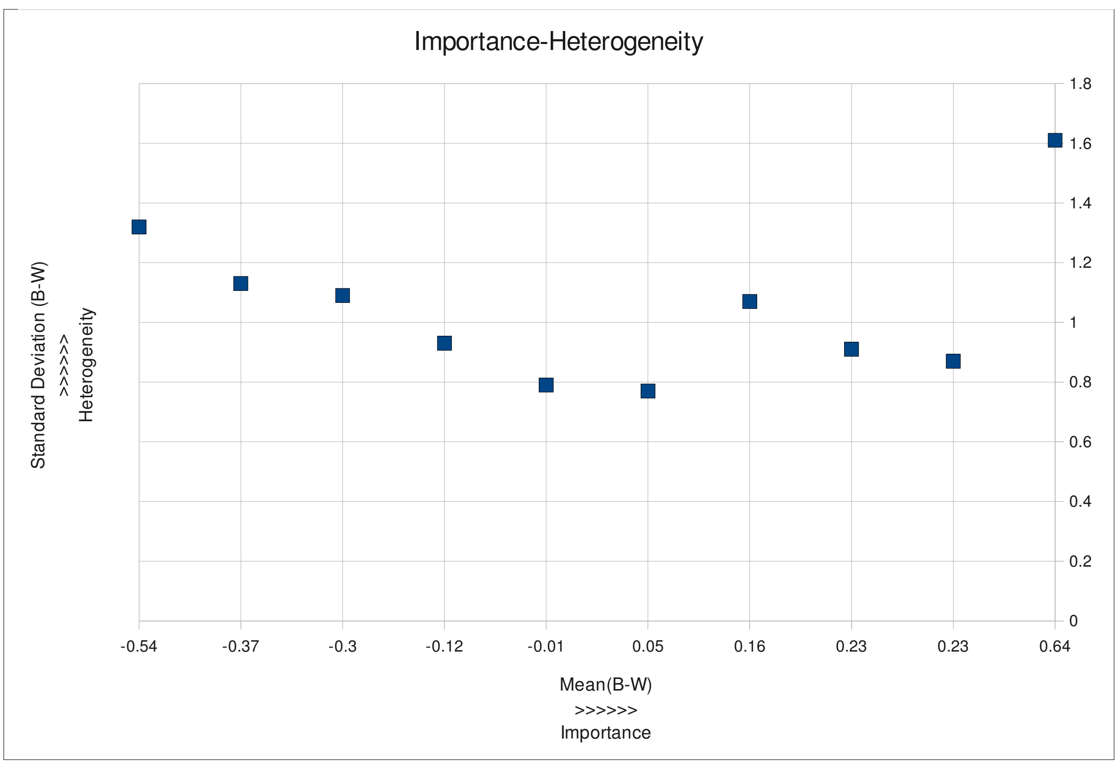
Table 3: Apple Production Attribute Importance Measures using Best-Worst Scaling

Attribute	Most	Least	Aggregate (B-W)	SQRT (B/W)	Standard Ratio	Mean (B-W)	STDEV B-W	Individual Ratio
Farm Size is small and Corporate Involvement is limited	759	120	639	2.51	100	0.64	1.61	100
Pollinator Management is Employed	388	136	252	1.69	67.16	0.23	0.87	35.94
Ground Cover and Area Management Practices are Employed	370	138	232	1.64	65.11	0.23	0.91	35.94
Production, distribution, and sale is done locally	392	229	163	1.31	52.02	0.16	1.07	25
Other Pests are controlled using preventative measures, and habitat controls	247	193	54	1.13	44.98	0.05	0.77	7.81
Fertilizer and Nutrient Materials are used minimally	196	205	-9	0.98	38.88	-0.01	0.79	-1.56
Little to No Chemical Herbicides are Used for Weed Management	223	341	-118	0.81	32.15	-0.12	0.93	-18.75
Michigan farmers are financially stable	135	438	-303	0.56	22.07	-0.3	1.09	-46.88
Little to No Chemical Pesticides are Used for Pest Management	162	531	-369	0.55	21.96	-0.37	1.13	-57.81
Consumer food prices are affordable	140	681	-541	0.45	18.03	-0.54	1.32	-84.38

502 respondents were randomly assigned to complete the survey on apple production practices. Each respondent answered one block of six best-worst questions, yielding 3,012 observations of most important and 3,102 observations of least important. The data reveals that the attribute corresponding to ”farm size is small and corporate involvement is limited” has the highest B-W score of 639, as well as having the highest frequency of choice as most important, 759. The next highest B-W score belongs to the attribute corresponding to ”pollinator management is employed” at 252, suggesting that pollination is valued highly by this sample of consumers for its contribution as an ecosystem service to the farming industry. However, pollinator management has a mean B-W score of 0.23 with a little more than a third of the magnitude of the farm size attribute at 0.64. The attribute corresponding to ”consumer food prices are affordable” has the lowest B-W score of -541. This score indicates that this attribute was chosen as least important more often in each scenario than any of the other attributes. ”Production, distribution, and sale is done locally” also had a relatively high B-W score of 163, ranking fourth among the attributes and about half as important as farm size.

When applied to a ratio scale, farm size is found to be more than five times as important to this sample of consumers than food prices. Also, for apple production the locality attribute ranks only half as important as farm size. The use of off-farm chemical inputs such as fertilize, herbicides, and pesticides all ranked consistently between 3 and four times less important than farm size and only marginally more important than consumer food prices. Overall, two of the four economic attributes were ranked as highly important to this sample of consumers, while the remaining two ranked as unimportant. Consumers appear to perceive apples produced on smaller farms that grow distribute locally to contribute more importance to the sustainable label. The financial well-being of farmers and consumer food prices were not considered by this population to be highly important to the sustainability of the farming practices.


Apple B-W Summary
CFP
CP
FF
CH
FN
OP
PL
GC
PM
FS
Attribute
-800
-600
-400
-200
0
200
400
600
800
Mean B-W


Importance-Heterogeneity
Standard Deviation (B-W)
>>>>>>
Heterogeneity
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.54
-0.37
-0.3
-0.12
-0.01
0.05
0.16
0.23
0.23
0.64
Mean(B-W)
>>>>>>
Importance

Graphs.pdf

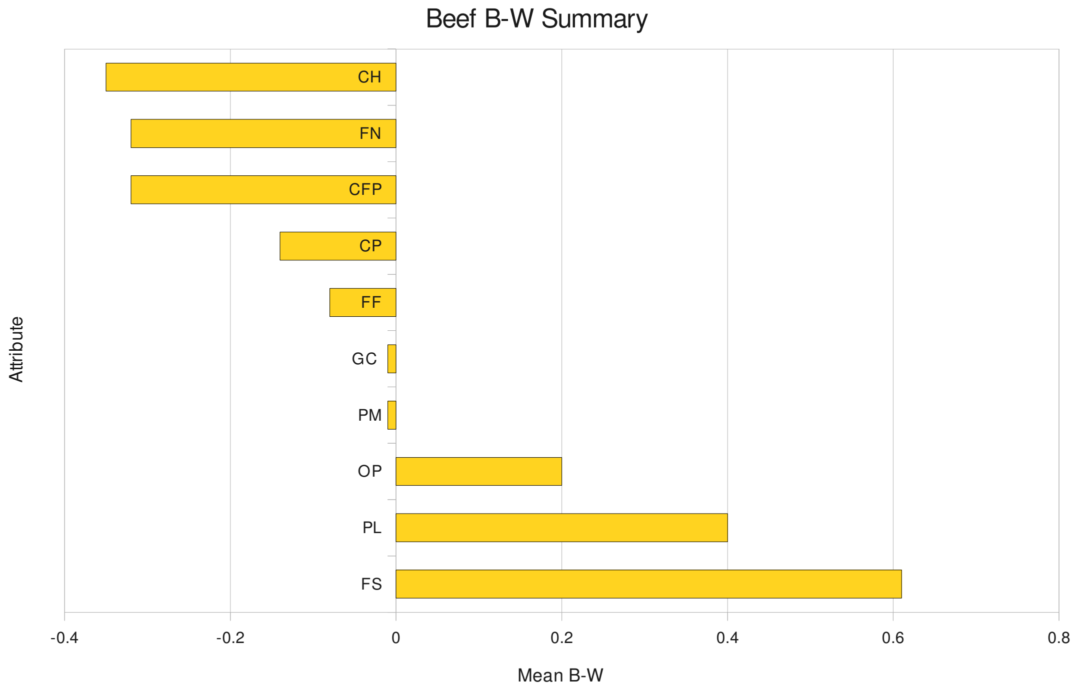
5.1.2 Beef Survey

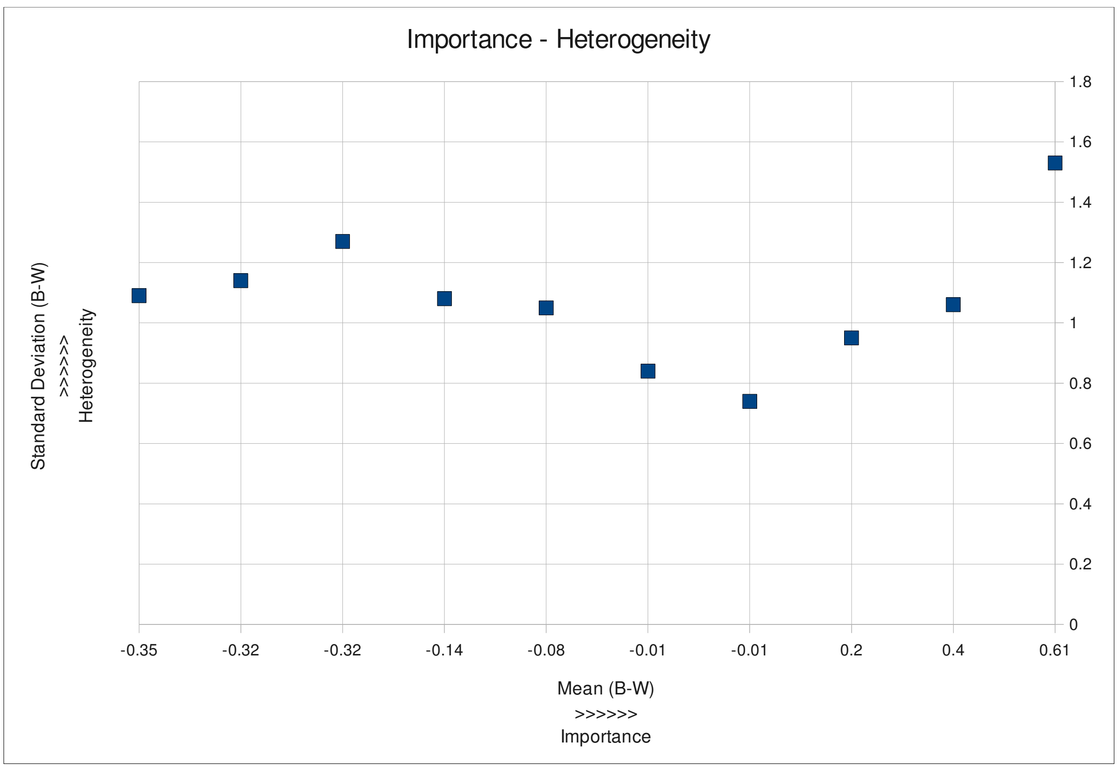
Table 4: Beef Production Attribute Importance Measures using Best-Worst Scaling

Attribute	Most	Least	Aggregate B-W	SQRT(B/W)	Standard Ratio	Mean Individual B-W	STDEV Individual B-W	Individual Ratio
Farm Size is small and Corporate Involvement is limited	729	117	612	2.5	100	0.61	1.53	100
Production, distribution, and sale is done locally	519	119	400	2.09	83.66	0.4	1.06	65.57
Pests are controlled using preventative measures, cultural and nutritional controls	351	153	198	1.51	60.68	0.2	0.95	32.79
Feed is Pasture Based and Waste Management Systems Employed	221	208	13	1.03	41.29	-0.01	0.74	-1.64
Prohibited use of sub-therapeutic antibiotics	216	227	-11	0.98	39.08	-0.01	0.84	-1.64
Michigan farmers are financially stable	223	304	-81	0.86	34.31	-0.08	1.05	-13.11
Prohibited use of genetically modified livestock	238	374	-136	0.8	31.96	-0.14	1.08	-22.95
Consumer food prices are affordable	194	517	-323	0.61	24.54	-0.32	1.27	-52.46
Prohibited use of growth hormones	181	506	-325	0.6	23.96	-0.32	1.14	-52.46
Animal Health and Safety are Protected	128	475	-347	0.52	20.8	-0.35	1.09	-57.38

Similarly, 502 respondents were randomly assigned to complete the survey on beef production practices. Each respondent answered one block of six best-worst questions, yielding 3,012 observations of most important and 3,102 observations of least important. The data reveals again that the attribute corresponding to "farm size is small and corporate involvement is limited" has the highest B-W score of 612, as well as having the highest frequency of choice as most important, 729. The next highest B-W score belongs to the attribute corresponding to "Production, distribution, and sale is done locally" at 400, suggesting that locality of meat production is valued more highly than locality of apple production by this sample of consumers. This result is likely linked to food traceability standards for meat being made a more public issue in the recent past. Locality has a mean B-W score of 0.4 with roughly three quarters of the magnitude of the farm size attribute at 0.61. The attribute corresponding to "animal healthy and safety is protected" has the lowest B-W score of -347. This score indicates that this attribute was chosen as least important more often in each scenario than any of the other attributes. "Consumer food prices are affordable" also had a relatively low B-W score of 323, along with "prohibited use of growth hormones" with a mean B-W score of -325.

When applied to a ratio scale, farm size is found to be more than five times as important to this sample of consumers than animal health and safety. Similar to the apple survey data, local production and distribution was ranked with high importance to sustainable farming practices. Like the results from the apple survey, the respondents of the beef survey ranked financial well-being of farmers and consumer food prices in the bottom third of importance on the ratio scale. The use of sub-therapeutic antibiotics and pasture based feed and waste systems have a near zero mean, indicating indifference among consumers. Overall, the same two of the four economic attributes were ranked as highly important to this sample of consumers, while the remaining two ranked relatively lower in importance. Consumers appear to also perceive small scale, local beef production to contribute higher importance to the sustainable label.


Beef B-W Summary
CH
FN
CFP
CP
FF
GC
PM
OP
PL
FS
Attribute
-0.4
-0.2
0
0.2
0.4
0.6
0.8
Mean B-W


Importance - Heterogeneity
Standard Deviation (B-W)
>>>>>>
Heterogeneity
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.35
-0.32
-0.32
-0.14
-0.08
-0.01
-0.01
0.2
0.4
0.61
Mean (B-W)
>>>>>>
Importance

Graphs.pdf

5.2 Attribute Heterogeneity

The mean B-W score for each attribute does not convey information about the similarity of importance placed on it by all consumers. A mean B-W score that ranks somewhere in the middle of all attributes may be caused by an averaging of respondents for whom the attribute is very important to some and very unimportant to others. This is a good illustration of the case of high consumer heterogeneity. The larger the range of consumer heterogeneity, the more the market will respond to targeting different consumers with variable channels of communication. The standard deviation of each individual B-W score across the sample measures the extent of variation amongst consumers over the relative importance of the attribute. A higher standard deviation indicates a wider variety of importance for a given attribute. Conversely, a smaller standard deviation is indicative of general agreement across consumers on the relative importance of the sustainable farming production attribute. Therefore, the standard deviation is a good statistical measure of consumer heterogeneity across the sample for the attributes in question.

The standard deviation is also bounded by $[-4, 4]$ as each attribute can be chosen as most or least important a maximum of four times. For our purposes, a standard deviation above 1 should be interpreted as high heterogeneity across consumers. Figure 3 shows a visual representation of importance heterogeneity. The nodes found higher on the y-axis correspond to attributes with higher consumer heterogeneity. As demonstrated here, mean B-W scores do not tell the entire story. Attributes at both ends of the mean B-W spectrum exhibit varying degrees of importance heterogeneity. Ideally, marketing managers should exploit the attributes with high importance but additionally use specialized communication channels for attributes with high heterogeneity. Attributes with high B-W mean scores and high heterogeneity indicate greater importance to a select subgroup of consumers. Also, attribute with low B-W mean scores and high heterogeneity have potential in niche markets since it appeals only to a small segment of consumers.

Distinct drivers of heterogeneity are importance to identify to determine which important attributes with high heterogeneity are related or jointly important for the same cluster of consumers. The variance-covariance matrix is useful here for outlining which attribute pairs vary simultaneously. If one attribute scores highly in B-W score, then an attribute that that is highly covariable will also exhibit high score in B-W for the same group. Additionally, attributes that are highly negatively correlated will likewise drive the same segment of consumers but in opposite directions. For this reason, it is often easier to interpret the correlation coefficients because it is bounded in $[-1, 1]$. The basis for the clustering analysis comes from attributes that tend to be tracked together over consumers. The more correlation coefficients that are found to be statistically significant imply a more structured market. The matrix of correlation coefficients is displayed in Figure 4.

Table 5: Apple Production Attribute Correlation Coefficients

	GC	FN	CP	CH	PM	OP	FS	PL	CFP	FF
GC	1									
FN	-0.06	1								
CP	-0.24	0.04	1							
CH	-0.15	0.1	0.45	1						
PM	0.15	-0.01	-0.22	-0.15	1					
OP	0.12	-0.05	0.1	-0.03	0.03	1				
FS	-0.11	-0.2	-0.31	-0.22	-0.12	-0.22	1			
PL	-0.04	-0.15	-0.28	-0.24	-0.09	-0.2	-0.01	1		
CFP	-0.2	-0.13	-0.15	-0.18	-0.28	-0.11	-0.24	-0.06	1	
FF	-0.15	-0.12	-0.19	-0.28	-0.07	-0.18	-0.17	-0.05	0.1	1

5.2.1 Apple Survey

The apple sustainable farming production attributes that exhibit the highest heterogeneity are farm size, consumer food prices, local distribution and production, pesticide use, and financial well-being of farmers, respectively. While consumer food prices is ranked as least important by mean B-W score, the high variance indicates that a small segment of consumers consider it very unimportant, while other segments disagree. Low mean and high variance pertain to products that have large potential in a niche market. In this case, it is likely that a small segment of consumers do not care at all (or nearly at all) about the price of the food as long as it is labeled "sustainably produced", while other consumer respondents indicated some relative importance of food prices. The high mean and high variance on the farm size attribute indicates that this attribute is very important to a smaller subset of consumers. The sample population contained a segment of consumers that placed a high value on small scale farming, while the other respondents gave mixed attribute importance measures. The authors hypothesize that both of these results point to the same or very similar small subset of consumers. It is possible that the consumers that highly value small production and limited corporate involvement understand that the minimized scale of production will cause prices to rise, and they are okay with that. For this small subgroup higher consumer food prices are a reasonable trade off for sustainably produced food products from smaller farms. It is suggested that this hypothesis be tested through factor analysis in the next step of this research.

5.2.2 Beef Survey

The beef sustainable farming production attributes that exhibit the highest heterogeneity are farm size, consumer food prices, use of growth hormones, animal health and safety, and use of genetically modified livestock, respectively. Similar to the apple survey data results, consumer food prices is ranked with low importance by mean B-W score, and the high variance indicates that a small segment of consumers consider it very unimportant, while other segments disagree. Again, low mean and high variance pertain to meat products that have large potential in a niche market. These results bring the authors to similar conclusions as stated above for apples. Additionally, three very contentious agricultural ethics

Table 6: Beef Production Attribute Correlation Coefficients

	SA	GH	GM	AS	PF	OP	FS	PL	CFP	FF
SA	1									
GH	0.19	1								
GM	-0.09	0.22	1							
AS	-0.09	-0.09	0.02	1						
PF	-0.02	-0.05	-0.1	0.06	1					
OP	-0.05	-0.13	-0.09	-0.05	0.06	1				
FS	-0.21	-0.37	-0.21	-0.22	-0.12	-0.15	1			
PL	-0.05	-0.21	-0.24	-0.2	-0.19	-0.07	0.03	1		
CFP	-0.14	-0.19	-0.26	-0.18	-0.15	-0.09	-0.16	-0.1	1	
FF	-0.21	-0.21	-0.19	-0.09	-0.18	-0.18	-0.05	-0.02	0.09	1

attributes exhibited high heterogeneity. The use of growth hormones, genetically modified livestock, and animal health and safety are issues that have been given a large amount of public attention by animal rights activist groups, organic farming advocates, and health officials alike. These three attributes hinge on several dimensions of consumer utility including food safety and human nutrition and health, while also considering intrinsic food values deeply connected to eco-responsibility movements and the emergence of socially alternative food markets. Umberger, McFadden, and Smith (2009) and Loureiro and Hine examine consumer valuation for hormone free and GM (genetically modified) free claims, respectively. These results support the aforementioned research findings that social dimensions of food values create distinct segments in the consumer market.

6 Discussion

Best-worst analysis was applied in this research to investigate the degree of importance consumers give to ten sustainable farming production attributes and in particular was used to determine behavioral differences across demographic subgroups of the population sample. The advantages of this methodology compared to more traditional stated preference analysis is evident in its higher discriminatory power for measuring trade-off decision making and in its wider applicability and interpretation outside of the survey context. While avoiding common rating bias, best-worst analysis results can be used in cross national and cross regional comparison studies on diverse populations and their judgment of similar attributes. This study gives credence to the strength of the best-worst method in yielding clear and simple interpretations. The simplicity of this analysis can be applied by marketing managers to gain insight into the evaluation behavior of different consumer sub-groups for targeted labeling.

The information gathered here from consumer data collected on perceptions about sustainable farming production practices holds large potential for marketing managers. The unique best-worst methodology provides additional insight into determinants of market behavior. In both beef and apple surveys, consumers indicated a strong perceptive correlation between sustainably produced labels and the size and locality of the farm of origin. The current available product differentiation schemes involve information and certifications related to variable production practices. This analysis suggests, similar to the findings of Onozaka and McFadden, that consumers perceive quality differences for locally grown and distributed products. Supporting studies, such as that of Bond, Thimany and Keeling-Bond (2008) give evidence that preferences for local food are significantly related to factors affecting farmer viability, sustaining local farm land, and contributing to smaller, local economies. Our work supports these findings that scale and geographic range factor heavily into consumer perceptions of sustainably labeled food products.

Increasing attention drawn towards smaller, local farms has been seen as a response to the widening awareness of the global food system business conduct. Distrust has been growing for imported foods, especially meat products with recent publicity on country of origin labeling requirements and other high profile food contamination cases. Sustainability claims on food targets many dimensions of consumer utility from quality and safety concerns to more intrinsic valuation connected to underlying food values such as fairness and the environment. In effect, some sustainability claims may be perceived by consumers as substitutable, while others are complimentary, a point emphasized by the work of Onozaka and McFadden. The value of a sustainable certification may only contribute marginally to the localness of a food product, while in other situations it may enhance the commitment to more well-rounded sustainable farming practices. It is important to consider how consumer willingness to pay varies for multiple combinations of sustainable farming production attributes. This body of work suggests that differentiating food claims on the level of local versus non-local and small scale farming versus corporate involvement may be a successful avenue of marketing. Sackett, Shupp, and Tensor plan to concentrate future research using this data set on willingness to pay measurements of sustainable labeling schemes in comparison to other niche markets such as local and organic.

7 Appendix

Table 7: Apple Sustainable Farming Attribute Key

I.D.	Attribute
GC	Ground Cover and Area Management Practices are Employed
FN	Fertilizer and Nutrient Materials are used minimally
CP	Little to No Chemical Pesticides are Used for Pest Management
CH	Little to No Chemical Herbicides are Used for Weed Management
PM	Pollinator Management is Employed
OP	Other Pests are controlled using preventative measures, and habitat controls
FS	Farm Size is small and Corporate Involvement is limited
PL	Production, distribution, and sale is done locally
CFP	Consumer food prices are affordable
FF	Farmers are financially stable

Table 8: Beef Sustainable Farming Attribute Key

I.D.	Attribute
SA	Prohibited use of sub-therapeutic antibiotics
GH	Prohibited use of growth hormones
GM	Prohibited use of genetically modified livestock
AS	Animal Health and Safety are Protected
PF	Feed is Pasture Based and Waste Management Systems Employed
OP	Pests are controlled using preventative measures, cultural and nutritional controls
FS	Farm Size is small and Corporate Involvement is limited
PL	Production, distribution, and sale is done locally
CFP	Consumer food prices are affordable
FF	Farmers are financially stable

8 References

Auger, P., Devinney, T., Louviere, J., 2007. Using best-worst scaling methodology to investigate consumer ethical beliefs across countries. Journal of Business Ethics 70, 299-326.

Bond, C.A., D. Thilmany, and J. Kelling-Bond, 2008. Understanding consumer interest in product and process based attributes for fresh produce. Agribusiness 24, 231-252.

Clonan, A., Holdsworth, M., Swift, J., and Wilson, P., 2010. UK consumers priorities for sustainable food purchases. http://purl.umn.edu/91948

Cohen, E., 2009. Applying best-worst scaling to wine marketing. International Journal of Wine Business Research 21, 8-23.

Darby, K., M. T. Batte, S. Ernst, and B. Roe, 2008. Decomposing Local: A conjoint analysis of locally produced foods. American Journal of Agricultural Economics 90, 476-486.

Detoni, D., and G.T. Tonsor, 2009. The direct and indirect effects of 'locally grown' on consumers attitudes towards agri-food products. Agricultural and Resource Economics Review 38, 384-396.

Howard, P.H. and P. Allen, 2010. Beyond organic and free trade? An analysis of ecolabel preferences in the United States. Rural Sociology 75, 244-269.

Finn, A., Louviere, J.J., 1992. Determining the appropriate response to evidence of public conern: the case of food safety. Journal of Public Policy and Marketing 11, 12-25.

Flynn, T.N., Louviere, J.J., et al. 2007. Best-worst scaling: What it can do for health care research and how to do it. Journal of Health Economics 26, 171-189.

Jaeger, S.R., A.S.Jorgensen, M.D. Aaslyng, W.L.P. Bredie, 2008. Best-worst scaling: An introduction and initial comparison with onadic rating for preference elicitation with food products. Food Quality and Preference 19, 579-588.

Louviere, J.J., 2001. What if consumer experiments impact variances as well as means: response variability as a behavioral phenomenon. Journal of Consumer Research 28, 506-522.

Lusk, J.L., and B. Norwood, 2005. Effect of experimental design on choice-based conjoint valuation estimates. American Journal of Agricultural Economics 87, 771-785.

Lusk, J.L. and Briggeman, B., 2009. Food values. American Journal of Agricultural Economics 91, 1-13.

Magidson, J. and Vermunt, J.K., 2001. Latent class factor and cluster models, bi-plots, and related graphical displays. Sociological Methodology 31, 223-264.

Marley, A.A.J., Louviere, J.J., 2005. Some probabilistic models of best, worst, and best-worst choices. Journal of Mathematical Psychology 49, 464-480.

McIntosh, E., Louviere, J.J., 2002. Separating weight and scale value: an exporation of best-attribute scaling in health economics, Health Economists' Study Group Meeting Notes, Brunel University.

Mueller, S., Francis, L., and Lockshin, L., 2010.Comparison of best-worst and hedonic scaling for the measurement of consumer wine preferences. Australian Journal of Grape and Wine Research.

Mueller, S., and Rungie, C., 2009. Is there more information in best-worst choice data? Using attitude heterogeneity structure to identify consumer segments. International Journal of Wine Business Research 21, 21-40.

Onozaka, Y. and D. Thilmany McFadden, 2010. Defining sustainable food market segments: do motivations and values vary by shopping locale? American Journal of Agricultural Economics; doi:10.1093/ajae/aaq152

Santimanon, T. and Weatherspoon, D., 2010. Hedonic Analysis of sustainable food products. International Food and Agribusiness Managment Review 13, 57-74.

Tonsor, G., and Shupp, R., 2009. Valuations of 'sustainably produced' labels on beef, tomato, and apple products. Agricultural and Resource Economics Review 38, 371-383.

Umberger, W. J., D. D. Thilmany McFadden, and A.R. Smith, 2009. Does altruism play a role in determining U.S. consumer preferences and willing to pay for natural and regionally produced beef? Agribusiness 25, 268-285.

Umberger, W.J., D.M. Feuz, C.R. Calkins, and K. Killinger-Mann, 2002. Consumer preference and willingness to pay for domestic corn-fed beef versus international grass-fed beed measured through an experimental auction. Agribusiness 18, 491-504.

Ward, C.E., J.L. Lusk, and J.M. Dutton, 2008. Implicit value of retail beef product attributes. Journal of Agricultural and Resource Economics 33, 364-381.