

Article

Application of Resolution Regression and Resolution Graphs in Evaluating Probability Forecasts Generated Using Binary Choice Models

Senarath Dharmasena *, David A. Bessler and Oral Capps, Jr.

Department of Agricultural Economics, Texas A&M University, College Station, TX 7743-2124, USA; d-bessler@tamu.edu (D.A.B.); oral.capps@ag.tamu.edu (O.C.J.)

* Correspondence: sdharmasena@tamu.edu; Tel.: +1-979-862-2894

Abstract

Binary choice models are widely used in econometric modeling when the dependent variable corresponds to discrete outcomes. With appropriate decision rules, these models provide predictions of binary choices generated from predicted probabilities. The accuracy of these predictions in terms of classifying probabilities to events that occurred versus those that did not is a key issue. The use of expectation-prediction success at present is the standard method used to assess the accuracy of these predictions. However, this method is limited in its ability to correctly classify probabilities in the absence of appropriate predetermined cut-off levels. We propose alternative methods to classify probabilities generated through binary choice models, namely resolution graphs and resolution regressions that measure the ability to sort predicted probabilities against observed outcomes. Using probabilities generated from the use of logit models applied to purchasing decisions of various non-alcoholic beverages made by U.S. households, we compare probability sorting power using expectation-prediction success as well as resolution graphs and resolution regressions. Based on expectation-prediction success, the logit models performed better at classifying outcomes related to purchasing isotonic drinks, regular soft drinks, diet drinks, bottled water, and tea. Based on resolution regressions, the null hypothesis of perfect sorting of probabilities was rejected for all non-alcoholic beverages. Although the logit models generated upward-sloping resolution graphs as expected, they were relatively flat compared to the 45-degree perfect sorting line. Going forward, we recommend using resolution regression and resolution graphs to capture sorting of probabilities in addition to the conventional metrics used in ascertaining the ability of binary choice models to predict out-of-sample behavior.

Keywords: binary choice models; probability forecasts; forecast evaluation; resolution regression; resolution graphs



Academic Editor: Julien Chevallier

Received: 22 October 2025

Revised: 28 January 2026

Accepted: 9 February 2026

Published: 24 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Qualitative choice models are widely used in econometric modeling when the dependent variable corresponds to discrete outcomes (Train, 2003). Binary logit and probit models, where the dependent variable corresponds to binary (0,1) outcomes, are popular among a wide range of¹ qualitative choice models available. The use of the probit/logit² analysis, particularly of binary choices, is well established in the economics and econometrics literature (Maddala, 1983; McFadden, 1984; Pindyck & Rubinfeld, 1998). Once

estimated, with appropriate decision rules, binary choice models are designed to provide predictions (probabilities) of the binary choices given a set of explanatory factors.

A key question relates to the accuracy of these predicted probabilities in terms of classifying them to events that occurred versus those that did not. Accuracy can be measured using a metric such as expectation-prediction success, which at present is the standard method to evaluate the predictive performance of qualitative choice models (Stock & Watson, 2007). With this metric, the percentage of correct (incorrect) predictions is calculated in comparison to the total number of predictions based on a predetermined probability cut-off level. However, the use of expectation-prediction success is limited in the ability to correctly classify the probabilities in the absence of appropriate predetermined cut-off levels³.

We propose resolution regression (also known as covariance regression) and resolution graphs (also known as covariance graphs), metrics that have not been used in the extant literature to measure the accuracy of predicted probabilities from binary choice models. Resolution graphs and resolution regression measure the ability of binary choice models to sort predicted probabilities against observed outcomes. To the best of our knowledge, our work is the first to evaluate forecasts developed from binary choice models (logit models in this study) using resolution regression and resolution graphs.

In this light, the general objective of this study is to propose alternative methods to evaluate the forecast performance of binary choice models. The specific objectives of this study are twofold:

- (1) to evaluate the out-of-sample prediction success of logit models using resolution regression and resolution graphs, and
- (2) to compare the use of resolution regression and resolution graphs with conventional prediction-success metrics that rely on predetermined cut-off values. To achieve these objectives⁴, we estimate logit models associated with purchases/non-purchases of non-alcoholic beverages made by U.S. households. We subsequently retrieve out-of-sample predicted probabilities of purchase and non-purchase. In turn, these predicted probabilities are then evaluated based on expectation-prediction success as well as using resolution regression and resolution graphs.

Based on sensitivity and specificity, the logit models performed better at classifying outcomes related to purchasing isotonic drinks, regular soft drinks, diet drinks, bottled water, and tea. The reverse was true for classifying outcomes related to not purchasing high-fat milk, low-fat milk, fruit drinks, fruit juices, and tea. Based on resolution regressions, the null hypothesis of perfect sorting of probabilities was rejected for all the non-alcoholic beverages considered. Although the respective logit models generated upward-sloping resolution graphs as expected, they were relatively flat compared to the 45-degree perfect sorting line. Except for isotonic drinks, the logit models were better at sorting forecasted probabilities associated with purchasing as opposed to not purchasing beverages. Based on our findings, we recommend using resolution regression and resolution graphs to capture sorting of probabilities in addition to the conventional metric used (expectation-prediction success) in ascertaining the ability of binary choice models to predict out-of-sample behavior.

2. Materials and Methods

To evaluate the effectiveness of the logit models to correctly classify the decision to purchase non-alcoholic beverages, we initially generate two random samples of observations. Then we estimate the respective logit models using one sample (the training sample) and reserve the data from the second sample (the testing sample) to obtain out-of-sample forecasted probabilities. In our analysis, we divide the sample of 7639 household-level

observations associated with purchases or non-purchases of non-alcoholic beverages in half, generating two random samples of data based on the use of the SAS 9.2 Enterprise Minor data mining software (SAS Institute, Inc., 2022). The first sample, labeled as sample A (the training sample), contains 3820 household-level observations, and the second sample, labeled as sample B (the testing sample), contains 3819 household-level observations. Simply put, we use sample A to estimate the logit model dealing with the decision to buy or not to buy a particular non-alcoholic beverage. Next, using the estimated coefficients of the explanatory variables from the logit model based on sample A, we rely on data from sample B to generate out-of-sample forecast probabilities of purchase and non-purchase. That is, we use the estimated coefficients obtained from sample A and the observations associated with the explanatory variables from sample B to generate the out-of-sample forecast probabilities. Following Equation (1), we depict the estimated logit model using sample A observations.

$$Z = X'\beta \quad (1)$$

where the cumulative distribution function, F_L , associated with the logistic distribution of Z is shown in Equation (2) below.

$$P = F_L(Z_i) = \frac{e^{Z_i}}{(1 + e^{Z_i})} \quad (2)$$

where,

Z = vector of index values generated for each observation through regression of the binary choice variable on the vector of explanatory variables;

X = vector of explanatory variables in the logistic regression (price of each product and other socio-economic and demographic variables in this study);

β = vector of estimated regression coefficients associated with each explanatory variable;

P = associated probability generated through assuming the index variable, Z , has a logistic distribution;

F_L = cumulative distribution function.

Expectation-prediction success tables are two-by-two contingency tables concerning expected outcomes conditional on predicted probabilities. These tables provide the number of instances in which the purchase of the non-alcoholic beverage is correctly and incorrectly predicted, and the number of instances in which the non-purchase of the non-alcoholic beverage is correctly and incorrectly predicted. For classification purposes, the cut-off probability value often has been assumed to be 0.5 in the extant literature. Hence, if the predicted probability generated via the estimation of the logit model is greater than 0.5, that observation is predicted to be associated with the event that occurred (i.e., purchase of the non-alcoholic beverage, outcome index equals 1). On the other hand, if the predicted probability generated via the estimation of the logit model is less than 0.5, that observation is predicted to be associated with an event that did not occur (i.e., non-purchase of the non-alcoholic beverage, outcome index equals 0). The overall percentage of correct predictions can be used as a measure of out-of-sample forecast performance (ability to correctly classify decision outcomes).

We illustrate the measurement of expectation-prediction success in Table 1. The various entries in Table 1 are associated with a predetermined cut-off value in conjunction with the predicted probabilities generated from binary choice models. Entry *TN* (true negative) depicts the number of occurrences where the predicted and observed outcomes match, that is, where the outcome index equals 0 (not purchasing non-alcoholic beverages in our study). Similarly, entry *TP* (true positive) depicts the number of occurrences where the predicted and observed outcomes also match, that is, where the outcome index equals

1 (purchasing non-alcoholic beverages in our study). Entries *FN* (false negative) and *FP* (false positive) depict the number of occurrences of incorrect classifications. Entry *FN* deals with the number of times the predictions were made of not purchasing non-alcoholic beverages, but in fact purchases were made, and in like manner, entry *FP* deals with the number of times the predictions were made of purchasing non-alcoholic beverages, but in fact purchases were not made.

Table 1. A two-by-two contingency table of expected-prediction success.

		Actual Outcome	
Predicted Outcome	0	0 True Negative	1 False Negative
	1	False Positive	True Positive

The number of correct out-of-sample predictions is given by $(TN + TP)$, the sum of the diagonal elements. The percentage of correct out-of-sample predictions is given by $\frac{(TN+TP)}{(TN+FN+FP+TP)} * 100\%$. The fraction of observations that are correctly predicted for which the outcome index equals 1 is called “*sensitivity*” in the literature and is depicted as $\frac{TP}{(TP+FN)}$. The fraction of observations that are correctly predicted for which the outcome index equals 0 is called “*specificity*” in the literature and is denoted by $\frac{TN}{(TN+FP)}$.

Another measure of forecast performance relates to the sum of the fraction of ones correctly predicted and the fraction of zeros correctly predicted, a number which should exceed unity if the classification of outcomes is of any value (Kennedy, 2003). In other words, the sum of sensitivity and specificity must exceed one.

In agreement with Greene (2012, p. 658), “in general any prediction rule will make two types of errors; it will incorrectly classify zeros as ones and ones as zeros.” For binary choice models, to the best of our knowledge, no benchmarks are provided regarding suitable prediction success rates or percentages of correct classifications. Without question, the choice of the cut-off probability level is critical to the decision to classify outcomes.

The choice of the cut-off probability level of 0.5 seems appropriate for discrete binary events that have realized relative frequencies close to 0.5 (the event occurs roughly 50% of the time). But suppose that the discrete binary events do not have realized frequencies of 50% or close to 50%. One example of this situation is the classification of a bull versus a bear market, wherein a bull market occurs 80% of the time, and a bear market occurs 20% of the time historically. This realized relative frequency can also be identified as the “market penetration.” In our analysis of non-alcoholic beverages, for example, the market penetration for bottled water is 70% (or probability of 0.7). Hence, the cut-off probability value that can be used to classify outcomes associated with the decision to purchase bottled water is 0.70 in this situation.

As noted previously, resolution is a metric of goodness of sorting power. In our work, we explicitly consider the ability of the logit models to sort probabilities into two classes, namely “high” probabilities associated with the purchase of non-alcoholic beverages and “low” probabilities associated with not purchasing non-alcoholic beverages. Ideally, we would like to see “high” probabilities associated with the purchase of a given non-alcoholic beverage, and “low” probabilities associated with the non-purchase of a given non-alcoholic beverage. Furthermore, for a perfect sorting model, we would like the probability to be close to 1 associated with the purchase of non-alcoholic beverages, and the probability to be close to 0 associated with the non-purchase of non-alcoholic beverages.

According to Sanders (1963), if any forecaster issues probability 1 for all events that occur (the purchase of non-alcoholic beverages) and probability 0 for all those events that do not occur (the non-purchase of non-alcoholic beverages), his/her sorting power is

perfect (or he/she issues perfectly sorted probabilities). Dawid (1986) further stated that “it is unreasonable in general to expect perfect sorting, because perfect sorting is equivalent to absolutely correct or absolutely incorrect categorical forecasting.” Resolution, according to Yates (1982), is a metric of “goodness-of-sorting” power and can be measured using *resolution regression*. Put simply, with resolution regression, the predicted probabilities generated from the use of binary choice models are regressed as a function of the outcome index (0 and 1). Mathematically, the Resolution regression is shown below in Equation (3).

$$\text{Prob}(Y|X) = \alpha + D\beta + e \quad (3)$$

where,

Y = predicted purchase probability generated from the binary choice model (logit model in this study);

X = vector of predictor variables used in the binary choice model (logit model in this study);

α = intercept term;

D = observed outcome index (zero or one) associated with predicted probabilities of the binary choice model;

β = estimated parameter associated with the outcome index;

e = error term of the resolution regression. This error term is distributed with a non-linear Beta distribution.

Equation (3) is estimated using a maximum likelihood method, where the error term of the resolution regression (Equation (3)) is distributed Beta. The reason for not using the ordinary least squares (OLS) estimation method to estimate resolution regression is that, since the dependent variable is a bounded dependent variable where the bounds are 0 and 1, the estimates may not be efficient and consistent due to potential heteroscedasticity issues. Consequently, we use a “beta regression” to mitigate this problem and estimate the slope and intercept coefficients of the resolution regression. In this way, the slope and intercept coefficients are consistent and efficient under the assumption of beta-distributed errors in the resolution regression.

A joint test of the null hypothesis of the slope coefficient associated with the outcome index equal to one, and intercept equal to zero, can be carried out using a chi-square test. Note that the intercept coefficient (α) captures the calibration bias, and the slope coefficient (β) captures the sorting slope. The joint hypothesis is shown below.

$$\begin{aligned} H_0 : \beta &= 1, \text{ and } \alpha = 0 \\ H_1 : \beta &\neq 1, \text{ and } \alpha \neq 0 \end{aligned} \quad (4)$$

Failing to reject this null hypothesis provides statistical evidence of perfect resolution (also known as perfect sorting) of probabilities for events that occurred and for probabilities of events that did not occur.

In addition, analysts can examine this sorting power visually by a plot of predicted probabilities on the y -axis and the outcome index on the x -axis. The line of perfect sorting coincides with a 45-degree line. In this case, all predicted probabilities are 0 when the outcome index equals 0, and all predicted probabilities are 1 when the outcome index equals 1. Some degree of sorting (or resolution) power of binary choice models is evident only in the case of upward-sloping graphs.

Resolution graphs provide visual representations of predicted probabilities and associated outcome indices, where a line can be drawn between the mean predicted probability associated with the outcome index of zero and the mean predicted probability associated with the outcome index of one. Upward-sloping graphs are expected for accurate classification of generated probabilities from binary choice models. This simple graphical

representation of the classification of generated probabilities provides a visual aid that would go hand in hand with resolution regression, supported by statistical metrics. The statistical test of the hypothesis related to Equation (4) is characteristic only of resolution regression, compared to standard methods such as the use of expectation-prediction success tables, and other methods explained in footnote #3.

In our analysis, we initially construct resolution graphs and then regress forecasted probabilities as a function of the outcome index to determine their sorting power⁵. Any deviation of the slope coefficients from one and the intercept coefficients from zero jointly in these respective regressions would be characterized as poorly resolved probabilities. Additionally, out-of-sample prediction success tables are constructed based on the standard cut-off value commonly used in the literature (0.5) as well as cut-off values corresponding to the proportion of households that purchased ten non-alcoholic beverages (market penetration). With this information, we are positioned to compare and contrast the use of resolution regression and resolution graphs with the use of prediction-success metrics.

3. Results

The estimation of the respective logit models was done using a maximum likelihood procedure using SAS, version 9.4 (SAS Institute, Inc., 2022). The parameter estimates, standard errors, and *p*-values associated with the ten estimated logit models are provided in Appendix A. The set of explanatory variables considered was household socio-demographic factors, as well as the price of the respective non-alcoholic beverages. Nevertheless, the sole focus of our analysis is assessing the ability of the logit models to correctly classify the decision to purchase or not to purchase non-alcoholic beverages using conventional methods and two novel approaches, namely resolution graphs and resolution regressions. Consequently, we do not present a discussion dealing with the determination of the key socio-demographic factors or the impacts of price on affecting the probability of purchase of each non-alcoholic beverage. Suffice it to say that the set of explanatory variables considered was influential in affecting the likelihood or probability of purchasing or not purchasing each of the ten non-alcoholic beverages selected. The goodness-of-fit statistics, based on McFadden's R^2 , ranged from 0.0437 (tea) to 0.1415 (coffee). The magnitudes of the McFadden R^2 values are consistent with those reported in the economics literature.

As exhibited in Table 2, the sensitivity, specificity, and the sum of sensitivity and specificity measures were calculated based on the out-of-sample probabilities generated using the 0.5 and the market penetration cut-off values for each beverage. For example, the market penetration value for diet soft drinks was calculated to be 65%. If the 0.5 cut-off was used to classify outcomes out-of-sample associated with the purchase or non-purchase of diet soft drinks, the sensitivity was calculated to be 0.90, and the specificity was calculated to be 0.19. However, if the market penetration was used as the cut-off value to classify the out-of-sample outcomes associated with the purchase or non-purchase of diet soft drinks, the sensitivity was calculated to be 0.73, and the specificity was calculated to be 0.46. The higher the sum of sensitivity and specificity, the better the out-of-sample forecast performance of the logit model concerning correct classifications. The respective sums of sensitivity and specificity were 1.09 and 1.19 for diet soft drinks. Hence, the better classification of outcomes occurred when the market penetration cut-off value was used. According to Table 2, this finding was perpetuated for each of the ten respective non-alcoholic beverages. Simply stated, the use of the market penetration cut-off yielded higher expectation-prediction success compared to the use of the 0.5 cut-off.

Further, the correct out-of-sample classification for purchases of non-alcoholic beverages ranged from 51% (high-fat milk) to 73% (diet soft drinks). On the other hand, the correct out-of-sample classification for non-purchases of non-alcoholic beverages varied

from 46% (diet soft drinks) to 71% (high-fat milk). The logit models did a better job of classifying outcomes related to purchases of isotonic drinks, regular soft drinks, diet drinks, bottled water, and tea because their specificity measures exceeded their sensitivity measures. But the sensitivity measures exceeded the specificity measures for high-fat milk, low-fat milk, fruit drinks, fruit juices, and tea. In these instances, the logit models were better at classifying outcomes related to non-purchases of these beverage categories.

Table 2. Summary of expectation-prediction success using the 0.5 cut-off value and the market penetration cut-off value.

Beverage Category	0.5 Cut-Off Value			Market Penetration Cut-Off Value		
	Sensitivity	Specificity	Sum of Sensitivity and Specificity	Sensitivity	Specificity	Sum of Sensitivity and Specificity
Isotonics	0.06	0.98	1.04	0.63	0.62	1.25
Regular Soft Drinks	1.00	0.00	1.00	0.68	0.64	1.35
Diet Soft Drinks	0.90	0.19	1.09	0.73	0.46	1.19
High-Fat Milk	1.00	0.00	1.00	0.51	0.71	1.22
Low-Fat Milk	0.87	0.27	1.15	0.55	0.67	1.22
Fruit Drinks	0.99	0.03	1.02	0.58	0.67	1.25
Fruit Juices	1.00	0.00	1.00	0.60	0.66	1.26
Bottled Water	0.98	0.07	1.05	0.66	0.54	1.20
Coffee	0.93	0.27	1.20	0.71	0.60	1.31
Tea	0.98	0.06	1.04	0.58	0.60	1.18

Source: Calculations made by the authors. Note: Sensitivity is the fraction of observations that are correctly predicted associated with the purchase of non-alcoholic beverages. Specificity is the fraction of observations that are correctly predicted associated with the non-purchase of non-alcoholic beverages. The sum of the sensitivity and the specificity is greater than 1 in all instances using the market penetration cut-off value. The higher the sum of sensitivity and specificity, the better the out-of-sample forecast performance of the logit model.

In Figure 1, we provide resolution graphs associated with the out-of-sample probability forecasts for each non-alcoholic beverage. As shown in Figure 1 as well as in Table 3, we observed relatively high probabilities associated with the outcome index of one (indicative of purchasing non-alcoholic beverages) for regular soft drinks, diet soft drinks, high-fat milk, low-fat milk, fruit drinks, fruit juices, bottled water, coffee, and tea. Indeed, the mean probability values were 0.91, 0.67, 0.82, 0.63, 0.77, 0.93, 0.72, 0.85, and 0.73, respectively, for each of these beverages. However, for isotonics, we observed a mean probability value of 0.29 associated with the outcome index of 1.

In instances where purchases of a non-alcoholic beverage do not occur, we expect to observe relatively low mean probability values. Only in the case of isotonic drinks did we find this pattern. Consequently, the respective logit models underperformed in the sorting of out-of-sample forecast probabilities associated with the outcome index of zero (indicative of not purchasing non-alcoholic beverages), except for isotonic drinks. Additionally, we observed lower mean probabilities associated with the outcome index of zero for all non-alcoholic beverage categories compared to those associated with the outcome index of one. Bottom line, except for isotonic drinks, the logit models were better at sorting forecasted probabilities associated with purchasing as opposed to not purchasing non-alcoholic beverages.

The median values associated with the generated out-of-sample probabilities for each beverage are as follows: isotonic drinks 0.18, regular soft drinks 0.93, diet soft drinks 0.69, high-fat milk 0.82, low-fat milk 0.63, fruit drinks 0.78, fruit juices 0.95, bottled water 0.72, coffee 0.79, and tea 0.74. The 95% confidence intervals associated with each out-of-sample probability are as follows: isotonic drinks 0.0039, regular soft drinks 0.0019, diet soft drinks 0.0038, high-fat milk 0.0028, low-fat milk 0.0037, fruit drinks 0.0033, fruit juices 0.0018, bottled water 0.0033, coffee 0.0057, and tea 0.0030.

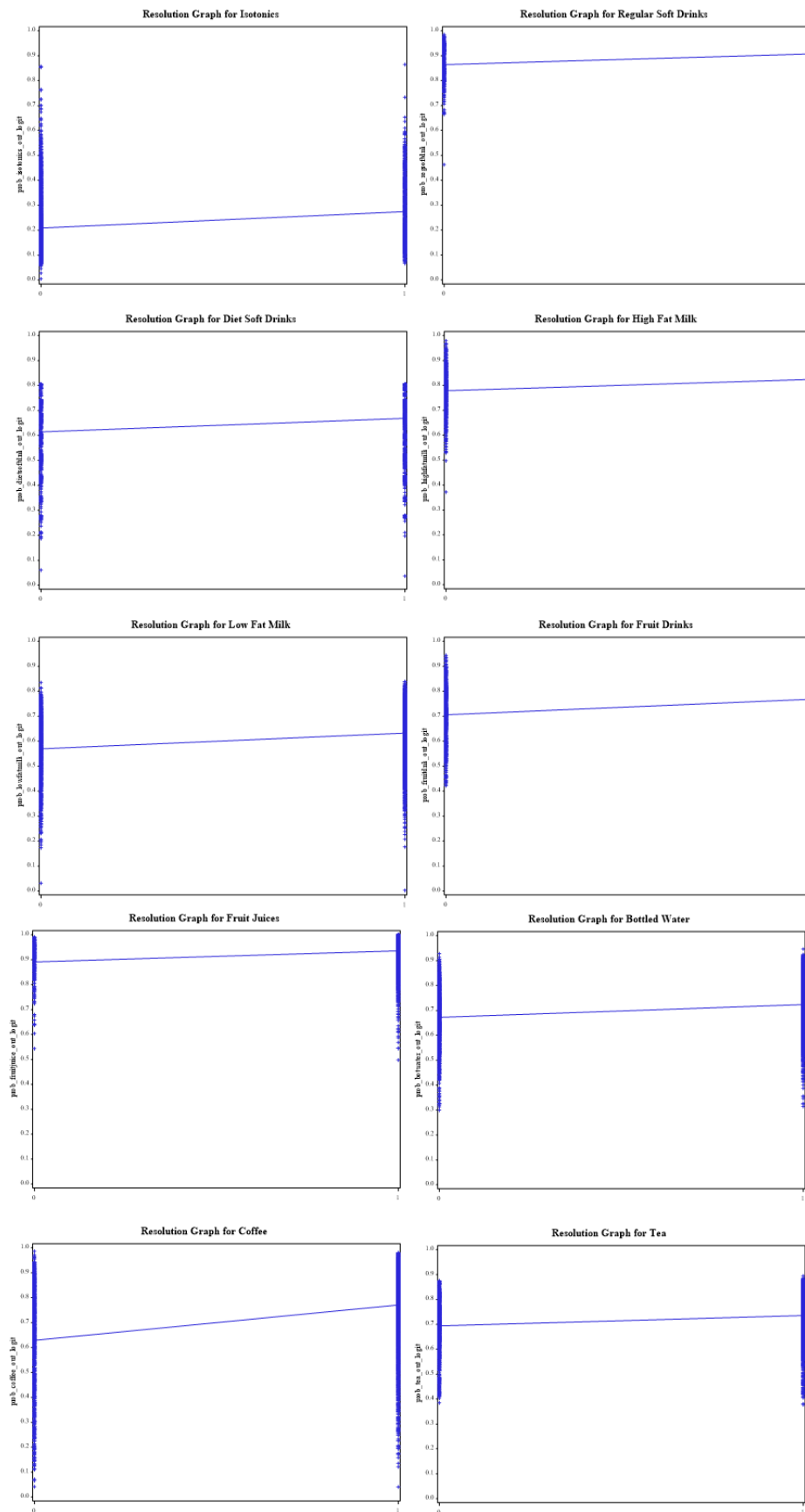


Figure 1. Resolution graphs of estimated probabilities generated from binary logit models association with various non-alcoholic beverages. Source: Construction by authors. Note: the binary outcome index, either 0 or 1, is on the x-axis. The estimated probability generated from the binary logit model is on the y-axis.

Table 3. Maximum Likelihood Estimates of the Parameters in the Resolution Regression (Beta Regression) and the Corresponding Mean Probabilities of Purchase and Non-Purchase.

Beverage Category	Intercept	Slope	Chi-Squared Statistics Associated with the Joint Test	Mean Probability for Purchases of Non-Alcoholic Beverages, Outcome Index = 1	Mean Probability for Non-Purchases of Non-Alcoholic Beverages, Outcome Index = 0
Regular Soft Drinks	1.8479 ^a (0.0338) ^b [0.0001] ^c	0.4185 (0.0355) [0.0001]	16,519.82 [0.0001]	0.91	0.84
Diet Soft Drinks	0.4527 (0.0150) [0.0001]	0.2356 (0.0177) [0.0001]	1994.99 [0.0001]	0.67	0.61
High-Fat Milk	1.2580 (0.0204) [0.0001]	0.2869 (0.0232) [0.0001]	6220.09 [0.0001]	0.82	0.77
Low-Fat Milk	0.2726 (0.0135) [0.0001]	0.2613 (0.0164) [0.0001]	2959.70 [0.0001]	0.63	0.57
Fruit Drinks	0.8761 (0.0169) [0.0001]	0.3191 (0.0197) [0.0001]	3039.86 [0.0001]	0.77	0.68
Fruit Juices	2.0894 (0.0497) [0.0001]	0.5715 (0.0514) [0.0001]	17,882.33 [0.0001]	0.93	0.89
Bottled Water	0.7128 (0.0147) [0.0001]	0.2471 (0.0174) [0.0001]	2343.81 [0.0001]	0.72	0.66
Coffee	0.5086 (0.0255) [0.0001]	0.6736 (0.0297) [0.0001]	539.81 [0.0001]	0.85	0.62
Tea	0.8160 (0.0146) [0.0001]	0.1991 (0.0168) [0.0001]	3127.16 [0.0001]	0.73	0.69
Isotonic Drinks	−1.3167 (0.0114) [0.0001]	0.3487 (0.0258) [0.0001]	20,539.25 [0.0001]	0.29	0.20

^a Estimated coefficient; ^b Estimated standard error; ^c *p*-value. Source: Estimation and calculations made by the authors. Note: All intercept and slope coefficients are statistically significant at the 0.01 level. The Chi-Square statistics correspond to the values associated with the joint test of the null hypothesis that the slope coefficient equals 1 and the intercept coefficient equals 0 (as shown in Equation (4)).

In addition, as exhibited in Table 3, each of the intercept coefficients of the resolution regressions for the set of non-alcoholic beverages was statistically different from zero, and each of the slope coefficients of the resolution regressions was statistically different from one at a level of significance of 0.01. Hence, as evidenced by the chi-squared statistics, the null hypothesis of perfect sorting of probabilities was rejected at a level of significance of 0.01. Although the respective logit models generated upward-sloping resolution graphs as expected, they were relatively flat compared to the 45-degree perfect sorting line.

4. Discussion

To demonstrate the application of resolution regression and resolution graphs, Nielsen Homescan scanner data concerning at-home purchasing behavior associated with ten non-

alcoholic beverages in U.S. households were used in this analysis. Monthly household purchases (expenditure and quantity information) of regular soft drinks, diet (low-calorie) soft drinks, high-fat milk (defined as whole milk and 2% milk), low-fat milk (defined as 1% milk and skim milk), fruit drinks, fruit juices, bottled water, coffee, tea, and isotonic drinks (sports drinks) were obtained over the period January 2003 through December 2003⁶ for 7639 U.S. households as part of a cooperative agreement with the Economic Research Service, U.S. Department of Agriculture.

The quantity data were standardized in terms of gallons per household for calendar year 2003, and expenditure data were expressed in terms of dollars per household for calendar year 2003. The prices associated with purchases were unit values derived as the ratio of expenditures (in dollars) to quantities (in gallons). However, some households may not have purchased non-alcoholic beverage products; in these situations, a weighted average unit value was generated by taking the ratio of the sum of total expenditures of all non-alcoholic beverages to the sum of quantities of all non-alcoholic beverages. This weighted average unit value was used as a proxy for the unobserved price of each non-alcoholic beverage considered in this study⁷. This weighted average unit value was positively correlated with the observed price (unit value) of each non-alcoholic beverage considered in this study⁸, justifying in part the appropriateness of the use of the weighted average unit value as a proxy for unobserved prices or unit values.

Demographic variables serve as the set of explanatory variables used in the respective logit models. Hill and Lynchehaun (2002) identified various cultural and socio-economic factors influencing consumer preferences, including age, ethnicity, income, education, gender, presence of children, region, and race. The detailed description of demographic categories of each variable used in this analysis is shown in Table 4. The demographic categories are age of the household head, employment status of the household head, education level of the household head, region where the household is located, race of the household head, Hispanic origin of the household head, age and presence of children, gender of the household head, and poverty status of the household (as a proxy for household income).

Table 4. Demographic characteristics considered in the binary logit models.

Demographic Characteristics	Categories
Age of Household Head	Less than 25 years
	25–29 years
	30–34 years
	35–44 years
	45–54 years
	55–64 years
	At least 65 years
Employment Status of Household Head	Household head employed full-time
	Household head employed part-time
	Household head not employed
Education Level of Household	Less than high school
	High school level
	Undergraduate level
	Post-college level
Region	Midwest
	South
	West
	East

Table 4. Cont.

Demographic Characteristics	Categories
Race of Household	Black
	Asian
	White
	Other
Hispanic Ethnicity of the Household Head	Hispanic Yes
	Hispanic No
Age and Presence of Children	Less than 6-years-old
	6–12-years-old
	13–17-years-old
	Less than 6-years-old and 6–12-years-old
	Less than 6-years-old and 13–17-years-old
	Less than 6-years-old, 6–12, and 13–17-years-old
Gender of the Household Head	No children under 18 years
	Household head male only
	Household head female only
Poverty Status of the Household	Households with both male and female
	Above poverty line of 185%
	Below poverty line of 185%

Source: Developed by authors.

The information associated with the two random samples of households, Sample A and Sample B, is presented in Table 5. The percentage of households in each demographic category is similar across the two samples of data. Balance tests of the covariates in Sample A (training) and Sample B (testing) are compared using a mean difference test (*t*-test). The associated *p*-values are also reported in Table 5. The covariates in Sample A and Sample B are not statistically different at the 0.05 significance level. As such, these results demonstrate a balance of the number of observations in the training (Sample A) and in the testing (Sample B).

Table 5. Percentage of Households associated with Various Demographic Characteristics in Sample A and Sample B and Covariate Balance Test Results.

Demographic Characteristics	Categories	Percentage Sample A	Percentage Sample B	<i>p</i> -Value for the Balance Tests
Age of Household Head	25–29 years	2.28	2.57	0.40
	30–34 years	5.84	6.50	0.23
	35–44 years	21.80	20.56	0.18
	45–54 years	27.88	27.31	0.57
	55–64 years	22.62	23.88	0.18
	At least 65 years	19.40	18.80	0.50
Employment Status of Household Head	Employed full-time	45.97	44.88	0.33
	Employed part-time	15.84	17.02	0.15
Education Level of Household	High school level	25.24	23.25	0.04
	Undergraduate level	60.50	62.00	0.18
	Post-college level	10.92	11.02	0.88
Region	West	21.13	21.21	0.93
	Midwest	19.24	17.99	0.15
	South	38.53	39.17	0.54
Race of Household	Black	12.36	13.62	0.10
	Asian	2.60	3.14	0.15
	Other	6.10	6.78	0.22

Table 5. Cont.

Demographic Characteristics	Categories	Percentage Sample A	Percentage Sample B	p-Value for the Balance Tests
Hispanic Ethnicity of the Household Head	Hispanic Yes	8.32	7.72	0.33
Age and Presence of Children	Less than 6-years-old	3.64	3.46	0.66
	6-12-years-old	5.76	6.49	0.19
	13-17-years-old	7.23	7.12	0.86
	Less than 6-years-old and 6-12-years-old	2.75	3.11	0.34
	Less than 6-years-old and 13-17-years-old	0.65	0.42	0.15
	6-12 and 13-years-old	4.66	4.14	0.26
	Less than 6-years-old, 6-12, and 13-17-years-old	0.89	1.00	0.63
Gender of the Household Head	Household head male only	27.85	27.26	0.85
	Household head female only	10.52	10.66	0.55
Poverty Status of the Household	Below poverty line of 185%	13.64	13.12	0.52

Note: Sample mean difference test was conducted to compare the two samples, training (Sample A) and testing (Sample B). The null hypothesis is “there is no difference in sample means between Sample A and Sample B. The cut-off *p*-value for the sample mean difference *t*-test is set at 0.01. Source: Developed by authors.

In Table 6, we show the number of households that purchased or did not purchase a given non-alcoholic beverage, as well as the corresponding respective market penetration numbers. Isotonic drinks or sports drinks have the lowest market penetration at 20–22%, while fruit juices have the highest market penetration at 93–94%. The market penetrations for regular soft drinks, coffee and tea, high-fat milk, and low-fat milk are 90–91%, 72–74%, 82–83%, and 61–64%, respectively.

Table 6. Number of households purchasing/not purchasing each beverage category and the accompanying market penetration.

Beverage Category	Sample A Number of Households			Sample B Number of Households		
	Purchase	Non-Purchase	Market Penetration	Purchase	Non-Purchase	Market Penetration
Isotonic drinks	846	2974	22%	763	3056	20%
Regular Soft Drinks	3444	376	90%	3477	342	91%
Diet Soft Drinks	2494	1326	65%	2499	1320	65%
High-Fat Milk	3121	699	82%	3161	658	83%
Low-Fat Milk	2332	1488	61%	2440	1379	64%
Fruit Drinks	2866	954	75%	2923	896	77%
Fruit Juices	3555	265	93%	3584	235	94%
Bottled Water	2693	1127	70%	2681	1138	70%
Coffee	2812	1008	74%	2746	1073	72%
Tea	2753	1067	72%	2785	1034	73%

Source: Calculations made by the authors. Note: Market penetration is defined as the number of households that bought the beverage as a percentage of the total number of households in each sample. The total number of households in Sample A is 3820, and the total number of households in Sample B is 3819.

5. Conclusions, Limitations, and Future Work

Binary choice models are widely used in econometric modeling when the dependent variable corresponds to discrete outcomes. With appropriate decision rules, these models provide predictions of binary choices generated from predicted probabilities. The accuracy

of these predictions in terms of classifying probabilities to events that occurred versus those that did not is a key issue. The use of expectation-prediction success at present is the standard method used to assess the accuracy of these predictions. However, this method is limited in its ability to correctly classify probabilities in the absence of appropriate predetermined cut-off levels. We propose alternative methods to classify probabilities generated through binary choice models, namely resolution graphs and resolution regressions that measure the ability to sort predicted probabilities against observed outcomes.

Using probabilities generated from the use of logit models applied to purchasing decisions of various non-alcoholic beverages, we compare probability sorting power using expectation-prediction success, resolution graphs, and resolution regression. Based on expectation-prediction success, the logit models performed better at classifying outcomes related to purchasing isotonic, regular soft drinks, diet drinks, bottled water, and tea. Based on resolution regressions, the null hypothesis of perfect sorting of probabilities was rejected for all non-alcoholic beverages. Although the logit models generated upward-sloping resolution graphs as expected, they were relatively flat compared to the 45-degree perfect sorting line. Bottom line, the use of resolution regression and resolution graphs yielded information not captured using prediction-success measures alone.

This additional evidence generated from resolution regressions and resolution graphs serves to complement the reporting of standard measures of prediction success. Going forward, for empirical analysis associated with binary choice models, we recommend that practitioners report the results from resolution regression and resolution graphs to capture sorting of probabilities in addition to the standard metrics used for the ability of the models to predict out-of-sample behavior.

This analysis was limited to purchase decisions made by U.S. households concerning ten non-alcoholic beverages for the calendar year 2003. Consequently, this analysis is tantamount to a case study. Since this study uses data from a representative sample of U.S. households for the calendar year 2003, this analysis does not capture the change in consumer behavior and market structure that has occurred in recent years. For future work, this study can be replicated using more current data, along with focusing on a myriad of other products purchased by households.

Author Contributions: Conceptualization, S.D. and D.A.B.; methodology, S.D. and O.C.J.; software, S.D.; validation, S.D. and O.C.J.; formal analysis, S.D. and O.C.J.; investigation, S.D. and D.A.B.; resources, S.D.; data curation, S.D.; writing—original draft preparation, S.D., D.A.B., and O.C.J.; writing—review and editing, S.D. and O.C.J.; visualization, S.D. and D.A.B.; supervision, S.D., D.A.B., and O.C.J.; project administration, S.D.; funding acquisition, not applicable. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: We wish to acknowledge the acquisition of the data for this analysis as part of a cooperative agreement with the U.S. Department of Agriculture (USDA). The results and conclusions associated with this analysis are solely those of the authors and not those of the USDA.

Acknowledgments: We wish to acknowledge comments and suggestions made by anonymous reviewers. Any remaining errors or omissions are those exclusively of the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Estimated coefficients, standard errors, and p -values associated with the respective logit models for each beverage category based on the use of sample A.

	Isotonic Drinks			Regular Soft Drinks			Diet Soft Drinks		
	Coef.	Std. Err.	$p > z $	Coef.	Std. Err.	$p > z $	Coef.	Std. Err.	$p > z $
Price	0.5659	0.1703	0.001	−0.1336	0.1238	0.280	0.2075	0.1291	0.108
price_2	−0.0620	0.0258	0.016	−0.0035	0.0129	0.783	−0.0400	0.0191	0.036
pov185	−0.0831	0.1301	0.523	0.3084	0.1896	0.104	−0.3136	0.1083	0.004
agehh2529	−2.0961	0.8768	0.017	1.3408	1.2209	0.272	−0.2959	0.8190	0.718
agehh3034	−2.0345	0.8570	0.018	1.5795	1.1544	0.171	0.2469	0.8029	0.758
agehh3544	−1.8535	0.8476	0.029	1.2769	1.1158	0.252	0.4807	0.7944	0.545
agehh4554	−2.2248	0.8467	0.009	0.7770	1.1098	0.484	0.2789	0.7927	0.725
agehh5564	−2.4878	0.8493	0.003	0.4738	1.1112	0.670	0.5285	0.7943	0.506
agehhgt64	−2.8128	0.8548	0.001	−0.1427	1.1167	0.898	0.2803	0.7977	0.725
emphhpt	−0.1266	0.1261	0.315	0.1332	0.1942	0.493	−0.0266	0.1118	0.812
emphhft	−0.2457	0.1072	0.022	−0.2434	0.1608	0.130	0.0149	0.0953	0.875
eduhhhs	−0.0739	0.2394	0.757	0.0937	0.3383	0.782	0.0652	0.2015	0.746
eduhhu	−0.2746	0.2342	0.241	−0.0485	0.3268	0.882	0.1820	0.1965	0.354
eduhhpc	−0.4562	0.2672	0.088	−0.6164	0.3504	0.079	0.1891	0.2213	0.393
reg_central	0.2124	0.1333	0.111	0.0648	0.1754	0.712	0.2067	0.1140	0.07
reg_south	0.3229	0.1137	0.005	0.0991	0.1539	0.519	−0.0128	0.0961	0.894
reg_west	0.1574	0.1331	0.237	0.0404	0.1700	0.812	−0.0432	0.1103	0.695
race_black	−0.2448	0.1338	0.067	0.6566	0.2029	0.001	−1.0445	0.1058	0.000
race_oriental	−0.4874	0.2742	0.075	−0.3890	0.3388	0.251	−0.7531	0.2178	0.001
race_other	−0.3464	0.2047	0.091	0.4517	0.3807	0.235	−0.3590	0.1823	0.049
hisp_yes	0.5823	0.1697	0.001	0.3200	0.3216	0.320	−0.0165	0.1624	0.919
agepct6_only	0.5880	0.2052	0.004				−0.0428	0.2013	0.831
agepc6_12only	0.8456	0.1613	0.000				0.2768	0.1676	0.099
agepc13_17only	0.8644	0.1443	0.000				0.3591	0.1517	0.018
agepct6_6_12only	0.6524	0.2261	0.004				0.0336	0.2335	0.885
agepct6_13_17only	0.9476	0.4229	0.025				−0.0680	0.4298	0.874
agepc6_12and13_17only	0.6141	0.1819	0.001				−0.0672	0.1810	0.71
agepct6_6_12and13_17	0.1304	0.3921	0.739				0.1052	0.4004	0.793
fhonly	−0.2975	0.1064	0.005	−0.9039	0.1309	0.000	−0.2864	0.0860	0.001
mhonly	−0.3086	0.1587	0.052	−1.1402	0.1631	0.000	−0.9961	0.1196	0.000
_constant	0.0843	0.9066	0.926	2.4588	1.1612	0.034	0.2491	0.8350	0.765
Number of Observations	3820			3820			3820		
McFadden R ²	0.0754			0.0925			0.0538		
Log Likelihood	−1867.50			−1114.98			−2333.56		
	High-Fat Milk			Low-Fat Milk			Fruit Drinks		
	Coef.	Coef.	$p > z $	Coef.	Std. Err.	$p > z $	Coef.	Std. Err.	$p > z $
price	−0.3177	−0.3177	0.001	0.7734	0.1378	0.000	0.1828	0.0986	0.064
price_2	0.0285	0.0285	0.016	−0.0932	0.0212	0.000	−0.0141	0.0117	0.229
pov185	0.2431	0.2431	0.523	−0.3362	0.1058	0.001	0.1430	0.1251	0.253
agehh2529	0.2620	0.2620	0.017	−0.3687	0.9085	0.685	0.9060	0.9316	0.331
agehh3034	1.2414	1.2414	0.018	−0.5077	0.8904	0.569	1.1157	0.8992	0.215
agehh3544	0.9501	0.9501	0.029	−0.4807	0.8824	0.586	0.8837	0.8784	0.314
agehh4554	0.9357	0.9357	0.009	−0.5101	0.8811	0.563	0.3185	0.8754	0.716
agehh5564	1.0403	1.0403	0.003	−0.4791	0.8823	0.587	−0.1568	0.8761	0.858
agehhgt64	0.9629	0.9629	0.001	−0.2143	0.8855	0.809	−0.5464	0.8797	0.534
emphhpt	0.2161	0.2161	0.315	−0.0791	0.1085	0.466	0.1905	0.1291	0.140
emphhft	0.2629	0.2629	0.022	−0.1247	0.0920	0.175	−0.1608	0.1069	0.132
eduhhhs	0.0172	0.0172	0.757	0.5064	0.1972	0.010	0.1636	0.2228	0.463
eduhhu	−0.3198	−0.3198	0.241	0.7174	0.1927	0.000	0.0645	0.2168	0.766
eduhhpc	−0.6136	−0.6136	0.088	0.9201	0.2181	0.000	−0.2107	0.2420	0.384
reg_central	0.2165	0.2165	0.111	−0.1621	0.1099	0.140	−0.0048	0.1238	0.969
reg_south	0.2485	0.2485	0.005	−0.2899	0.0946	0.002	−0.0390	0.1081	0.718
reg_west	−0.0294	−0.0294	0.237	−0.2657	0.1091	0.015	−0.1240	0.1224	0.311

Table A1. Cont.

	High-Fat Milk			Low-Fat Milk			Fruit Drinks		
	Coef.	Coef.	$p > z $	Coef.	Std. Err.	$p > z $	Coef.	Std. Err.	$p > z $
race_black	0.2237	0.2237	0.067	−0.8052	0.1053	0.000	1.0953	0.1535	0.000
race_oriental	−0.0749	−0.0749	0.075	−0.3283	0.2200	0.136	0.1695	0.2752	0.538
race_other	−0.1828	−0.1828	0.091	−0.3934	0.1777	0.027	0.2654	0.2310	0.251
hisp_yes	0.5899	0.5899	0.001	−0.0579	0.1571	0.713	0.3246	0.1998	0.104
agepctl6_only	2.1597	2.1597	0.004	0.2563	0.2039	0.209			
agepc6_12only	1.0907	1.0907	0.000	0.0637	0.1579	0.687			
agepc13_17only	0.4222	0.4222	0.000	0.1760	0.1407	0.211			
agepctl6_6_12only	1.1356	1.1356	0.004	0.2517	0.2325	0.279			
agepctl6_13_17only	0.9242	0.9242	0.025	−0.2017	0.4221	0.633			
agepc6_12and13_17only	0.3981	0.3981	0.001	0.1064	0.1771	0.548			
agepctl6_6_12and13_17	0.5935	0.5935	0.739	0.4642	0.3997	0.246			
fhonly	−0.4079	−0.4079	0.005	−0.2874	0.0837	0.001	−0.4075	0.0929	0.000
mhonly	−0.7848	−0.7848	0.052	−0.5944	0.1192	0.000	−1.0045	0.1236	0.000
_constant	1.0970	1.0970	0.926	−0.4634	0.9183	0.614	0.7146	0.9107	0.433
Number of Observations	3820	3820		3820			3820		
McFadden R ²	0.0615	0.0615		0.0455			0.0862		
Log Likelihood	−1706.05	−1706.05		−2437.65			−1961.99		
	Fruit Juice			Bottled Water					
	Coef.	Std. Err.	$p > z $	Coef.	Std. Err.	$p > z $			
price	0.9083	0.1458	0	0.2251	0.0967	0.02			
price_2	−0.0568	0.0142	0	−0.0157	0.012	0.19			
pov185	−0.1683	0.1881	0.371	−0.4521	0.1108	0			
agehh2529	−13.121	744.3962	0.986	−0.5998	1.1273	0.595			
agehh3034	−11.9996	744.3962	0.987	−0.2816	1.1114	0.8			
agehh3544	−13.0856	744.3961	0.986	−0.6098	1.0996	0.579			
agehh4554	−13.1990	744.3961	0.986	−0.8437	1.0979	0.442			
agehh5564	−13.1709	744.3961	0.986	−0.9895	1.0987	0.368			
agehhgt64	−12.8456	744.3961	0.986	−1.4964	1.101	0.174			
emphhpt	0.2945	0.239	0.218	0.1689	0.1165	0.147			
emphhft	−0.3136	0.1771	0.077	0.0948	0.0992	0.339			
eduhhhs	0.3233	0.3133	0.302	−0.1668	0.2128	0.433			
eduhhu	0.7058	0.3062	0.021	−0.1832	0.208	0.378			
eduhhpc	0.6285	0.3589	0.08	−0.2899	0.2334	0.214			
reg_central	−0.0387	0.224	0.863	−0.1941	0.115	0.092			
reg_south	−0.3086	0.1917	0.107	−0.1032	0.1009	0.307			
reg_west	−0.7758	0.2062	0	0.0894	0.1181	0.449			
race_black	0.845	0.2575	0.001	0.3303	0.1211	0.006			
race_oriental	0.3293	0.4855	0.498	0.1079	0.2617	0.68			
race_other	−0.3499	0.3239	0.28	0.4361	0.2171	0.045			
hisp_yes	0.3386	0.3241	0.296	0.142	0.183	0.438			
agepctl6_only				0.1246	0.2332	0.593			
agepc6_12only				0.3047	0.1922	0.113			
agepc13_17only				0.1678	0.1584	0.289			
agepctl6_6_12only				−0.0782	0.2547	0.759			
agepctl6_13_17only				0.3251	0.5574	0.56			
agepc6_12and13_17only				0.3115	0.2147	0.147			
agepctl6_6_12and13_17				0.5746	0.5035	0.254			
fhonly	−0.6982	0.1572	0	−0.0720	0.0895	0.421			
mhonly	−1.3657	0.1849	0	−0.7892	0.122	0			
_constant	14.1394	744.3961	0.985	1.6086	1.122	0.152			
Number of Observations	3820			3820					
McFadden R ²	0.0864			0.0559					
Log Likelihood	−879.50			−2187.69					

Table A1. Cont.

	Coffee			Tea		
	Coef.	Std. Err.	$p > z $	Coef.	Std. Err.	$p > z $
price	−0.9775	0.1161	0.000	0.1480	0.0949	0.119
price_2	0.0491	0.0148	0.001	−0.0101	0.0112	0.368
pov185	−0.3628	0.1257	0.004	0.0616	0.1178	0.601
agehh2529	0.4237	0.8358	0.612	−0.1606	0.9200	0.861
agehh3034	0.5214	0.8179	0.524	0.1605	0.9029	0.859
agehh3544	1.2749	0.8096	0.115	0.1724	0.8932	0.847
agehh4554	1.6393	0.8085	0.043	0.3366	0.8916	0.706
agehh5564	1.9975	0.8111	0.014	0.1239	0.8926	0.890
agehhgt64	2.3876	0.8175	0.003	0.0574	0.8964	0.949
emphhpt	−0.0661	0.1281	0.606	−0.0080	0.1214	0.947
emphhft	−0.0031	0.1091	0.977	−0.3663	0.1015	0.000
eduhhhs	−0.5881	0.2945	0.046	0.2429	0.2088	0.245
eduhhu	−0.6844	0.2875	0.017	0.3716	0.2034	0.068
eduhhpc	−0.6825	0.3084	0.027	0.3356	0.2295	0.144
reg_central	−0.3497	0.1289	0.007	−0.9535	0.1206	0.000
reg_south	0.0480	0.1130	0.671	−0.3878	0.1105	0.000
reg_west	−0.3768	0.1252	0.003	−0.7964	0.1211	0.000
race_black	−0.6734	0.1151	0.000	−0.0020	0.1169	0.986
race_oriental	0.0185	0.2454	0.94	−0.2995	0.2268	0.187
race_other	−0.1813	0.2144	0.398	0.1562	0.1964	0.426
hisp_yes	0.4499	0.1959	0.022	−0.2554	0.1676	0.128
agepct6_only	−0.0085	0.2133	0.968	0.5069	0.2442	0.038
agepc6_12only	0.2355	0.1787	0.188	0.1340	0.1765	0.448
agepc13_17only	0.0344	0.1575	0.827	0.2465	0.1594	0.122
agepct6_6_12only	0.2818	0.2541	0.267	−0.1853	0.2408	0.442
agepct6_13_17only	−0.1492	0.4462	0.738	0.1725	0.5116	0.736
agepc6_12and13_17only	−0.3532	0.1881	0.060	−0.0493	0.1921	0.797
agepct6_6_12and13_17	0.9204	0.5173	0.075	0.4005	0.4706	0.395
fonly	−0.6690	0.0982	0.000	−0.2944	0.0905	0.001
monly	−1.0535	0.1322	0.000	−0.8637	0.1216	0.000
_constant	2.7379	0.8684	0.002	1.0259	0.9240	0.267
Number of Observations	3820			3820		
McFadden R ²	0.1415			0.0437		
Log Likelihood	−1892.47			−2163.69		

Source: Estimation done by the authors using the software package SAS, Version 9.4. Note: Calculation of a goodness-of-fit measure like the McFadden R-squared statistic is typical of binary choice models. Given the variability in cross-sectional data (household-level purchase data) used in this analysis, these measures, not surprisingly, are relatively close to zero, ranging from 0.0437 for tea to 0.1415 for coffee.

Brier Score: The magnitude of the Brier Score is inversely related to forecast performance. The Brier Scores in this analysis ranged from 0.06 to 0.22. The Brier scores were lowest for fruit juices and regular soft drinks but highest for low-fat milk and diet soft drinks.

Kullback–Leibler Information Criterion: According to this approach, we examine the “closeness” or deviation of the respective logit models in generating out-of-sample probabilities relative to the true data-generating probability distribution. The Kullback–Leibler Information Criterion (KLIC) uses the calculated log-likelihood function values associated with each logit model in the analysis. The KLIC values varied from −0.62 to −0.21. According to this criterion, the logit model associated with fruit juices generated probabilities that moved most closely with the true data-generating process.

Table A2. Results from Area under the ROC curve, Brier Score, and Out-of-Sample Log-likelihood Function Values Associated with the Probabilities Generated by the Respective Logit Models.

Beverage	Area Under ROC Curve	Brier Score	Kullback–Leibler Information Criteria
Isotonic drinks	0.67	0.15	−0.47
Regular soft drinks	0.69	0.08	−0.28
Diet soft drinks	0.64	0.22	−0.62
High-fat milk	0.67	0.14	−0.43
Low-fat milk	0.66	0.22	−0.62
Fruit drinks	0.68	0.17	−0.50
Fruit juices	0.74	0.06	−0.21
Bottled water	0.63	0.20	−0.58
Coffee	0.73	0.18	−0.53
Tea	0.63	0.19	−0.56

Note: Receiver Operating Characteristic (ROC): This measure evaluates forecast probabilities considering a wide range of probability cut-off values (not just the standard 0.5 cut-off). Logit models with the highest calculated area under the ROC chart provide evidence of the best forecast performance out-of-sample. In this analysis, the respective ROC metrics ranged from 0.63 (bottled water and tea) to 0.74 (fruit juices).

Accuracy of the out-of-sample predicted probabilities from the observed outcome indices was compared against other methods (ROC curves, Brier Score, and Kullback–Leibler criteria) across the logit models for all beverages. When market penetration cut-off values were used using expectation-prediction success tables, the sum of sensitivity and specificity was highest for regular soft drinks (1.35), followed by coffee (1.31). This measure was lowest for tea (1.18). When the ROC charts were used, the logit model for fruit juice resulted in the highest ROC value at 0.74, followed by the logit model for coffee at 0.73. The Brier Scores were lowest for the logit model for fruit juice at 0.06 and for the logit model for regular soft drinks at 0.08. Based on the Kullback–Leibler criteria, probabilities generated from the logit model for fruit juice, followed by those from the logit model for regular soft drinks, produced the best out-of-sample probabilities. Bottom line, the results produced from these various metrics were not uniform in providing the best out-of-sample probabilities generated by the respective logit models. That said, based on these criteria, logit models associated with fruit juices, regular soft drinks, and coffee provided the best out-of-sample forecasts of the generated probabilities.

Although the Brier score indicates the degree of the ability of the logit models to forecast accurately, the components of the Yates decomposition of the Brier score provide more detailed information. Larger values of the Brier scores were primarily due to larger variances of outcome indices. In this analysis, the bias component was negligible for all forecast probabilities associated with the respective non-alcoholic beverages. Scatter and minimum variance directly contributed to the variance of the forecast probabilities. The lowest variance of output index, scatter, and minimum variance were associated with fruit juices. However, the highest covariance of outcome index and forecast probabilities occurred for coffee.

Table A3. Results from Yates Decomposition of the Brier Score for Probabilities Generated by the Logit Models for Non-alcoholic Beverages.

Beverage	Yates Decomposition of the Brier Score				
	Variance of Outcome Index	Minimum Variance Probabilities	Scatter	Bias	Covariance of Outcome Index and Probability
Isotonic drinks	0.16	0.0007	0.01	0.00048	0.02
Regular soft drinks	0.08	0.0001	0.004	0.000062	0.007
Diet soft drinks	0.23	0.0007	0.01	0.000025	0.02
High-fat milk	0.14	0.0003	0.008	0.00014	0.01
Low-fat milk	0.23	0.0009	0.01	0.00086	0.03
Fruit drinks	0.18	0.0007	0.01	0.00014	0.02
Fruit juices	0.06	0.0001	0.003	0.000038	0.005
Bottled water	0.21	0.0006	0.01	0.000034	0.02
Coffee	0.20	0.0040	0.03	0.00013	0.06
Tea	0.20	0.0003	0.009	0.000031	0.02

Note: Yates Decomposition of the Brier score: The Yates decomposition separates the Brier Score measure into various components--variance of the observed outcome index, minimum variance of the associated probability forecast, scatter of the associated probability forecast, miscalibration measured using bias, and covariance between the outcome index and associated probability forecast. Using this measure, better forecasting models are linked to lower variances of outcome indices as well as lower scatter, bias, and minimum variance. Additionally, for better forecasting models, we expect higher covariances of the forecast probabilities generated with their outcome indices.

Notes

- ¹ A wide range of qualitative choice models are available, such as probit, logit, mixed logit, ordered probit, generalized extreme value, nested logit, multinomial probit, multinomial logit, etc.
- ² The probit and logit models yield similar results in the case of binary choice models. Additionally, since the logistic density function closely resembles the t-distribution with seven degrees of freedom (Hanushek & Jackson, 1977), the logit and probit formulations are quite similar. The only difference is that the logistic density has a slightly heavier tail than the standard normal density. For probit models, the error term is assumed to follow a *standard normal distribution*, and for logit models, the error term is assumed to follow a *logistic distribution*.
- ³ Receiver Operating Characteristics (ROC) curves (Hsieh & Turnbull, 1996; Reiser & Faraggi, 1997) offer relief for this issue in classifying probabilities by calculating and plotting probability outcomes based on a wide range of cut-off probabilities. Alternatively, a log-likelihood function approach, which selects models closest to the true data-generating process based on the Kullback–Leibler Information Criterion, has been used to assess the performance of models generating probabilities (Stone, 1977; Shao, 1993; Norwood et al., 2004). Additionally, other techniques such as calibration, calibration graphs, and scoring rules such as the Brier Score (Brier, 1950) and the Yates partition of the Brier Score (Yates, 1982) have been used to measure the accuracy of these types of predictions (Zellner et al., 1991). In Appendix A, we provide two tables associated with these metrics applied to out-of-sample probabilities generated in this study. In this way, we provide a more inclusive picture of additional methods available to evaluate forecast performance of binary choice models other than the measures proposed in this study (resolution regression and resolution graphs).
- ⁴ Although we rely on the use of binary logit models, the analysis can nevertheless be carried out using binary probit models. Because no implicit ordering is associated with the decision to purchase non-alcoholic beverages, we rely on the use of binary logit models in this analysis. Also, we performed the same analysis using binary probit models using the same data. Since the probit model analysis resulted in very similar outcomes/results, for brevity, we do not produce the probit model results here.
- ⁵ Alternatives to selecting one cut-off probability value to correctly classify probabilities are the receiver operating characteristic curve (ROC) and the cumulative accuracy profile (CAP) chart. The use of ROCs and CAP charts is beyond the scope of this analysis and is reserved for future research. The theoretical framework for ROC and CAP charts can be found in Mann and Whitney (1947) and Bamber (1975).
- ⁶ The use of these data is for demonstration purposes only in this analysis. No claims are made regarding the current market behavior of households in the decision to purchase or not to purchase non-alcoholic beverages.

- ⁷ Alternatively, to impute prices, some studies used auxiliary regressions based on observed unit values as a function of household income, household size, and region.
- ⁸ The Pearson correlation coefficients of the weighted average unit value and the observed unit values of each non-alcoholic beverage are shown below. The null hypothesis of “no correlation” is soundly rejected. *p*-values are given in parentheses.

	Observed Unit Value									
	Isotonic drinks	Regular soft drinks	Diet soft drinks	High-fat milk	Low-fat milk	Fruit drinks	Fruit juice	Bottled water	Coffee	Tea
Weighted average unit value	0.06 −0.0001	0.25 −0.0001	0.11 −0.0001	0.12 −0.0001	0.15 −0.0001	0.19 −0.0001	0.25 −0.0001	0.19 −0.0001	0.07 −0.0001	0.23 −0.0001

References

- Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, 12(4), 387–415. [\[CrossRef\]](#)
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1–3. [\[CrossRef\]](#)
- Dawid, A. P. (1986). *Probability forecasting: Encyclopedia of statistical sciences* (vol. 7). Wiley.
- Greene, W. H. (2012). *Econometric analysis* (7th ed.). Pearson.
- Hanushek, J., & Jackson, E. (1977). *Statistical methods for social scientists*. Academic Press.
- Hill, H., & Lynchehaun, F. (2002). Organic milk: Attitudes and consumption patterns. *British Food Journal*, 104(7), 526–542. [\[CrossRef\]](#)
- Hsieh, F., & Turnbull, B. W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Annals of Statistics*, 24(1), 25–40. [\[CrossRef\]](#)
- Kennedy, P. (2003). Limited dependent variables. In *A guide to econometrics* (5th ed.). MIT Press.
- Maddala, G. S. (1983). *Limited dependent and qualitative variables in econometrics*. Cambridge University Press.
- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1), 50–60. [\[CrossRef\]](#)
- McFadden, D. (1984). Econometric analysis of qualitative response models. In Z. Griliches, & M. Intriligator (Eds.), *Handbook of econometrics* (vol. 2, pp. 1395–1457). Elsevier.
- Norwood, F. B., Lusk, J. L., & Brorsen, B. W. (2004). Model selection for discrete dependent variables: Better statistics for better steaks. *Journal of Agricultural and Resource Economics*, 29(3), 404–419.
- Pindyck, R. S., & Rubinfeld, D. L. (1998). *Econometric models and economic forecasts* (4th ed.). McGraw-Hill.
- Reiser, B., & Faraggi, D. (1997). Confidence intervals for the generalized ROC criterion. *Biometrics*, 53, 644–652. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sanders, F. (1963). On subjective probability forecasting. *Journal of Applied Meteorology*, 2(2), 191–201. [\[CrossRef\]](#)
- SAS Institute, Inc. (2022). *SAS/STAT version 9.4, user guide*. SAS Institute, Inc.
- Shao, J. (1993). Linear model selection by cross-validation. *Journal of American Statistical Association*, 88, 486–494. [\[CrossRef\]](#)
- Stock, J. H., & Watson, M. W. (2007). *Introduction to econometrics* (2nd ed.). Pearson Education Inc.
- Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike’s criterion. *Journal of Royal Statistical Society, Series B (Methodological)*, 39(1), 44–47. [\[CrossRef\]](#)
- Train, K. E. (2003). *Discrete choice methods with simulation* (1st ed.). Cambridge University Press.
- Yates. (1982). External correspondence: Decompositions of the mean probability score. *Organizational Behavior and Human Performance*, 30, 132–156. [\[CrossRef\]](#)
- Zellner, A., Hong, C., & Min, C. (1991). Forecasting turning points in international output growth rates using bayesian exponentially weighted autoregression, time-varying parameter, and pooling techniques. *Journal of Econometrics*, 49, 275–304. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.