CONSUMER DEMAND FOR FOOD DIVERSITY

Jonq-Ying Lee and Mark G. Brown

Abstract

In this paper, consumer demand for food diversity is measured by the entropy and Simpson indices for budget shares. Results show that consumer demand for food diversity is related to total food expenditures and household size and composition.

Key words: food diversity, entropy, demand, Simpson index.

Consumer demand for variety in consumption has received increasing attention in recent years (Theil and Finke; Jackson; Shonkwiler et al.; Lee). It has been noted that as incomes increase, consumers tend to increase the number of goods consumed. Theil and Finke used the entropy and the Hirschman-Herfindahl indices of budget shares to study consumer demand for diversity. Jackson proposed a hierarchic purchasing model and used a flexible functional form to study variety in consumption. Both the Theil and Finke and the Jackson studies concentrated on broad commodity aggregates. Their results suggest that increases in expenditures on a single broad consumption category are accompanied by increases in the number of individual goods consumed in that category.

Shonkwiler et al. investigated how the number of individual foods consumed by a household is affected by the household's preferences and food expenditure. A theoretical model suggested by household production theory (Becker; Lancaster; Deaton and Muellbauer) was adapted to empirical data from a cross-section of households. The results indicated that the number of individual foods consumed responded strongly to household expenditure on all foods. A similar result was found by Lee using econometric models for count data.

Understanding variety in food consumption can be important in several areas. First, food variety can be important for nutrition. Nutrient levels vary between foods, and understanding factors affecting demand for food variety can be important for understanding nutrient intake levels. Factors such as the food stamp program may be particularly important for nutrition policy. Studying variety in food consumption may also reveal consumption patterns useful for marketing. In general, the features of modeling variety in consumption may be important for accurately specifying consumer demand equations.

The major purpose of this study is to examine the household demand for a varied diet using the entropy and Herfindahl or Simpson indices and to investigate the relationship between household composition and demand for a varied diet. The impacts of the household food budget and the food stamp program are also analyzed. The present study expands on the work by Theil and Finke who analyzed diversity in consumption across countries focusing on broad commodity groups (e.g., all foods were lumped together in one group) and the impact of income. The following sections review the relevant economic theory about the demand for variety and the commonly used measurements of diversity. The demand relationship between a varied diet and household characteristics is then estimated.

THEORETICAL BACKGROUND

The analysis in this study is based on traditional consumer demand theory. Maximization of utility subject to a budget constraint and nonnegativity conditions yields demand equations

(1) $q_i = g_i(p, m, z), \; i = 1, \ldots, n,$

Jonq-Ying Lee and Mark G. Brown are Research Economists, Florida Department of Citrus, and Adjunct Professor and Adjunct Associate Professor, respectively, Department of Food and Resource Economics, University of Florida.

Florida Agricultural Experiment Station Journal Series No. R-00163.

where q_i is quantity demanded of commodity i, p is a vector of prices, m is total expenditure, and z is a vector of variables indicating consumer preferences. Based on the nonnegativity conditions, quantity demanded is equal to or greater than zero ($q_i \geq 0$), allowing for corner solutions.

Assuming weak separability, subgroup demand equations for foods can be written as:

$$(2)\quad q_{Fi} = g_{Fi}(p_F, m_F, z),$$

where the F subscript indicates the food subgroup (i.e., p_F is a vector of food prices, and m_F is total food expenditure).

A simple measure of diversity in food consumption is the number of food items where $q_{Fi}>0$. Based on specification (2), this number is a function of food prices p_F, total food expenditures m_F, and preferences variables z.

Alternatively, diversity can be measured based on demand expenditure shares defined by

$$(3)\quad w_{Fi} = \frac{p_{Fi}q_{Fi}}{m_F} = h_{Fi}(p_F, m_F, z),$$

where w_{Fi} is the fraction of food expenditure used in purchasing food i and depends on p_F, m_F, and z. Using equation (3), diversity in food consumption can be specified as a function of food expenditure shares,

$$(4)\quad d = d(h_F(p_F, m_F, z)) = f(p_F, m_F, z),$$

where d is a measure of diversity, and h_F is a vector of food shares. In reduced form, diversity d is a function of food prices, total food expenditures, and preferences.

Diversity could also be measured using quantity shares ($q_i/\Sigma q_j$) instead of expenditure shares. In this case, a common measure (e.g., pounds) for aggregating across foods would be required. In the present study, quantities were not available and measuring diversity using quantity shares was not analyzed.

MEASUREMENTS OF DIVERSITY

Research on variety in food consumption has focused on the number of food items actually consumed (Jackson; Shonkwiler et al.; Lee). By this criterion, if a person consumes some of each food, then maximum food variety is being elected. However, such an approach does not provide any information about the distribution of individual food quantities consumed or, alternatively, the distribution of expenditure shares for the food items. The purpose of the present paper is to examine measures of diversity that do account for the latter type of distributions in the context of food consumption.

Diversity has been an important concept in ecological theory and application, appearing in the biological, physical, social, and management sciences under various names. The diversity concepts used in economics involve measurements of industrial concentration (e.g., Horowitz and Horowitz; Hexter and Snow) and income distribution. Economic theory suggests that the vigor of competition is related directly to the number of firms in the relevant industry, other things being equal. However, to the extent that monopoly power is correlated positively with both fewness of sellers and inequality in their sizes, it makes a difference whether, in an industry with 100 firms, each firm controls 1 percent of the industry's output or four firms control 80 percent while the remaining 96 account for only 20 percent. A similar argument can be used in measuring diversity in food consumption. The food item counts used in the Jackson, Shonkwiler et al., and Lee studies do not account for the magnitudes of individual food expenditure shares.

An analytical tool commonly used to study the concentration of an industry is the Herfindahl index, which is given by the formula

$$(5)\quad H = \Sigma w_i^2,$$

where, here, w_i is the market share of the i^{th} firm. When the industry is occupied by only one firm (pure monopolist), the index attains its maximum value of 1.0. The value declines with increases in the number of firms and increases with rising inequality among any given number of firms. The range of the Herfindahl index is [1/n, 1], where n is the number of firms in the industry of interest, with the upper limit being attained when one firm has total control of the market as mentioned above, and the lower limit when n firms have an equal share. The Herfindahl index thus measures diversity inversely. Alternatively, the Simpson index (Patil and Taillie; Bhargava and Uppuluri; Good) is defined as 1 minus the Herfindahl index, or

$$(6)\quad \text{Simpson index} = 1 - \Sigma w_i^2.$$

The range of the Simpson index is [0, 1-(1/n)] and measures diversity directly.

Another commonly used measure of industrial concentration is the entropy of market shares (Theil; Hart; Hall and Tideman). This measurement is also called Shannon's index (Patil and Taillie) or Gibb's index (Good). The entropy index (E) is defined as

$$(7)\quad E = -\Sigma w_i \log(w_i);$$

the entropy or the Shannon index varies from zero (when one market share equals 1 and hence the n-1 others vanish) to a maximum of log n (when all shares equal 1/n). As discussed by Theil, in information theory, entropy is the expected information content of a message.

DATA, VARIABLES, AND SPECIFICATION

As an example, diversity in food consumption was analyzed based on data from the 1981 consumer expenditure survey conducted by the Bureau of Labor Statistics (BLS). The survey involved roughly 10,000 households and was by design representative of the urban United States. The present example focuses on households in the northeastern region of the United States which completed two weekly diaries. The number of households in this group was 1,061.

Variables analyzed are food expenditure, household size, and household composition. Food expenditure is the money value of food purchased by the household during the two survey weeks. Foods purchased and eaten away from home are included in food expenditure as a separate category. The dependent variables are the entropy and Simpson indices calculated from the expenditure shares of 19 different food groups (see Appendix I for detailed food expenditure shares).

The independent variables chosen for the study include the number of household members in 12 age-sex categories, household-size-squared, total food expenditure in dollars during the two survey weeks, and the annual value of food stamps. Descriptive statistics for these variables are given in Table 1.

The independent variables listed above do not include price variables because of the cross-section nature of the data and absence of price information (over a two-week period, prices probably did not vary much; restricting the analysis to the northeastern region also reduces price variation associated with location). The BLS data provide information on food expenditure but do not indicate quantities purchased, so that implicit prices could not be calculated by dividing expenditures by quantities. If implicit prices could be calculated from more complete data, quality issues would also need to be addressed in estimating price responses (Deaton).

TABLE 1. DESCRIPTIVE STATISTICS FOR VARIABLES USED IN THE EMPIRICAL ANALYSIS

Variable	Mean	Standard Error
Entropy[a]	2.06	0.60
Herfindahl	0.78	0.18
Household Composition:		
M≤3[b]	0.11	0.38
F≤3[b]	0.10	0.35
4≤M≤10	0.16	0.45
4≤F≤10	0.15	0.45
11≤M≤15	0.15	0.45
11≤F≤15	0.16	0.45
16≤M≤25	0.23	0.47
16≤F≤25	0.27	0.48
26≤M≤50	0.36	0.49
26≤F≤50	0.40	0.50
M>50	0.24	0.44
F>50	0.32	0.48
Household Size	2.66	1.60
Food Expenditures	$112.12	80.51
Food Stamp Value	$ 64.62	281.59

[a]The maximum of the entropy index for this study is ln 19 = 2.944.
[b]Represents the number of males (M) or females (F) present in the household less than or equal to three years old. Same notations were used for other age-sex groups.

A six-digit code system is used by the BLS to identify different food items in its public user tape. Each food item can be reclassified into a broader category by using the truncated food code. For example, if the six-digit codes are used, cakes, cookies, and bread are different food items. When the first three digits of the food codes are used, these three food items will have the same food code and belong to the bakery product category. For illustrative purposes, in this study, food items are defined by the first three digits of the six-digit food code used by the BLS. The six-digit code might be useful for other studies.

For the present study, relationship (4) can be written as

$$(8)\quad d_t = f(hc_t, m_{Ft}),$$

where d_t is either the entropy or the Simpson index[1] for household t, and hc_t is the set of

[1]Theil and Finke used Working's model for expenditure shares to analyze the entropy and Herfindahl indices; expenditure share equations were estimated separately as functions of income. In the present study, a reduced formulation is used to analyze the latter two indices; values for the indices were calculated based on observable expenditure shares, and then regressed on a set of explanatory variables.

TABLE 2. PARAMETER ESTIMATES FOR EQUATION (9) WITH ENTROPY AND SIMPSON INDICES AS DEPENDENT VARIABLES

Variable	Entropy		Simpson	
	Estimate	S. E.[a]	Estimate	S. E.
Intercept	0.67187	0.08214	0.48911	0.02629
Household Composition:				
M≤3	0.24607	0.05267	0.07804	0.01686
F≤3	0.15071	0.06071	0.04612	0.01943
4≤M≤10	0.16798	0.05432	0.05718	0.01739
4≤F≤10	0.15429	0.05276	0.05224	0.01689
11≤M≤15	0.08724	0.05620	0.03269	0.01799
11≤F≤15	0.08912	0.05738	0.03091	0.01837
16≤M≤25	0.07015	0.05063	0.03317	0.01620
16≤F≤25	0.11400	0.04889	0.04780	0.01565
26≤M≤50	0.05102	0.05340	0.02464	0.01709
26≤F≤50	0.29669	0.05068	0.10009	0.01622
M>50	0.09736	0.05228	0.03639	0.01673
F>50	0.48391	0.05390	0.15879	0.01725
(Household Size)^2	-0.01246	0.00486	-0.00424	0.00156
Log(Food Exp.)	0.23125	0.01972	0.03626	0.00631
Food Stamps	0.00018	0.00006	0.00005	0.00002
Adjusted R^2	0.2852		0.1853	

[a]Represents the coefficient standard error estimate.

household characteristics shown in Table 1.

Theory has, at present, little to say on the rate at which commodities enter or leave the set of foods which are purchased. In the present study, a linear specification with logarithmic food expenditures was found to best fit the data, that is,

$$(9)\quad d_t = c + \sum_j \alpha_j f_{jt} + \beta HZ_t^2 + \varphi \log(m_{Ft}) + \gamma FSP_t + e_t ,$$

where f_{jt} is the number of persons of the j[th] type in household t; $HZ_t = \Sigma_j f_{jt}$ is the household size; FSP_t is the annual dollar value of food stamps; e_t is the disturbance term; and c, α, β, φ, and γ are parameters to be estimated. Household size squared was included in specification (9) to allow for economy of scale type impacts. The inclusion of the annual value of food stamps in the analysis provides information about the impacts of both participation and level of participation in the food stamp program. Equation (9) was estimated with the ordinary least squares method using the entropy and the Simpson index as dependent variables.

RESULTS

Entropy and Simpson index estimates for specification (9) are presented in Table 2. Based on the t-ratios, all parameter estimates are significantly different from zero at the α=.10 level of significance, except the estimate for male household members between the ages of 26 and 50 for the entropy specification.

All parameter estimates for age-sex composition variables are positive. The effect of an additional male household member decreases with age, reaching a trough for men between the ages of 26 and 50, and then increases slightly for men more than 50 years old. The effect of an additional female household member has a different pattern with a trough for members between the ages of 11 and 15. Moreover, an additional female has a smaller impact during her earlier years and a greater impact during her later years on the demand for a varied diet than an additional male of comparable age. The estimates indicate females in the homemaker age range are important for diversity in food consumption. The finding may indicate females in this age group have a greater interest in preparation of various meals and, perhaps, nutrition. The latter possibility may be particularly important to the extent that diversity in food consumption is related to a household's nutritional requirements and well-being.

The parameter estimates for the household-size-squared variable suggest economies of scale associated with diversity in food consumption. The negative coefficients for household-size-squared indicate that the positive impact of additional household members is reduced with increases in household size. For example, the positive impact of the arrival of a male infant to a two-person household would have a stronger impact on the diversity of food expenditures than the arrival of the same infant to a four-person household. Marginal impacts can be straightforwardly obtained by differentiating equation (9) and using the results in Table 2.

The estimates for the expenditure variables indicate increases in both total food expenditure and food stamp income increase diversity in food consumption. Given the semilogarithmic specification, the marginal impact of total food expenditure decreases at higher levels of total food expenditure. At the sample means, the marginal impacts of total food expenditure were .0021 and .0003 for the entropy and Simpson indices, respectively. Provided food stamp income is used during the two-week period in question, an increase in food stamp income can be expected to increase total food expenditure, so that the marginal impact of food stamp income would involve both the coefficient estimates for the logarithm of total food expenditure and food stamp income. The marginal impact of food stamp income cannot be determined precisely as the data were not complete with respect to food stamp income used during the two-week survey period. That is, to determine the marginal impact of food stamp income on diversity in food consumption, the marginal impact of food stamp income on total food expenditure must first be determined.

In comparison with the age-sex composition variables, the marginal impacts of total food expenditure are much lower and indicate that household composition is relatively more important for diversity in food consumption at the margin. For example, in the entropy model, the increase in food expenditure needed to increase diversity by .25, the marginal impact of a female between the ages of 26 and 50 in a two-person household, is about $119 per two weeks using the .0021 marginal impact for total food expenditure as an approximation.

The patterns of household effects shown in Table 2 are different from the ones found by Lee. Lee found that the effect of an additional household member, regardless of sex, initially increases with age, reaching a peak for members between the ages of 16 and 25, and then decreases slowly. To investigate these differences, the same data were used to estimate a Poisson regression model and a compound Poisson model with the weighted least squares method (see equations [9] and [10] in Lee's study). The coefficient estimates for the age-sex and expenditure variables for the Poisson models follow patterns similar to those for the entropy and Simpson models in Table 2. This finding suggests that the particular results in the present paper are dependent on the type of data (expenditure versus quantity) used.[2] The particular functional forms used—entropy, Simpson, or Poisson specifications—are not imposing undue restrictions. However, the entropy and Simpson indices might be preferred based on the theoretical properties of the indices in comparison to simply counting food items consumed.

CONCLUDING REMARKS

In this study, the consumer's demand for a variety of foods is measured by the entropy and Simpson indices for budget shares. The results of this study show that the demand for a diverse diet is positively related to the number of household members in different age-sex groups. The change in the demand for a diverse diet increases at a decreasing rate as household size increases. In addition, the results of this study suggest that increases in household food expenditure and participation in the food stamp program increase the demand for a diverse diet. These results are generally consistent with those found in studies by Theil and Finke; Jackson; Shonkwiler et al.; and Lee.

Diversity in food consumption is important to the extent that food variety is related to nutrition. In general, nutrient levels vary between foods, and understanding factors affecting demand for food variety can be important for understanding nutrient intake levels. The results of the present study indicate that household composition, particularly with respect to the number of older females, is relatively more important than the food expenditure variables for diversity in food consumption.

[2] Food consumption data were used in Lee's study, while expenditure data were used in this study.

REFERENCES

Becker, G. "A Theory of the Allocation of Time." *Econ. J.*, 75(1965): 493–517.

Bhargava, T. N., and V.R.R. Uppuluri. "On an Axiomatic Derivation of Gini Diversity with Applications." *Metron*, 33(1975):1–13.

Bureau of Labor Statistics. "1981 Urban Diary Survey Public Use Tape."

Deaton, A. "Quality, Quantity and Spatial Variation of Price." *Amer. Econ. Rev.*, 78(1988): 418–30.

Deaton, A., and J. Muellbauer. *Economics and Consumer Behavior.* Cambridge: Cambridge University Press, 1980.

Good, I. J. "Diversity as a Concept and Its Measurement: Comment." *J. Amer. Stat. Assoc.*, 77(1982): 561–63.

Hall, M., and N. Tideman. "Measures of Concentration." *J. Amer. Stat. Assoc.*, 62(1967):162–68.

Hart, P. E. "Entropy and Other Measures of Concentration." *J. Royal Stat. Society*, ser. A, 134(1971):73–85.

Hexter, J. L., and J. W. Snow. "An Entropy Measure of Relative Aggregate Concentration." *South. Econ. J.*, 36(1970):239–43.

Horowitz, A., and I. Horowitz. "Entropy Markov Processes and Competition in the Brewing Industry." *J. Ind. Econ.*, 16(1968):196–211.

Jackson, L. F. "Hierarchic Demand and the Engel Curve for Variety." *Rev. Econ. and Stat.* 66(1984):8–15.

Lancaster, K. J. "A New Approach to Consumer Theory." *J. Polit. Econ.*, 74(1966): 132–57.

Lee, J. "The Demand for a Varied Diet with Econometric Models for Count Data." *Amer. J. Agr. Econ.*, 69(1987): 687–92.

Patil, G. P., and C. Taillie. "Diversity as a Concept and Its Measurement." *J. Amer. Stat. Assoc.*, 77(1982): 548–61.

Shonkwiler, J. S., J. Lee, and T. G. Taylor. "An Empirical Model of the Demand for a Varied Diet." *Applied Econ.*, 19(1987): 1403–10.

Theil, H. *Statistical Decomposition Analysis.* Amsterdam: North-Holland, 1972.

Theil, H., and R. Finke. "The Consumer's Demand for Diversity." *Eur. Econ. Rev.*, 23(1983): 348–59.

APPENDIX I. DESCRIPTIVE STATISTICS FOR FOOD EXPENDITURE SHARES USED IN THE EMPIRICAL ANALYSIS

Food Group	Mean	Standard Deviation	Food Group	Mean	Standard Deviation
Cereal and Cereal Products	2.82	3.01	Fresh Fruits	3.65	5.20
Bakery Products	7.46	6.51	Fresh Vegetables	3.29	3.40
Beef	8.26	9.11	Processed Fruits	2.89	4.31
Pork	4.12	5.65	Processed Vegetables	1.73	2.27
Other Meats	3.75	4.88	Sugar and Other Sweets	2.55	3.83
Poultry	3.47	4.90	Nonalcoholic Beverages	6.44	6.80
Fish and Seafood	2.29	4.27	Fats and Oils	1.78	2.38
Eggs	1.59	2.27	Miscellaneous Foods[a]	27.51	24.15
Fresh Milk and Cream	6.17	8.09	Food Away From Home	5.38	5.90
Other Dairy Products	4.85	5.59			

SOURCE: Bureau of Labor Statistics.

[a]Includes soup, frozen meals, frozen and prepared food items, potato chips and snacks, nuts, condiments, and baby food.