Estimating Desirable Cattle Traits Using Latent Class and Mixed Logit Models: A Choice Modeling Application to the U.S. Grass-Fed Beef Industry

Isaac Sitienei[*]
Jeffrey Gillespie[*]
Robert W. Harrison[*]
Guillermo Scaglia[†]

[*]Dept. of Agricultural Economics and Agribusiness
Louisiana State University Agricultural Center
Martin D. Woodin Hall
Baton Rouge, LA 70803

[†]Iberia Research Station
Louisiana State University Agricultural Center
Jeanerette, LA 70544

Emails: isitie1@lsu.edu; jmgille@lsu.edu; wharrison@agcenter.lsu.edu; gscaglia@agcenter.lsu.edu

Selected Paper prepared for presentation at the Southern Agricultural Economics Association's 2015 Annual Meeting, Atlanta, Georgia, January 31-February 3, 2015

Abstract

This study examines preferences for cattle traits using mixed logit and latent class models. Choice experiment data from a 2013 mail survey of 1,052 U.S. grass-fed beef producers were used. Preliminary results indicate that producers prefer lower-priced, heavy, black, and easy-to-handle feeders backgrounded from their own cows.

Introduction

Despite the relatively small market for grass-fed beef in the U.S., consumer demand has been slowly but consistently rising in recent years. The U.S. revenue from the sales of grass-fed beef was projected to increase from just under $5 million in 1998 to over $1 billion in 2010 (Food Market Institute, 2005). The uptick in consumer demand has motivated an increase in the production of grass-fed beef.

To increase productivity, technical, institutional, and genotypic factors affecting the grass-fed beef industry need to be considered. Breed improvement programs are key drivers for improving productivity in the beef industry. Every cattle producer faces the common question, "which animal should I choose?" The answer to this question remains at large for most categories of producers. The term breed can be used to describe animals with a common origin and selection history. Different cattle breeds possess different traits, and these traits can generally be classified as either production or market-related. Traits associated with the management aspects of cattle while still on the farm may be classified as production-related and they include: temperament, weight gain, coat color, gender, and source. Market-related traits, on the other hand, are those traits that describe the final product entering the market. For product sold as meat, these would include color, tenderness, juiciness, and flavor. For product sold as live animal, these would include coat color, weight, and body frame. Producers at different production segments value these traits differently. Genetic improvement programs should focus on broader concepts of breed selection and cattle traits rather than specializing on one or just a few.

A significant portion of the important factors that determine the value that farmers receive for their beef products can be controlled by the farmer and are generally tied to specific cattle breeds. Some of these factors include body frame, gender, weight, body condition, and temperament (Smith et al. 2000). Bretschneider (2005) found that castration of bulls at weaning or before attaining 551 pounds in body weight resulted in higher carcass quality, rapid growth, efficient food utilization, and less aggressive behavior. Other desirable traits can be incorporated into the herd through the influence of the farm operator. However, there are some extraneous factors such as extreme weather conditions, fluctuating beef prices, and animal diseases that are out of the farmer's control. A farmer can only adopt production practices that help mitigate the negative effects of these extraneous factors on the farm.

Most stated preference studies have focused on meat characteristics. Consumer willingness to pay for specific meat attributes such as tenderness, juiciness, and/or marbling has been widely investigated. However, little has been done to evaluate producer preference for cattle-related attributes such as daily weight gain, sex, temperament, and/or coat color. We are unaware of previous studies that have focused specifically on the preferences for cattle traits in the grass-fed beef segment. The objectives of this study are to: (1) assess the selection characteristics used by U.S. grass-fed beef producers and (2) determine grass-fed producer preferences for cattle traits. To achieve these objectives, a study was conducted in the fall of 2013 with a sample of U.S. grass-fed beef producers. Data collection, methods used, results, and discussion are provided in the following sections.

Data and Methods

Data used for this study were collected via a mail survey conducted during Fall, 2013, following the Tailored Design Method as recommended by Dillman et al. (2007). A total of 1,052 grass-fed beef producer names were collected via an extensive Internet search. Sources included www.eatwild.com, the American Grass-fed Association, Market Maker, and other individually advertised grass-fed beef farms. Using first-class mail, producers were initially mailed a survey package containing a cover letter, ten-page questionnaire, and postage-paid return envelope. A postcard reminder followed two weeks later. After a second round of mailing and another postcard reminder, 384 usable responses were obtained. The adjusted response rate was 41%, considering the 384 completed responses, 117 from producers no longer in the grass-fed beef business, and non-deliverable returns.

Respondents were presented with the nine choice scenarios with 18 hypothetical profiles similar to that represented in Figure 1 consisting of three alternatives "A", "B", and "Neither". The choice questions asked respondents to select one of the two animals they would retain/purchase for forage finishing as shown in Figure 1. The "Neither" alternative was an "opt-out" option for those producers who preferred neither "A" nor "B". The "opt-out" option is a characteristic that makes choice experiments desirable and more flexible for use in conjoint analysis.

Of seven attributes considered, five consisted of three levels each and the remaining had two levels each, as indicated in Table 1. These attributes were: (1) weight in pounds (lbs) at which the animal is introduced to the forage finishing phase (550, 650, and 750 lbs); (2) body frame, referring to the animal's skeletal size based on its hip height (small, medium, and large); (3) temperament, referring to how easy or difficult it is to handle the animal; (4) gender or sex of the animal (heifer, steer, or intact male); (5) the source of the feeder animal for grass-finishing (retained from own cows, purchased from an auction, or purchased via private treaty); (6) the animal's color, referring to the coat color of the animal which was generalized for ease of analysis to two levels, black or non-black; and (7) the price representing the value of the animal per hundredweight (cwt), indicating the price to purchase the animal or the market value of the retained animal for producers who background their animals. Based on prevailing market prices, three price levels were chosen ($120, $140, and $160). A full factorial design would yield 972 (3^5 x 2^2) profiles, which would be practically infeasible to work with, resulting in respondent fatigue. An orthogonal fractional factorial design having 18 profiles was therefore used, which reduced the number of choice sets while at the same time maintaining orthogonality—independence among the hypothetical products' levels (Harrison, Stringer, and Prinyawiwatkul, 2002; Hair et al., 2006).

Econometric Models

The choice modeling framework arose from the consumer theory developed by Lancaster (1966), which states that preference for goods is derived from characteristics embodied in those goods rather than the goods per se. Utility of a good can therefore be decomposed into separate utilities comprising its constituent characteristics or traits. Accordingly, for purposes of our study, grass-fed beef producers derive their utility for cattle from cattle attributes. A choice-based conjoint analysis experiment was conducted to assess the grass-fed beef producer preference for several attributes of beef cattle. Choice-based conjoint analysis has been widely used in marketing, health, transportation, and environmental research. It is a multi-attribute judgmental method that integrates conjoint analysis and discrete choice models in evaluating producer preference.

McFadden's random utility model can be specified for an individual grass-fed beef producer.

$$U_{nj} = V_{nj} + \epsilon_{nj}, \tag{1}$$

where U_{nj} is the *n*th producer's utility associated with choosing alternative *j*, V_{nj} is the non-stochastic portion determined by the cattle traits and their value levels, and ϵ_{nj} is the stochastic element. The probability that producer *n* chooses alternative *j* is given by

$$Prob\{V_{nj} + \epsilon_{nj} \geq V_{nk} + \epsilon_{nk}; for\ all\ k \in C_n\}, \tag{2}$$

where C_n is the choice-set of respondent *n* comprising alternatives A, B, and Neither (*Animal A*, *Animal B*, and *Neither* options in our choice set).

Assuming that individual *n* receives utility *u* from choosing alternative *j*, a utility function comprising two components, deterministic and stochastic, can be specified as shown below

$$U_{njt} = \beta_n x_{njt} + \epsilon_{njt}, \tag{3}$$

where x_{njt} is a vector of alternative-specific traits at observation *t*, coefficient β_n is unobserved for each *n* and varies with individuals within a population defined by the density function $f(\beta_n|\theta^*)$, and ϵ_{njt} is the random term. If the random term is independently and identically distributed, then the relatively easy-to-handle conditional logit model generally known as the multinomial logit model would be specified. The probability that individual *n* will choose alternative *j* is then represented by the following simple conditional logit model

$$P_{nj} = \frac{exp(\beta x_{nj})}{\sum_{l=1}^{M} exp(\beta x_{nl})}, \tag{4}$$

where β is the coefficient to be estimated and x_{nj} is a vector of alternative-specific traits.

Despite providing an easy-to-handle estimation process, the conditional logit model is limited by its assumption of independence of irrelevant alternatives (IIA). The model assumes that the introduction of a third alternative should not affect the probability of choosing the first or second alternative. Another important limitation with the conventional conditional logit model is the assumption that all respondents share the same β coefficient. More advanced and flexible models have been developed to circumvent this limitation. Random parameters logit and latent class models relax the (IIA) assumption and introduce heterogeneity of the taste parameters estimated.

Mixed Logit (MLM)

The mixed logit is a well-known model where the kernel is the logit formula for a given choice or repeated choices made by an agent (Revelt & Train, 1998; Train 2008). McFadden and Train (2000) showed the advantage of using the mixed logit to approximate a random utility model to any degree of accuracy with clear specification of variables and mixing distribution. It is a flexible logit model that allows parameters associated with the observed variable to vary across individuals where there is a known population distribution. The probability of respondent *n* choosing alternative *j* on occasion *t* is given by the following logit formula:

$$P_{ijt} = \frac{exp\left(\beta_n x_{njt}\right)}{\sum_{l=1}^{M} exp(\beta_n x_{nlt})} = P_{it}|\beta_i, \tag{5}$$

where β_n is a random parameter with unconditional density $f(\beta_n|\theta)$, θ is the distribution of β_n, and as specified earlier, x_{njt} is a vector of alternative-specific traits.

Latent Class Model (LCM)

The LCM theory suggests that an individual's choice behavior depends on observable and unobservable heterogeneity that vary with factors that cannot be observed by the analyst (Greene & Hensher, 2003). Unlike mixed logit models that cannot account for the sources of heterogeneity, LCM does a good job of specifying the source of such heterogeneity. The probability that individual *n* belonging to class *s* chooses alternative *j* in the *t*th choice situation is

$$P_{njt|s} = \prod_{t=1}^{T} \frac{\exp(\beta_s x_{njt})}{\sum_{l=1}^{M} \exp(\beta_s x_{nlt})} \quad s = 1, \ldots, S, \tag{6}$$

where β_s is the class-specific parameter used to capture heterogeneity in preference across classes, x_{njt} is a vector of alternative-specific traits for individual *n*, and *t* the number of choice occasions for individual *n*.

Variables Used in the Study

Table 1 shows the attributes and their respective levels used in the choice experiment. Most comparisons made in this study concern the signs and magnitude of the coefficients obtained for these attributes. The following are the membership variables that were used in the Consistent Akaike Information Criterion (CAIC) and the Bayesian Information Criterion **(**BIC) to decide the optimal number of classes for the LCM analysis: *Certified Organic* indicates whether or not the farmer operated a certified organic farm, *Cow-calf Producer* is a dummy variable indicating whether or not the respondent produced weaned calves, *Total Number of Cattle* indicates the total number of cattle raised on the farm, *Age* is the age of the producer, and *College* is a dummy variable indicating whether or not the farmer held a 4-yr college degree. Regional variables indicated where the farm was located in the U.S. Coefficients of the above listed membership variables were not the interest of the current study.

Results

Table 2 shows the summary statistics of the membership variables used in the LCM. Eighty percent of the respondents were cow-calf producers. Only 10% of the respondents were certified organic. The average total number of cattle raised was 127 with a standard deviation of 372. The average age of producers was 55 years and 70% of the respondents held a 4-year college degree. The majority of the respondents (32%) were located in the Midwest, 17% were in the Northwest, 14% were in the Southeast, and 9% were in the Southwest.

Comparison between CLM, MLM, and LCM

The simulated maximum likelihood estimates for the mixed model are reported in Table 3. Of interest is the standard deviation of each random parameter estimate. A highly significant standard deviation of a random parameter indicates heterogeneity in the population. The standard deviations for the weight 550 lbs and medium body frame coefficients were not significant, indicating a lack of heterogeneity in parameter estimates for these variables among the respondents. All coefficients for the remaining set of random parameters were highly significant, indicating that these coefficients were indeed heterogeneous among the respondents. Results from the MLM are summarized in Table 3. Grass-fed beef producers generally preferred 650-lb animals that were small-to-medium framed and easy to handle for finishing. Relative to intact males, steers and heifers were preferred. Animals that were retained from their own calves were preferred. Black, lower-priced feeder cattle were also preferred. The negative sign on the auction coefficient indicates disutility associated with that method of procuring animals for grass-finishing.

Table 4 summarizes the estimated results of CLM, MLM, and LCM. Individual-specific characteristics are not presented in the current study. All alternative-specific attributes considered in the current study are random with a normal distribution. Specification of a normal distribution allows *a priori* the assumption of a possibility for both positive and negative parameter estimates. A positive sign suggests that individual *n* would be willing to pay for an increase or presence of the associated attribute.

From results in Table 4, based on the log-likelihood values, the hypothesis that the CLM is favored to either the MLM or the LCM can be safely rejected. With the exception of weight and color, 550 lbs and non-black, respectively, all of the other coefficients of the CLM are significant. The only coefficient that is not significant in the MLM is the weight 550 lbs. Given that these two models are nested, a comparison based on the likelihood ratio test is meaningful (Pacifico, 2011). The signs of coefficients for both the MLM and CLM are consistent with our expectations and those obtained in the MLM reported in Table 3. However, the magnitude of the coefficients is significantly different for the MLM and CLM, a clear indication of the bias produced by the IIA assumption of the standard CLM (Bhat, 2003). To choose the optimal number of classes, we employed the CAIC and BIC proposed by Boxall and Adamowicz (2002). CAIC and BIC are both minimized at 4 classes as shown in Table 5.

Discussion and Conclusions

This paper employed the conditional logit, mixed logit, and latent class models to examine United States grass-fed beef producer preferences for cattle traits. Signs on coefficients obtained on all the models estimated provide insights into the utility related with each trait investigated. For instance, small-to-medium sized but 650-lb and easy-to-handle feeder cattle were found to be the most preferred traits by U.S. grass-fed beef producers. The results from the mixed logit model revealed significant preference heterogeneity among grass-fed beef producers for most cattle traits with the exception of 550-lb level in the weight trait. Sources of preference heterogeneity is not the primary objective/focus of the current study and will be investigated in future studies. Production systems and regions variables will be investigated as the main potential sources of heterogeneity.

References

Bhat, C. R. (2003). Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B: Methodological*, *37*(9), 837-855.

Boxall, P.C., and W.L. Adamowicz. (2002). Understanding Heterogeneous Preferences in Random Utility Models: a Latent Class Approach. *Environmental and Resource Economics* 23(4): 421-446.

Bretschneider, G. (2005). Effects of age and method of castration on performance and stress response of beef male cattle: A review. *Livestock Production Science*, *97*(2), 89-100.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2007). Internet, mail, and mixed-mode surveys: The Tailored Design Method. 3rd ed. John Wiley Sons, New York, NY.

Food Marketing Institute. (2005). SuperMarket Research. Food Marketing Institute. 7(1). Washington DC. **http://www.fmi.org/newsletters/super_research/index.cfm?year=2005**

Greene, W. H., & 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.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). Multivariate Data Analysis, 6th ed. Upper Saddle River, NJ: Pearson Education, Inc.

Harrison, R. W., Stringer, T., & Prinyawiwatkul, W. (2002). An analysis of consumer preferences for value-added seafood products derived from crawfish. *Agricultural and Resource Economics Review*, *31*(2), 157-170.

Lancaster, K. J. (1966). A new approach to consumer theory. *The journal of political economy*, 132-157.

McCluskey, J. J., Wahl, T. I., Li, Q., & Wandschneider, P. R. (2005). US Grass-Fed Beef: Marketing Health Benefits. *Journal of Food Distribution Research*, *36*(3).

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of applied Econometrics*, *15*(5), 447-470.

Pacifico, D. (2011). *Estimating nonparametric mixed logit models via EM algorithm* (No. 0663).

Revelt, D., & Train, K. (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of economics and statistics*, *80*(4), 647-657.

Smith, S. C., Gill, D. R., & Bess, C. (2000). *Effect of selected characteristics on the sale price of feeder cattle in Eastern Oklahoma*. Oklahoma Cooperative Extension Service, Division of Agricultural and Natural Resources, Oklahoma State University

Tables

Table 1. Cattle Traits and Levels Used in the Choice Experiment

Attribute	Level Codes
Weight	550, 650, or 750 lbs
Body Frame	Small, Medium, or Large
Temperament	Easy or Difficult
Gender	Heifer, steer, or Intact Male
Source	Retained, Private Treaty, or Auction
Color	Black or Non-black
Price	$ 120/cwt, $ 140/cwt, or $ 160/cwt

Table 2. Membership Variables Used in the Latent Class Model.

Independent Variables	Unit Description	Mean	Std Dev
Certified organic	= 1 if certified organic, 0 otherwise	0.10	0.30
Cow-calf	= 1 if cow-calf producer, 0 otherwise	0.80	0.40
Total number of cattle	Total number of grass-fed beef animals	126.78	371.69
Age	Age of the producer	54.66	13.73
College degree	= 1 if held a 4-year college degree	0.70	0.46
Northeast	= 1 if farm was located in the Northeast	0.21	0.41
Southeast	= 1 if farm was located in the Southeast	0.14	0.34
Northwest	= 1 if farm was located in the Northwest	0.17	0.38
Southwest	= 1 if farm was located in the Southwest	0.09	0.28
Midwest	= 1 if farm was located in the Midwest	32.29	0.47

Table 3. Simulated Maximum Likelihood Estimates from the Mixed logit Model.

Cattle Traits	Mean Coefficient	Standard Deviation
Weight 550 lbs	-0.2326 (0.1618)	0.1804 (0.1695)
Weight 650 lbs	0.4186** (0.2108)	0.8737*** (0.1920)
Small	1.1312*** (0.1946)	0.6304** (0.2495)
Medium	0.7931*** (0.1394)	0.3160 (0.2492)
Easy	3.4514*** (0.1952)	1.2282*** (0.1442)
Heifer	1.3075*** (0.2196)	0.5814* (0.3364)
Steer	1.3534*** (0.1668)	-1.1040*** (0.1745)
Retained	0.7750*** (0.1568)	1.0762*** (0.1587)
Auction	-1.1652*** (0.1721)	0.8789*** (0.1876)
Non-black	-0.2714** (0.1324)	0.8210*** (0.1499)
Price	-0.0279*** (0.0013)	-0.0144*** (0.0010)

Table 4. Comparing CLM, MLM, and LCM Parameter Estimates.

			LCM			
Cattle traits	**CLM**	**MLM**	**Class 1**	**Class 2**	**Class 3**	**Class 4**
Weight						
550 lbs	-0.0524	-0.2326	-0.2414	0.5030	0.3646	0.8592*
	(0.1431)	(0.1618)	(0.3832)	(0.7319)	(0.4175)	(0.4994)
650 lbs	0.3293**	0.4186**	0.3060	0.8589	0.0310	2.3918***
	(0.1867)	(0.2108)	(0.4438)	(0.8018)	(0.5917)	(0.8292)
Body frame						
Small	1.1837***	1.1312***	2.0153***	0.3936	1.0303**	0.8270
	(0.1715)	(0.1946)	(0.4629)	(0.8979)	(0.5008)	(0.6119)
Medium	0.9961***	0.7931***	1.0967***	0.2887	0.4886	0.7155*
	(0.1328)	(0.1394)	(0.2873)	(0.6258)	(0.4345)	(0.4041)
Temperament						
Easy	3.4532***	3.4514***	3.4062***	1.6822***	1.7539***	1.8514***
	(0.1413)	(0.1952)	(0.3964)	(0.6525)	(0.5744)	(0.4462)
Sex						
Heifer	1.1722***	1.3075***	-0.0216	1.0127	0.4021	0.5596
	(0.1841)	(0.2196)	(0.4485)	(0.6660)	(0.6729)	(0.7817)
Steer	1.1598***	1.3534***	0.7445**	2.2210***	0.5102	1.2799***
	(0.1297)	(0.1668)	(0.3020)	(5834)	(0.4659)	(0.3824)
Source						
Retained	1.0101***	0.7750***	0.6483**	0.5111	0.0335	0.8392***
	(0.1299)	(0.1568)	(0.2677)	(0.3602)	(0.4337)	(0.3174)
Auction	-1.0202***	-1.1652***	-1.4739***	-0.4464	-0.4145	-2.4674***
	(0.1548)	(0.1821)	(0.3425)	(0.6078)	(0.5597)	(0.7818)
Color						
Non-black	-0.0216	-0.2714**	-0.0766	-1.0292**	-0.6709	1.1901***
	(0.1130)	(0.1324)	(0.3007)	(0.5116)	(0.5401)	(0.4048)
Price	-0.0307***	-0.0279***	-0.0174***	0.0027	-0.0269***	-0.0347***
	(0.0009)	(0.0013)	(0.0024)	(0.0040)	(0.0039)	(0.0029)
LR Test	3267.95***	724.07***	-	-	-	-
Log likelihood	-4846.3083	-2347.5902	-2200.743	-2200.743	-2200.743	-2200.743

Table 5. Latent Class Logit Specification Using CAIC and BIC

Classes	**Log Likelihood**	**CAIC**	**BIC**
2	-2354.90	4924.86	4893.86
3	-2273.87	4901.56	4850.56
4	**-2200.74**	**4894.05**	**4823.05**
5	-2169.41	4970.13	4879.13
6	-2132.21	5034.49	4923.49
7	-2125.53	5159.88	5028.88
8	-2080.39	5208.35	5057.35
9	-2062.10	5310.52	5139.52
10	-2052.44	5429.95	5238.95

Figure 1. Sample of a Choice Experiment Question

Choice 1

Attributes	Animal A	Animal B
Weight	550 lbs	650 lbs
Body frame	Small	Small
Temperament	Easy	Difficult
Gender	Heifer	Heifer
Source	Retained	Auction
Color	Non-black	Non-black
Price	$120/cwt	$160/cwt

❖ **Which animal would you retain/purchase for forage finishing if these were the only feeders available?**

□ Animal A

□ Animal B

□ Neither