The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Decomposition of Discrete Choice Model Generated Probabilities and their Robustness to Changing Substantive Knowledge (Conditioning Variables)

Senarath Dharmasena
Department of Agricultural Economics
Texas A&M University
College Station, TX 77843-2124
sdharmasena@tamu.edu

David A. Bessler
Department of Agricultural Economics
Texas A&M University
College Station, TX 77843-2124
d-bessler@tamu.edu

Selected Paper prepared for presentation at the Southern Agricultural Economics Association's 2016 Annual Meeting, San Antonio, Texas, February 6-9, 2016

Copyright 2016 by Senarath Dharmasena and David A. Bessler. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Decomposition of Discrete Choice Model Generated Probabilities and their Robustness to Changing Substantive Knowledge (Conditioning Variables)

Abstract

Clear understanding of "goodness" and how substantive knowledge contributes to such goodness is generally absent from the economist's use of probability. Although probability forecast from either subjective experts or from data based on prior theory and models can be generated, it is more problematic to generate a "good probability forecast" with a crisp understanding of what constitutes "good". Further it is generally not clear to economists how different conditioning information affects this measure of "good."

Heretofore probability forecasts have been evaluated using the Brier Score and its Yates partition. Our work contributes by exploring how different sets of substantive information affect the Brier score and each component of the Yates partition. We will explore partitions associated with a set of observational data on beverages and the associated consumer decision to purchase. Probabilities are modeled using discrete choice models.

Results show the higher the substantive knowledge, higher the model's ability to offer a high probability for events occurred versus low probability for events that did not occur. Also, this model gave rise to lower Brier Score (lower the better) and higher covariance between probabilities offered and events observed. Better sorting of probabilities was demonstrated in the model with more substantive knowledge.

Keywords: Brier Score, Yates partition, probabilities, discrete choice model, beverages

JEL Classification: C18

Decomposition of Discrete Choice Model Generated Probabilities and their Robustness to Changing Substantive Knowledge (Conditioning Variables)

Problem Statement and Justification:

Probability has been at the center of economics and decision making. Yet a clear understanding of "goodness" and how substantive knowledge (conditioning variables) contributes to such goodness is generally absent from the economist's use of probability. While it is, perhaps, simple to generate a probability forecast from either subjective experts or from data based on prior theory and estimated models, it is more problematic to generate a "good probability forecast" with a crisp understanding of what constitutes "good." Further it is generally not clear to economists how different conditioning information sets affect our measure of "good."

Heretofore probability forecasts have been evaluated using the Brier Score (Brier 1950). Some work in economics has explored (rather superficially) this metric and its Yates partition (Yates, 1988). Our work seeks to contribute to the knowledge-base by exploring how different sets of substantive information affect the Brier score and each component of the Yates partition. In particular, we will explore partitions associated with a set of observational data on non-alcoholic beverages and the associated consumer decision to purchase. Beverages studied are regular soft drinks, diet soft drinks, isotonics (sports drinks), high-fat milk, low-fat milk, fruit drinks, fruit juices, bottled water, coffee and tea. Our interest is to explore for beverages having good Brier score components over bad Brier score components. Probabilities will be modeled using discrete choice models.

Preliminary work by Dharmasena (2010) and Dharmasena et al., (2014), on evaluating probabilities generated through discrete choice models in modeling the decision to purchase a given non-alcoholic beverages by a sample of U.S. households, suggests following mixed results. Probabilities generated through probit model in the decision to purchase fruit juices gave the lowest Brier score and low-fat milk had the highest Brier score, demonstrating good and poor probability forecasts, respectively. However, as far as the Yates partition of the Brier score is concerned, the most important component, the covariance between the probabilities generated through the model and outcome index, as identified by Yates (1988) is the heart of the forecasting problem. The highest covariance is associated with beverage coffee and the lowest covariance is associated with fruit juices. This result suggests the probability forecasts associated with fruit juices to be worse compared to such probability forecasts associated with coffee. However, according to the Brier score estimate, fruit juices had the best model. These results are contradictory. Therefore, our effort will be geared to find potential effects of the substantive knowledge (information we place on the right-hand side of discrete choice models) on generating "good" probability forecasts, measured through the Brier score and various components of the Yates partition of the Brier score.

Objectives:

Specific objectives of this study are to: (1) study the robustness of the Brier score to changes in substantive knowledge; (2) study the robustness of the components of the Yates partition to the Brier score to changes in substantive knowledge; both using probabilities generated through discrete choice models.

Methodology:

For each beverage, we generate the Brier score and its Yates partition for varying combinations of substantive knowledge variables (price, age, education status, employment status, region, race, Hispanic status, age and presence of children, poverty status and gender of the household head). The behavior of the components of the Yates partition of the Brier score (such as variance of the indicator variable, bias, scatter, minimum variance and the covariance of probabilities and outcome indexes) and robustness of each part to varying combinations of substantive knowledge were explored. This would allow us to bring a meaningful interpretation to each component of the Yates partition in terms of evaluating probabilities, as well as to discover the most important component of the Yates partition in sorting probabilities.

The Brier Score and the Yates Partition of the Brier Score

The following discussion on Brier score (BS), and the Yates partition of the Brier score follows from Brier (1950), Yates (1982), Yates and Curley (1985) and Yates (1988).

Let *f* represent the probabilistic forecast for an event that the forecaster is trying to predict (in our analysis, probabilities are generated using qualitative choice models). Let *d* represent the outcome index where, $d = 1$ if the event occurs and $d = 0$ if the event does not occur. As shown in equation (1), the probability score *(PS)* is formally defined as the squared difference between *f* and *d*.

$$PS(f,d) = (f - d)^2 \qquad (1)$$

The *PS*is bounded $0 \le PS \le 1$. Over *N* occasions, indexed by $i = 1,....., N$, the mean of the PS (or $\overline{PS}$ or the Brier score) is given by

$$\overline{PS}(f,d) = \frac{1}{N}\sum_{i=1}^{N}(f_i - d_i)^2 \qquad (2)$$

Sanders (1963) and Murphy (1972a, 1972b, 1973) have decomposed the Brier score into various components including measures of calibration and resolution. However, Yates (1982), Yates and Curley (1985), and Yates (1988) further decomposed the Brier score into its variance and covariance components allowing for additional analysis. His formulation called *"covariance decomposition"* is given as follows.

$$\overline{PS}(f,d) = Var(d) + MinVar(f) + Scat(f) + Bias^2 - 2 * Cov(f,d) \quad (3)$$

The various components of $\overline{PS}$ on the right hand side of equation (3) have following definitions and interpretations. *Var(d)* represents the variance of the outcome index and defined as:

$$Var(d) = \bar{d}(1-\bar{d}) \quad (4)$$

where

$$\bar{d} = \frac{1}{N}\sum_{i=1}^{N} d_i \quad (5)$$

Equation (4) shows the relative frequency or the "base rate" with which the target event occurs, where the target event for our analysis would be the decision to *buy* coffee. This decision is completely out of control of the forecaster (in our analysis the forecaster is the qualitative choice model), hence the *Var(d)* is not determined through our model. The remaining terms reflect the factors that are under the model's control. Thus we want to minimize, $Scat(f)$ and $Bias^2$, while maximizing $Cov(f,d)$ for an allowable minimum variance ($MinVar(f)$) to obtain the lowest $\overline{PS}$. It should be noted that our objective is to minimize the $\overline{PS}$ in evaluating probabilities, because the lower the Brier score, the higher the ability of the model to correctly classify probabilities.

Bias is defined as follows.

$$Bias = (\bar{f} - \bar{d}) \quad (6)$$

where

$$\bar{f} = \frac{1}{N}\sum_{i=1}^{N} f_i \quad (7)$$

In the equation (6), $\bar{f}$ is the mean of the probabilities generated from the model. Bias reflects the overall miscalibration of the forecast. The square of the bias, which is what actually appears in the covariance decomposition (equation 3), reflects the calibration error regardless of the direction (+ or -) of the error.

The $Cov(f, d)$ term is defined as follows.

$$Cov(f, d) = [slope][Var(d)] \quad (8)$$

The slope is defined as the difference between the means of conditional probability of events that actually occurred and conditional probability of events that actually did not occur. Algebraically the slope is defined as follows.

$$Slope = (\bar{f}_1 - \bar{f}_0) \quad (9)$$

where

$$\bar{f}_1 = \frac{1}{N_1}\sum_{j=1}^{N_1} f_{1j} \quad (10)$$

$$\bar{f}_0 = \frac{1}{N_0}\sum_{j=1}^{N_0} f_{0j} \quad (11)$$

Here $\bar{f}_1$ represents the conditional mean probability forecast for event under consideration over the N_1 occurrences for which the event actually occurs; $\bar{f}_0$ represents the conditional mean probability for event under consideration over the N_0 occurrences that the event does not occur, with $N = N_1 + N_0$. The maximum value that Slope can have is 1, which occurs when the model always reports $f = 1$ when the target event is going to occur and $f = 0$ when it is not. Furthermore, Slope is the gradient of the regression line when probabilities generated through the model are regressed on outcome indexes. For a perfect forecast, all the probabilities associated with events that do not occur must have probabilities equal to zero and all probabilities associated with events that did occur must have probabilities equal to one, resulting in a slope equal to one. Therefore, it makes sense for Slope to contribute to mean probability score negatively. In other words, steeper the Slope, the more appropriate the classification of probabilities for events that occurred and that did not occur (high probabilities for event that occurred and lower probabilities for events that did not occur, the smaller the Brier score the better).

Covariance between the probabilities generated through the model and outcome index $Cov(f,d)$ is the heart of the forecasting problem (Yates, 1988). It reflects the ability of the model to make distinctions between individual occasions in which the event occurs or does not occur. In other words, it represents how responsive the forecast is to information related to the event. Our objective with respect to minimum variance is that the model needs to maximize the value associated with the $Cov(f,d)$ to achieve a lower Brier score.

Scatter is defined as the mean of the weighted variances of probabilities associated with events that occurred and that did not occur. The algebraic representation of scatter is depicted in equation (12) below.

$$\mathrm{Scat(f)} = \frac{1}{\mathrm{N}}[\mathrm{N_1 Var(f_1)} + \mathrm{N_0 Var(f_0)}] \qquad (12)$$

where

$$\mathrm{Var(f_1)} = \frac{1}{\mathrm{N_1}}\sum_{i=1}^{\mathrm{N_1}}(\mathrm{f_{1i}} - \bar{\mathrm{f}}_1)^2 \qquad (13)$$

and

$$\mathrm{Var(f_0)} = \frac{1}{\mathrm{N_0}}\sum_{i=1}^{\mathrm{N_0}}(\mathrm{f_{0i}} - \bar{\mathrm{f}}_0)^2 \qquad (14)$$

$\mathrm{Var(f_1)}$ is the conditional variance of the probabilities generated from the model associated with the events on those N_1 occasions when the event actually occurred and $Var(f_0)$ is the conditional variance of the probabilities generated from the model associated with the events on those N_0 occasions when the event actually did not occur. $Var(f_1)$ and $Var(f_0)$ measure variability in model generated probabilities which is unrelated to whether or not the target event occurs. Scatter can be interpreted as an index of overall noise contained in model generated probabilities. It is expected that the Scatter will be minimized to achieve a lower mean probability score.

$MinVar(f)$ is defined as follows.

$$\mathrm{MinVar(f)} = \mathrm{Var(f)} - \mathrm{Scat(f)} \tag{15}$$

where $Var(f)$ is the variance of the entire collection of probabilities generated for the target event. Minimum variance can also be shown as follows.

$$\mathrm{MinVar(f)} = (\bar{f}_1 - \bar{f}_0)^2[\bar{d}(1-\bar{d})] \tag{16}$$

which contains the elements of the covariance of judgments and outcome indexes (Yates, 1988). To give more perspective to the relationship between minimum variance and overall variance of the probabilities generated through the models, we can rearrange equation (15) as follows.

$$\mathrm{Var(f)} = \mathrm{MinVar(f)} + \mathrm{Scat(f)} \tag{17}$$

Minimum variance can also be defined as the variance of probabilities on top of scatter that contributes toward the overall variance, i.e. $Var(f)$.

Since $Var(f)$ contributes to the Brier score positively, one would want to minimize it. That is to say, in the equation (17), we have to minimize the components in the right hand side, i.e. $MinVar(f)$ and $Scat(f)$. It would make sense to minimize $Scat(f)$ of probabilities as lower the $Scat(f)$ the tighter the distribution of probabilities around conditional means of probabilities for events that actually occurred and events that did not occur the better the model's ability to sort probabilities for events that occurred versus events that did not occur. However, it would not make sense to minimize the $MinVar(f)$ in trying to minimize the overall variance of the probabilities generated. This is clear when one looks at the equation (16). $MinVar(f)$ is a function of Slope and variance of index variable, where the latter is not determined through the model that we used to generate probabilities. The only manipulatable component is the Slope, which is a function of conditional probabilities. What is desired is to have a maximum slope of one at the extreme in minimizing the Brier score. However, in trying to minimize the $Var(f)$, if one minimizes the $MinVar(f)$, it will eliminate the Slope, which is not desirable. Therefore, we need to have some Slope, hence some $MinVar(f)$ in the model, in minimizing $Var(f)$ and trying to achieve the minimum Brier score. Therefore, $MinVar(f)$ essentially reflects the *maximum allowable model variability* (or amount of model variability that must be tolerated) which is required to minimize the $Var(f)$, hence the Brier score.

Since $Cov(f,d)$ and $MinVar(f)$ are both functions of Slope, $(\bar{f}_1 - \bar{f}_0)$ and Variance of outcome index, $(\bar{d}(1-\bar{d})$, we can establish a relationship between $Cov(f,d)$ and $MinVar(f)$ as follows. Equation (16) can be rearranged to represent the Slope as follows;

$$(\bar{f}_1 - \bar{f}_0) = \sqrt{\frac{\mathrm{MinVar(f)}}{\mathrm{Var(d)}}} \quad (18)$$

Substituting (18) into (8) and after simplification we arrive at the following relationship that combines Covariance of forecast probabilities, Minimum Variance and Variance of outcome index as follows.

$$\mathrm{Cov(f,d)} = \sqrt{\mathrm{MinVar(f)} * \mathrm{Var(d)}} \quad (19)$$

According to equation (19), variance of outcome index and Minimum Variance are positively related to the covariance of forecast probabilities and outcome index. It is an obvious fact that variance of the outcome index, $Var(d)$ is beyond the control of the forecasting model and only determined externally by the actual observations. Therefore, the only model generated variable that affect the $Cov(f,d)$ is $MinVar(f)$. We can conclude that higher the Slope, the higher the $MinVar(f)$, the higher the $Cov(f,d)$. In other words, high $MinVar(f)$ is associated with high $Cov(f,d)$. This result has leverage in explaining the forecasting model's sorting power (resolution) and $Cov(f,d)$. We also can conclude that, high resolution is associated with high $Cov(f,d)$.

It is important to note that, although the Brier score gives an overall indication of the ability of the model to forecast (the lower the Brier score, the better the forecast), the components of the covariance decomposition of the Brier score provides a clearer indication of the ability of the model to forecast as well as to sort probabilities.

Data:

Data are observational, collected by Nielsen HomeScan scanner data for household purchases of various non-alcoholic beverages (expenditure, quantity information as well as household demographics) for calendar years 1998 through 2012.

Results and Discussion:

Our work would give us a sense of how substantive knowledge contributes to the representation of good probability forecasts. Some preliminary results are as follows. Model's ability to offer a high probability for events occurred versus low probability for events that did not occur (also known as resolution) is higher, the higher the substantive knowledge (see Figure 1 and Figure2). Also, the model with more substantive knowledge gave rise to lower Brier Score (lower the better) and higher covariance between probabilities offered and events observed (see Table 1). Overall, better sorting of probabilities was demonstrated in the model with more substantive knowledge vis-à-vis the model with less substantive knowledge. Future work will be conducted to generate Brier Score and Yates partition of Brier Score for many possible combinations of substantive knowledge.

Table 1: Substantive Knowledge Information in Logit Models and Associated Brier Score and Covariance

Model	*Substantive Knowledge Information*	*Brier Score*	*Covariance (f,d)*
1	Price	0.06388	0.0010
2	Price, Poverty	0.06382	0.0011
3	Price, Poverty, Age	0.0636	0.0015
4	Price, Poverty, Age, Employment	0.0633	0.0021
5	Price, Poverty, Age, Employment, Education	0.0632	0.0023
6	Price, Poverty, Age, Employment, Education, Region	0.0627	0.0031
7	Price, Poverty, Age, Employment, Education, Region, Race	0.0626	0.0035
8	Price, Poverty, Age, Employment, Education, Region, Race, Hispanic	0.0625	0.0036
9	Price, Poverty, Age, Employment, Education, Region, Race, Hispanic, Children	0.0621	0.0046
10	Price, Poverty, Age, Employment, Education, Region, Race, Hispanic, Children, Gender	0.0612	0.0066

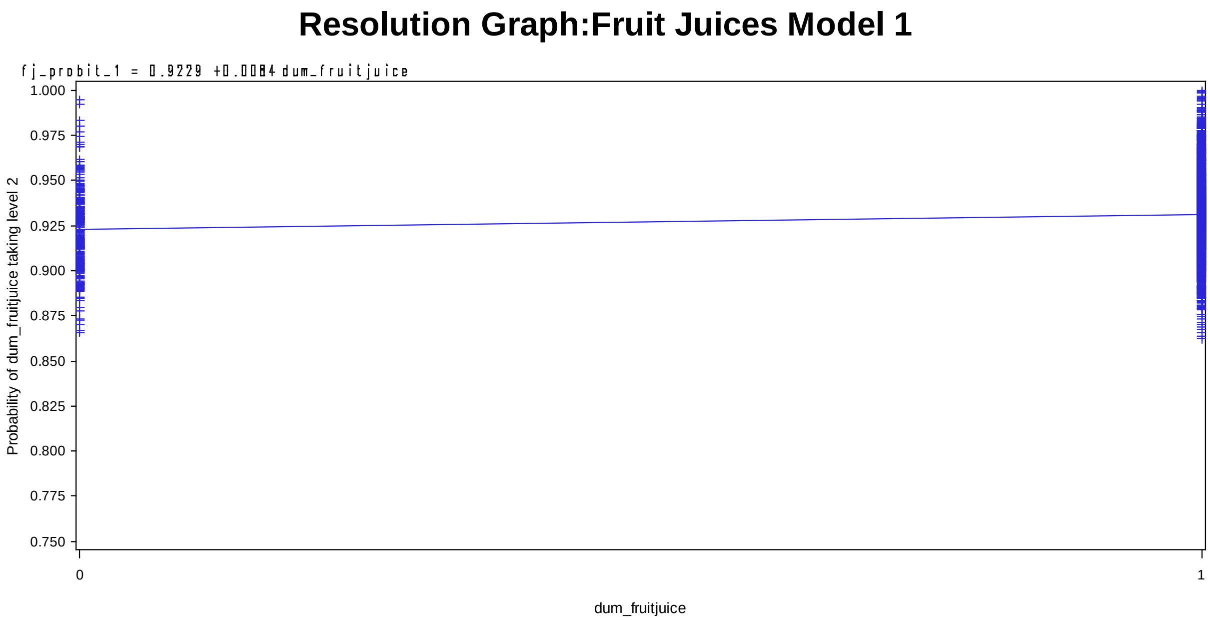

Figure 1: Resolution graph for Model 1: (substantive knowledge variable, Price)

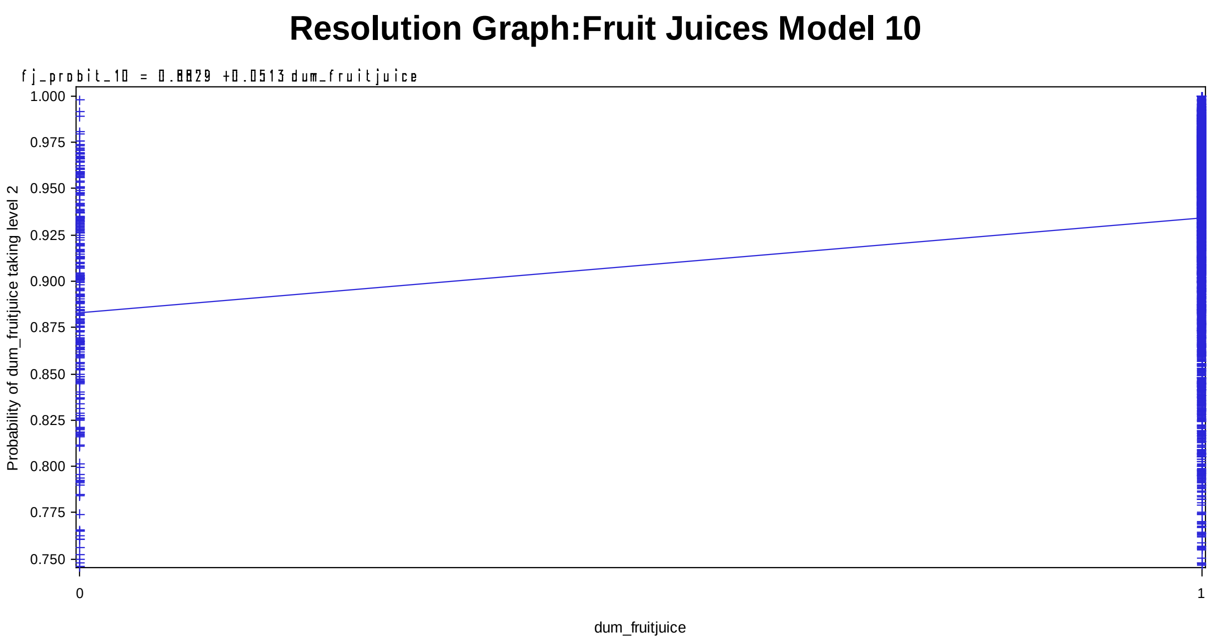

Figure 2: Resolution graph for Model 10 (substantive knowledge variables: Price, Poverty, Age, Employment, Education, Region, Race, Hispanic Status, Presence of Children, Gender of household head)

References:

Brier, G.W., January 1950. "Varification of Forecasts Expressed in terms of Probability." *Monthly Weather Review* 78(1):1-3.

Dharmasena, S., 2010. *The Non-alcoholic Beverage Market in the United States: Demand Interrelationships, Dynamics, Nutrition Issues and Probability Forecast Evaluation*, unpublished Ph.D. Dissertation, Texas A&M University, May 15.

Dharmasena, S., D.A. Bessler, and O. Capps, Jr., 2014. "On the Evaluation of Probability Forecasts: An Application to Qualitative Choice Models." Working Paper, Department of Agricultural Economics, Texas A&M University.

Yates, J.F., 1988. "Analyzing the Accuracy of Probability Judgements for Multiple Events: An Extension of the Covariance Decomposition." *Organizational Behavior and Human Performance* 41:281-299.