The Apple and Your Eye: Visual and Taste Rank-Ordered Probit Analysis with Correlated Errors *

Laura Schechter

August 10, 2009

Abstract

I look at data from an experiment in which people rank apples according to how they think they will taste. They are then blindfolded and rank how they actually taste. I estimate a multinomial rank-ordered probit model with correlated errors between the taste and visual rankings. I find that the errors for visual characteristics are correlated based on coloring, while the errors for taste are correlated based on sweetness and tartness. Allowing for correlation between the errors in the two regressions shows that, although people often misperceive apple taste based upon visual cues, they do so in systematic ways. People who prefer the looks of an apple they think to be tart (Granny Smith), will like the taste of other apples which are also tart but less well-known (Jonagold).

1 Introduction

Few people have the ability to choose a good apple from the many varieties and sizes available at supermarkets today. Often, a consumer will pick the

*I am indebted to Ethan Ligon for collecting the data and allowing me to use it and to Ken Train for his invaluable advice and help. I appreciate comments from Jennifer Alix-Garcia, Yanhong Jin, and Guanming Shi. Alex Yuskavage provided able research assistance.

Table 1: Shares of Rankings, $n = 135$

	Visual					Taste				
	1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
Granny Smith	0.24	0.24	0.19	0.27	0.07	0.27	0.15	0.27	0.16	0.16
Fuji	0.16	0.19	0.31	0.23	0.12	0.30	0.24	0.15	0.19	0.13
Golden Del.	0.10	0.27	0.27	0.22	0.14	0.08	0.19	0.17	0.25	0.31
Jonagold	0.43	0.20	0.17	0.12	0.08	0.16	0.24	0.23	0.21	0.15
Red Del.	0.07	0.10	0.07	0.17	0.59	0.19	0.19	0.18	0.19	0.25

prettiest, most colorful, unblemished apple in the hope that the apple will taste good as well. In this paper I look at rankings of 5 different varieties of apple. The data comes from an experiment carried out on Cal Day[1] in 1999. Participants were shown the apples and asked to rank them in the order of how good they thought they would taste. Then, the participants were blindfolded and ate a piece of each apple and ranked them again in the order of how good they actually tasted. Throughout this paper I will label these 'visual' and 'taste' rankings, although what I call a 'visual' ranking is actually a visual assessment of how good the respondent thinks the apple will taste; it is not an assessment of the apple's visual characteristics.

Although data on willingness-to-pay are not available, McCluskey et al. (2007) show that subjective evaluation by consumers of apple quality based on taste has higher predictive power for willingness to pay than do objective tests of apple quality such as firmness or soluble solids. Thus, subjective evaluations of apples do contain economic content. The reader should keep in mind that this is a 'simplified' experiment in that it does not re-create a real market setting. For example, visual and taste perceptions may interact with price information or knowledge of apple varietals in unknown ways which we can not capture with our data. Future experiments using the statistical methodology laid out in this paper could use a more realistic experimental set-up.

135 participants tasted Granny Smith, Golden Delicious, Jonagold, Fuji, and Red Delicious apples (although they did not actually know what variety

[1] Cal Day is a day for friends, alumni, parents, and students at UC Berkeley to enjoy demonstrations, lectures, and other activities.

Table 2: Typical apple characteristics

Variety	Color	Taste	Introduced
Granny Smith	Bright green	Very tart and crisp	1850
Fuji	Red with yellow and green highlights	Very sweet and crisp	1962
Golden Del.	Golden yellow	Mildly sweet and creamy	1900
Jonagold	Yellow with orange and red highlights	Sweet-tart and crisp	1968
Red Del.	Bright red	Very sweet and crisp	1872

each apple was). The shares of people ranking each apple first through fifth in terms of visual and taste preference are presented in Table 1. Some typical characteristics of apples of each variety are presented in Table 2. One sees that people have an aversion to the looks of the Red Delicious apple. This aversion may be caused by Red Delicious apples' reputation for having little taste. On the other hand, people are not so averse to the taste of the Red Delicious apple, suggesting that the apple's bad reputation may be partially unfounded. Conversely, the participants liked the look of the Jonagold apple, but were less impressed by the Jonagold's taste. The Fuji tasted better than expected. It is reasonable to expect that people are not able to distinguish by sight which apple will taste the best between apples of the same variety, but here one finds that people can not discern which apple they will like best, even among apples of distinct varieties. Allowing for correlations between unobserved characteristics in the taste and visual rankings will make it possible to find patterns in this mis-ranking.

Other papers which compare taste and visual evaluations include Melton et al. (1996) and Nalley et al. (2006) who conduct auctions on pork chops and sweet potatoes respectively and find little consistency in bids between the pre-tasting and post-tasting bids. Bredahl et al. (1998) and Banović et al. (2009) find that assessments of beef before and after consumption diverge.[2] Opposing evidence is found by Bello Acebrón and Calvo Dopico (2000) and Bredahl (2004) who find that assessments of pork and beef before and after consumption are largely in accord with one another. The last two differ from the others in that they are ex-post assessments of meats which the consumer himself chose to purchase rather than items chosen by the experimenter.

[2] Poole et al. (2007) find that bids in auctions for oranges change after first seeing the oranges whole, then peeling them, and again after tasting them.

Ranked data contains much more information than do surveys which ask for only the preferred alternative. Thus ranked data surveys are extremely cost-effective since fewer observations are necessary for a given level of precision. On the other hand, some economists have pointed out that top ranked choices may be ranked with more precision than bottom ranked choices (Hausman and Ruud, 1987; Ben-Akiva et al., 1992). Consumers are daily confronted with situations in which they choose their first choice, but are not often asked to rank the remaining alternatives, possibly resulting in decreasing precision of ranked choices.

Carson et al. (1994) discuss fatigue effects (boredom with the choice tasks) as well as the use of hypothetical experimental choice sets as causes of decreasing precision in ranks. Because our data are on taste and visual rankings of real apples, which the subject can see and enjoy tasting, (as opposed to varying hypothetical mobile phone attributes) these issues should be attenuated here. In addition, although, from a theoretical perspective, the number of choices presented to respondents should not affect the reliability of the data; in practice, too many choices may add to respondent fatigue and make the ranking exercise more difficult as the options become more similar. The literature does not tell us how the number of choices affects the reliability of ranked data, but a choice between five apples is not excessive compared to other experiments using ranked data. Recent evidence from Caparrós et al. (2008) suggests that ranking experiments which only use data regarding the favorite choice and preferred choice experiments give similar outcomes, and that the divergence found in previous studies is largely due to differences in experimental design rather than the participants' cognitive processes.

This paper uses rank-ordered probit to analyze ranked data while most previous papers have used a rank-ordered logit model (Beggs et al., 1981; Ben-Akiva et al., 1992; Bradlow and Fader, 2001; Hausman and Ruud, 1987). These papers look at rankings of transportation, car phones, and even the *Billboard* top 100 list. The logit specification exhibits the Independence from Irrelevant Alternatives (IIA) assumption. The IIA assumption makes evaluating a rank-ordered logit model quite simple. One can "explode" each individual's ranking into a series of independent choice situations. First the individual picks his first choice. Then, with that alternative eliminated, the participant picks his second choice, given the remaining alternatives. In this way ranked data can be decomposed into a series of statistically independent choices.

IIA is a very convenient property, but the maintained assumption of IIA may be quite restrictive. IIA implies that the ratio of the probabilities of any two alternatives is constant, no matter what other alternatives are presented to the individual. The rank-ordered probit model, on the other hand, allows each alternative to have a random component with a complete variance-covariance structure.

More importantly, rankings from best to worst are not compatible with ranking from worst to best under the multinomial logit specification. Luce and Suppes (1965) call this the "impossibility theorem" (Theorem 51 on page 357). For an individual's ranking from best to worst to be compatible with the individual's ranking from worst to best, the probability of each alternative must be $\frac{1}{N}$ (where N is the number of alternatives). The IIA property of rank-ordered logit is the critical piece which causes this incompatibility. Rank-ordered probit does not suffer from Luce and Suppes' impossibility theorem. Thus, if one thinks that people may not necessarily think through their decisions from best to worst,[3] then the rank-ordered probit specification is preferable to that of rank-ordered logit.

Yao and Böckenholt (1999) use Gibbs sampling in a Bayesian framework to estimate a rank-ordered probit model. Chan and Bentler (1998) estimate a rank-ordered probit model with a full variance/covariance matrix using GLS and partition maximum likelihood. GLS and partition maximum likelihood are limited information estimators and are consistent, but not efficient. Maximum likelihood is a full information procedure, and maximum simulated likelihood is efficient when the number of draws rises faster than the square root of the number of observations (i.e. respondents) (Hajivassiliou and Ruud, 1994). Hajivassiliou and Ruud (1994) use maximum likelihood to estimate a rank-ordered probit model with a complete variance/covariance matrix with Monte Carlo data as an example. Riddel and Schwer (2006) is the only paper thusfar to estimate a rank-ordered probit using maximum likelihood with real data.[4]

The contribution of the current paper is that I estimate a rank-ordered probit model using maximum likelihood over multiple characteristics of apples with errors correlated between the two regressions. By using rank-ordered probit, I can efficiently relax the IID assumption which has hampered previous work using ranked data on both single and multiple characteristics.

[3]Perhaps they choose their most and least favorite alternatives first, and then figure out the rankings in between.

[4]The authors do not directly state that they use maximum likelihood, but one can infer that they probably do.

In addition, I show how one can allow the errors between the multiple rankings (in this case, taste and visual) to be correlated with one another in the rank-ordered probit framework. This technique could also be useful when combining stated and revealed preference. Ben-Akiva and Morikawa (1990) and Adamowicz et al. (1994) carry out a logit model which maintains the IID assumption within the revealed preference grouping and within the stated preference grouping, but allows for a correlation between the two.

Allowing a correlation between taste and visual rankings tells us, in this case, that people who like the taste of the Jonagold (a newer and possibly less well-known tart apple) often incorrectly think they will prefer the Granny Smith apple when they rank the apples visually. On the other hand, people who prefer the Jonagold visually are often mistaken, preferring the taste of the sweeter apples and perhaps not realizing by look that the Jonagold is a tart apple.[5] This means that, although people mistakenly think when looking at apples that they will prefer the taste of apples they don't actually prefer, there is a systematic basis to these errors. This type of analysis can be applied to combine any data with rankings over multiple attributes of the same good or to combine stated and revealed preference, in order to design products with characteristic sets well-attuned to the desires of specific sectors of the consumer population.

The rest of this paper is organized as follows: Section 2 presents the choice model and simulation methods used in the analysis, Section 3 presents the results assuming no correlation between the taste and visual rankings while Section 4 presents the results allowing correlation between errors in the taste and visual regressions. Section 5 gives some concluding remarks.

2 Econometric Model

The probit model specifies utility with an observed and an unobserved component. Let the utility that individual i receives from alternative n be denoted as

$$U_{in} = V_{in} + e_{in}$$

[5]This finding is similar to that of Thybo et al. (2004) who find that children who claim to prefer green apples do actually prefer them in taste tests, while children who claim to prefer red apples do as well. The red apple in their experiment is sweet while the green apple is tart, which implies that children are at least able to predict whether they will like tart or sweet apples.

where V is the observed portion and e the unobserved portion of utility. The vector $\widetilde{e}_n = \{e_{in}\}_n$ of error terms is normally distributed with a mean of 0 and variance-covariance matrix Ω. In cases in which there is only information on the individual's preferred alternative, one calculates the probability that the utility associated with the chosen alternative is higher than the utility associated with all other alternatives. If there are J alternatives, there will be $J-1$ non-redundant inequalities. One would then calculate their joint probability.

$$P_{in} = \text{Prob}(V_{in} + e_{in} > V_{ij} + e_{ij}) \; \forall j \in J, j \neq n$$

In the rank-ordered case, one only has to consider the $J-1$ non-redundant inequalities for each individual inspired by their ranking. For example, if there are 5 choices which the individual ranks in the order 3, 2, 1, 4, 5, there will be 4 inequalities.

$$\begin{aligned} P_{in} = \text{Prob}(\quad & V_{i3} + e_{i3} > V_{i2} + e_{i2} \\ & \& \; V_{i2} + e_{i2} > V_{i1} + e_{i1} \\ & \& \; V_{i1} + e_{i1} > V_{i4} + e_{i4}) \\ & \& \; V_{i4} + e_{i4} > V_{i5} + e_{i5}) \end{aligned}$$

This likelihood function can be written more formally as

$$P_{in} = \int_{e_{i3}=V_{i2}+e_{i2}-V_{i3}}^{\infty} \int_{e_{i2}=V_{i1}+e_{i1}-V_{i2}}^{e_{i3}} \int_{e_{i1}=V_{i4}+e_{i4}-V_{i1}}^{e_{i2}} \int_{e_{i4}=V_{i5}+e_{i5}-V_{i4}}^{e_{i1}} \phi(\widetilde{e}_{in}) de_{i4} de_{i1} de_{i2} de_{i3}$$

where $\phi(\widetilde{e}_{in})$ is the joint normal density function.

Using maximum likelihood estimation, I maximize the sum over individuals of the logs of each likelihood function to find the parameters of the model. Because calculating the closed form value of this integral directly is not possible, I approximate the integral with simulations. Hajivassiliou et al. (1996) conducted a survey of 11 different Monte Carlo techniques used for simulation of the multinomial probit likelihood. They find that the Geweke-Hajivassiliou-Keane (GHK) simulator is the most reliable of all the simulators, especially for simulating probabilities. The GHK simulator takes samples from recursive truncated normals after a Choleski transformation. Let U, V, and e be column vectors of U_{in}, V_{in}, and e_{in} respectively Then the utility functions can be expressed as $U = V + e$.

In implementing the GHK simulator with rank-ordered data, one premultiplies the U and V, matrices by a transformation matrix M and transforms Ω by $M\Omega M'$. This matrix allows one to translate the original model to the model with only $J-1$ equations. The first row of this matrix will contain all 0's except for a -1 in the column of the individual's first choice and a 1 in the column of his second choice. The second row will have a -1 in the column of the individual's second choice and a 1 in the column of his third choice. For example, with our previous ranking (3,2,1,4,5), one would use the following transformation matrix.

$$M_r = \begin{vmatrix} 0 & 1 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{vmatrix}$$

After transforming the utility functions to utility differences one has $M_rU = M_rV + M_re$, where M_r represents the transformation matrix when ranking order r is chosen. Let C_r be the Choleski factor of $M_r\Omega M_r'$. Then the utility differences can be rewritten as $M_rU = M_rV + C_r\eta$ (where η is a vector of standard normal deviates). The GHK simulator will randomly choose a value for η_{i1} for individual i and alternative 1, which satisfies the first inequality. Then, given that value of η_{i1}, the simulator will randomly choose a value of η_{i2} which satisfies the second inequality and so on.

The simulation process can often be very slow because a large number of random draws is needed for precision in estimation. Random draws are not as efficient as some types of 'intelligent' draws. Quasi-random ('intelligent') draws exhibit more uniform coverage of the space over which one randomizes. Train (2000) and Bhat (2001) have researched the use of draws from Halton sequences in simulation procedures. They have found that simulation variance is lower with only 100 quasi-random Halton draws than it is with 1000 more traditional random draws and so in this paper I also use 100 Halton draws.

3 Results - Uncorrelated Errors between Taste and Visual Regressions

To analyze the visual and taste rankings over apples I first use a rank-ordered multinomial probit model with uncorrelated errors between the two sets of rankings. I compute the probability integral using a GHK simulator with 100 Halton quasi-random draws. I use four alternative specific constants and normalize the coefficient on Granny Smith to 0.[6]

Table 3: General variance/covariance matrix

0.000	0.000	0.000	0.000	0.000
0.000	1.000	$\frac{s_{11}+s_{23}-s_{13}-s_{12}}{s_{11}+s_{22}-2s_{12}}$	$\frac{s_{11}+s_{24}-s_{14}-s_{12}}{s_{11}+s_{22}-2s_{12}}$	$\frac{s_{11}+s_{25}-s_{15}-s_{12}}{s_{11}+s_{22}-2s_{12}}$
0.000	$\frac{s_{11}+s_{23}-s_{13}-s_{12}}{s_{11}+s_{22}-2s_{12}}$	$\frac{s_{11}+s_{33}-2s_{13}}{s_{11}+s_{22}-2s_{12}}$	$\frac{s_{11}+s_{34}-s_{14}-s_{13}}{s_{11}+s_{22}-2s_{12}}$	$\frac{s_{11}+s_{35}-s_{15}-s_{13}}{s_{11}+s_{22}-2s_{12}}$
0.000	$\frac{s_{11}+s_{24}-s_{14}-s_{12}}{s_{11}+s_{22}-2s_{12}}$	$\frac{s_{11}+s_{43}-s_{14}-s_{13}}{s_{11}+s_{22}-2s_{12}}$	$\frac{s_{11}+s_{44}-2s_{14}}{s_{11}+s_{22}-2s_{12}}$	$\frac{s_{11}+s_{45}-s_{15}-s_{14}}{s_{11}+s_{22}-2s_{12}}$
0.000	$\frac{s_{11}+s_{25}-s_{15}-s_{12}}{s_{11}+s_{22}-2s_{12}}$	$\frac{s_{11}+s_{53}-s_{15}-s_{13}}{s_{11}+s_{22}-2s_{12}}$	$\frac{s_{11}+s_{54}-s_{14}-s_{15}}{s_{11}+s_{22}-2s_{12}}$	$\frac{s_{11}+s_{55}-2s_{15}}{s_{11}+s_{22}-2s_{12}}$

Interpreting the variance-covariance matrices for the error terms is quite difficult. In order for the model to be identified, one must normalize the first row and column to contain only zeros, and the entry in the second row and second column to be 1. Define s_{ij} to be the correlation between the unobserved characteristics of the i^{th} and j^{th} apples. Table 3 presents the formula for the entries in the variance-covariance matrix after normalization. Note that in the case of homoskedasticity and zero correlation between apples, the matrix would have 1's on the diagonal and .5's in all of the other positions (except for the first row and column of 0's).[7]

Table 4 reports the estimates of the normalized alternative-specific constants using Granny Smith as a base.[8] Although the Red Delicious apple has a significant negative coefficient in the visual regression, its coefficient is almost 0 in the taste regression. Many people may have had such bad experiences with Red Delicious apples that they no longer think that a Red Delicious would taste good when they see it and so predict it will yield them lower utility than Granny Smith. The Jonagold looks the best to everyone, although people don't seem to be especially keen on its taste. The Fuji apple

[6] Note that the normalization of an alternative-specific constant to 0 is common in all discrete choice analysis and is true of estimation techniques that both do and do not assume IIA.

[7] If there is no correlation between apples then $s_{12} = s_{13} = s_{14} = s_{15} = s_{23} = s_{24} = s_{25} = s_{34} = s_{35} = 0$ and if there is homoskedasticity then $s_{11} = s_{22} = s_{33} = s_{44} = s_{55}$. So, for example, the entry in the third row and second column of table 3 is $\frac{s_{11}+s_{23}-s_{13}-s_{12}}{s_{11}+s_{22}-2s_{12}}$. If there is no correlation between apples this simplifies to $\frac{s_{11}}{s_{11}+s_{22}}$ and if there is also homoskedasticity this further simplifies to $\frac{1}{2}$.

[8] In this case, results using multinomial logit give quite similar coefficients. Thus, the "impossibility theorem" may not always be important in practice.

Table 4: Multinomial rank-ordered probit results, n=135 - No correlation between taste and visual

Variable	Visual	Taste
Fuji	-0.1192	0.1087
	(0.0975)	(0.0968)
Golden Delicious	-0.1628^{**}	-0.3409^{***}
	(0.0797)	(0.0982)
Jonagold	0.3022^{***}	-0.0607
	(0.1120)	(0.0784)
Red Delicious	-0.9153^{***}	-0.1298
	(0.1532)	(0.1039)
Log-Likelihood	-572.9	-625.9

Granny Smith coefficient normalized to 0. Numbers in parenthesis are standard errors.
*-90%, **-95%, and ***-99% significant.

tastes the best, while the Golden Delicious tastes the worst.

Tables 5 and 6 contain the variance/covariance matrix for visual and taste unobservable characteristics. In order to interpret these covariance matrices more easily, one can look to Table 7. In that table I forecast the probability that each apple is chosen as the best apple both with all five apples, and after taking out any one apple from the choice set.[9] Each column sums to 100%, with a bullet point representing the omitted apple.

Remember that in the case of zero correlation between apples and homoskedasticity, all off-diagonals (other than the first row and column) would be 0.5. The high number, 0.681, in the (2,4) position of Table 5 is caused by the high correlation between the Fuji and the Jonagold in unobserved visual characteristics. This high correlation is to be expected, as the apples have similar colorings (both are a mixture of red, green, and yellow). When one looks at the substitution patterns in Table 7, one finds that after taking Jonagold out of the choice set, the probit predicts that almost 6 percentage points more of the people switch to preferring Fuji than would have under the assumption of IIA.

The low number 0.310 in the (3,5) position of Table 5 would lead one

[9]To forecast these probabilities one must use the preferred-choice transformation matrix rather than the rank-ordered transformation matrix (M_r). This is because I forecast preferred choice, and not permutations of rankings.

Table 5: Variance/covariance matrix for visual characteristics

	Gr. Sm.	Fuji	Go. Del.	Jona.	Red Del.
Gr. Sm.	0.000	0.000	0.000	0.000	0.000
Fuji	0.000	1.000	0.447	0.681	0.497
Go. Del.	0.000	0.447	0.622	0.350	0.310
Jona.	0.000	0.681	0.350	1.187	0.343
Red Del.	0.000	0.497	0.310	0.343	1.411

Table 6: Variance/covariance matrix for taste characteristics

	Gr. Sm.	Fuji	Go. Del.	Jona.	Red Del.
Gr. Sm.	0.000	0.000	0.000	0.000	0.000
Fuji	0.000	1.000	0.596	0.363	0.571
Go. Del.	0.000	0.596	0.912	0.406	0.527
Jona.	0.000	0.363	0.406	0.639	0.475
Red Del.	0.000	0.571	0.527	0.475	1.169

to believe that the correlation between Red Delicious and Golden Delicious was very low, but the substitution patterns in Table 7 tell a different story. The substitution patterns in Table 7 show what share of respondents would be predicted to choose each apple as their first choice if one of the five apples were omitted from the choice set. So, looking at the column labeled Granny Smith, we see that if Granny Smith had not been one of the options then 20.92% of respondents are predicted to have preferred the looks of the Golden Delicious apple rather than the 12.9% who are predicted to have preferred the looks of Golden Delicious when Granny Smith was also an option. The number 0.0422 in parentheses tells us the difference between the share of respondents predicted to prefer Golden Delicious under the rank-ordered probit assumption and the share predicted under the IIA assumption. So, we see that this prediction of 21% is 4 percentage points higher than the prediction we would have estimated (17%) under IIA. Thus, the low number 0.310 in Table 5 arises because the visual correlation between Granny Smith (the omitted apple) and Golden Delicious are high. These apples are green and yellow respectively, and are the only two apples with no red coloring in the choice set. Thus, the apples with low visual correlations are of different

Table 7: Substitution patterns - No correlation between taste and visual

			Gr. Sm.	Fuji	Go. Del.	Jona.	Red Del.
Visual	Actual	Forecast	Forecast After Omission				
Gr. Sm.	0.2370	0.2274	•	0.2560	0.2790	0.3673	0.2519
				(-0.0090)	(0.0179)	(-0.0273)	(0.0052)
Fuji	0.1556	0.1418	0.1792	•	0.1724	0.3063	0.1589
			(-0.0043)		(0.0096)	(0.0602)	(0.0051)
Go. Del.	0.1037	0.1290	0.2092	0.1624	•	0.2152	0.1432
			(0.0422)	(0.0121)		(-0.0086)	(0.0033)
Jona.	0.4296	0.4237	0.5088	0.4906	0.4608	•	0.4460
			(-0.0396)	(-0.0031)	(-0.0257)		(-0.0136)
Red Del.	0.0741	0.0781	0.1031	0.0912	0.0879	0.1112	•
			(0.0020)	(0.0002)	(-0.0018)	(-0.0243)	
Taste	Actual	Forecast	Forecast After Omission				
Gr. Sm.	0.2667	0.2398	•	0.3262	0.2598	0.3007	0.2866
				(-0.0157)	(-0.0038)	(0.0133)	(-0.0153)
Fuji	0.2963	0.2987	0.3742	•	0.3328	0.3394	0.3716
			(-0.0187)		(0.0045)	(-0.0186)	(-0.0044)
Go. Del.	0.0815	0.0902	0.1219	0.1586	•	0.1138	0.1164
			(0.0033)	(0.0300)		(0.0057)	(0.0028)
Jona.	0.1630	0.1657	0.2540	0.2296	0.1859	•	0.2254
			(0.0360)	(-0.0067)	(0.0038)		(0.0168)
Red Del.	0.1926	0.2057	0.2502	0.2858	0.2216	0.2462	•
			(-0.0204)	(-0.0075)	(-0.0045)	(-0.0003)	

Each column sums to 100. The last five columns represent the omission of a different apple from the choice set. Numbers in parentheses are the difference between shares predicted by rank-ordered probit and shares under IIA. For example, if one omitted Granny Smith, then 51% of people are predicted to prefer the looks of Jonagold and this is 4% less than IIA would have predicted.

colors while those with high correlations are of more similar coloring.

In Table 6, the high .596 in the (2,3) position shows a high correlation in the unobserved taste characteristics between Fuji and Golden Delicious, and the high 1.169 in the (5,5) position show a low correlation between the taste characteristics of the Granny Smith and the Red Delicious apples. Both of these results are born out in the substitution patterns. When Fuji is removed almost 3 percentage points more of the people switch to Golden Delicious than would have, and when Granny Smith is removed almost 2 percentage points less switch to Red Delicious than would have under IIA.

For the marketing of apples in stores which carry different varieties of apples, these correlations are extremely important, and could not be discovered while using a simple logit specification given the IIA assumption. When choosing an apple based on looks, consumers tend to choose based on color. On the other hand, when they have the opportunity to taste the apples, they tend to differentiate on the basis of sweetness versus tartness.

In Table 8, I estimate the visual and taste characteristic regressions (under the assumption that the taste and visual unobserved characteristics are uncorrelated). Each column in this table sums to 100%, but this table now includes much more detail than did Table 7 since there are 25 different combinations of visual and taste first choices possible.

Previously I stated that after taking Jonagold out of the choice set, almost 6 percentage points more people are predicted to switch to Fuji as their visual first choice than would have under the assumption of IIA. The second to last column in Table 8 forecasts what will happen when removing Jonagold from the choice set. Of the 6 percentage points more people who switch to Fuji as their visual first choice, only 26% preferred the taste of Fuji. Most (36%) preferred the taste of Granny Smith. While Fuji and Jonagold are similarly colored apples, Fuji is sweet while Jonagold is tart. Thus, taking Jonagolds out of the choice set often induces people to mistakenly predict that the Fuji will taste better based on visual characteristics, when these people actually prefer the taste of the Granny Smith. Similarly, people who like the taste of the Granny Smith are more likely to switch to liking the looks of the Fuji, perhaps thinking that it looks like it would be a tart apple.

4 Results - Correlated Errors between Taste and Visual Regressions

This analysis can be taken one step further by allowing the errors to be correlated between the taste and visual ranking regressions. Rather than estimating 9 of the entries in the symmetric 5x5 correlation matrix shown in table 3 as before, we must now estimate 34 of the entries in the symmetric 10x10 correlation matrix (9 of the entries for taste rankings, 9 of the entries for visual rankings, and 16 of the entries for the correlation between taste and visual characteristics). It would be possible to carry out such an analysis with a third characteristic (e.g., smell) but we would then have to estimate 75 of the entries in the symmetric 15x15 correlation matrix (9 of the entries for taste rankings, 9 of the entries for visual rankings, 9 of the entries for

Table 8: Substitution patterns - No correlation between taste and visual

				Gr. Sm.	Fuji	Go. Del.	Jona.	Red Del.
Visual	Taste	Actual	Forecast	Forecast After Omission				
Granny Smith	Gr. Sm.	0.0963	0.0545	•	0.0835 (-0.0071)	0.0725 (0.0037)	0.1104 (-0.0030)	0.0722 (-0.0023)
	Fuji	0.0296	0.0679	•	•	0.0929 (0.0071)	0.1246 (-0.0166)	0.0936 (0.0008)
	Go. Del.	0.0222	0.0205	•	0.0406 (0.0065)	•	0.0418 (-0.0009)	0.0293 (0.0013)
	Jona.	0.0593	0.0377	•	0.0588 (-0.0038)	0.0519 (0.0043)	•	0.0568 (0.0053)
	Red Del.	0.0296	0.0468	•	0.0731 (-0.0046)	0.0618 (0.0028)	0.0904 (-0.0069)	•
Fuji	Gr. Sm.	0.0296	0.0340	•	•	0.0448 (0.0019)	0.0921 (0.0214)	0.0456 (-0.0009)
	Fuji	0.0963	0.0424	0.0670 (-0.0051)	•	0.0574 (0.0039)	0.1040 (0.0159)	0.0591 (0.0012)
	Go. Del.	0.0000	0.0128	0.0218 (0.0000)	•	•	0.0349 (0.0082)	0.0185 (0.0010)
	Jona.	0.0074	0.0235	0.0455 (0.0055)	•	0.0320 (0.0024)	•	0.0358 (0.0037)
	Red Del.	0.0222	0.0292	0.0448 (-0.0048)	•	0.0382 (0.0014)	0.0754 (0.0148)	•
Golden Delicious	Gr. Sm.	0.0296	0.0309	•	0.0530 (0.0016)	•	0.0647 (0.0004)	0.0410 (-0.0012)
	Fuji	0.0370	0.0385	0.0783 (0.0126)	•	•	0.0730 (-0.0071)	0.0532 (0.0006)
	Go. Del.	0.0074	0.0116	0.0255 (0.0057)	0.0258 (0.0064)	•	0.0245 (0.0003)	0.0167 (0.0008)
	Jona.	0.0222	0.0214	0.0531 (0.0167)	0.0373 (0.0018)	•	•	0.0323 (0.0031)
	Red Del.	0.0074	0.0265	0.0523 (0.0071)	0.0464 (0.0023)	•	0.0530 (-0.0022)	•
Jonagold	Gr. Sm.	0.0889	0.1016	•	0.1600 (-0.0088)	0.1197 (-0.0085)	•	0.1278 (-0.0109)
	Fuji	0.1259	0.1265	0.1903 (-0.0251)	•	0.1533 (-0.0063)	•	0.1657 (-0.0071)
	Go. Del.	0.0519	0.0382	0.0620 (-0.0031)	0.0778 (0.0143)	•	•	0.0519 (-0.0003)
	Jona.	0.0593	0.0702	0.1292 (0.0097)	0.1126 (-0.0040)	0.0857 (-0.0029)	•	0.1005 (0.0047)
	Red Del.	0.1037	0.0871	0.1273 (-0.0211)	0.1402 (-0.0046)	0.1021 (-0.0079)	•	•
Red Delicious	Gr. Sm.	0.0222	0.0187	•	0.0297 (-0.0014)	0.0228 (-0.0008)	0.0334 (-0.0055)	•
	Fuji	0.0074	0.0233	0.0386 (-0.0011)	•	0.0293 (-0.0002)	0.0377 (-0.0108)	•
	Go. Del.	0.0000	0.0070	0.0126 (0.0006)	0.0145 (0.0028)	•	0.0127 (-0.0020)	•
	Jona.	0.0148	0.0129	0.0262 (0.0042)	0.0209 (-0.0006)	0.0164 (0.0000)	•	•
	Red Del.	0.0296	0.0161	0.0258 (-0.0016)	0.0261 (-0.0006)	0.0195 (-0.0008)	0.0274 (-0.0060)	•

Each column sums to 100. The last five columns represent the omission of a different apple from the choice set. Numbers in parentheses are the difference between shares predicted by rank-ordered probit and shares under IIA. For example, if one omitted Granny Smith, then 13% of people are predicted to prefer the looks and taste of Jonagold and this is 1% more than IIA would have predicted.

Table 9: Multinomial rank-ordered probit results, n=135 - Correlated errors between taste and visual

Variable	Visual	Taste
Fuji	-0.1399	0.0790
	(0.1072)	(0.1021)
Golden Delicious	-0.1705^{*}	-0.3511^{***}
	(0.0892)	(0.1024)
Jonagold	0.2888^{**}	-0.0756
	(0.1187)	(0.0890)
Red Delicious	-0.9367^{***}	-0.1591
	(0.2044)	(0.1161)
Log-Likelihood	-1184.3	

The coefficients for the Granny Smith apple has been normalized to 0. Numbers in parenthesis are standard errors, *-90%, **-95%, and ***-99% significant.

smell rankings, 16 of the entries for the correlation between taste and visual, 16 of the entries for the correlation between taste and smell, and 16 of the entries for the correlation between visual and smell). One would thus need data from many respondents to be able to estimate models with three characteristics, and data requirements for models with even more characteristics would become yet more onerous.

Table 9 shows the coefficients from this new regression which, as expected, are quite similar to those shown in Table 4 when assuming the errors in the taste and visual regressions were not correlated. What changes more than the coefficients on the apples, are the substitution patterns between the apples. Table 10 is the counterpart to Table 7 after allowing the errors between the taste and visual regressions to be correlated. We see differences between the two tables in forecasts. If one omits Jonagold, more people switch to Granny Smith and less to Fuji as both their first visual choice and their first taste choice. Although the Jonagold and Fuji may look similar, the Jonagold is more similar in flavor to the Granny Smith. Granny Smith and Jonagold are both tart apples, while the other three are sweeter apples. Thus, the similar coloring of the Fuji and Jonagold apples masks two very different apples. Since the respondents were not told the varieties of the apples they were looking at, it is not clear how much of this confusion would have been cleared up by informing participants of the name of the variety, and how

Table 10: Substitution patterns - Correlated errors between taste and visual

			Gr. Sm.	Fuji	Go. Del.	Jona.	Red Del.
Visual	Actual	Forecast	Forecast After Omission				
Gr. Sm.	0.2370	0.2311	•	0.2589	0.2822	0.3742	0.2564
				(-0.0097)	(0.0170)	(-0.0245)	(0.0051)
Fuji	0.1556	0.1397	0.1763	•	0.1699	0.2985	0.1576
			(-0.0054)		(0.0096)	(0.0575)	(0.0057)
Go. Del.	0.1037	0.1285	0.2088	0.1615	•	0.2150	0.1429
			(0.0417)	(0.0121)		(-0.0067)	(0.0032)
Jona.	0.4296	0.4204	0.5095	0.4862	0.4580	•	0.4432
			(-0.0372)	(-0.0025)	(-0.0244)		(-0.0140)
Red Del.	0.0741	0.0804	0.1057	0.0937	0.0901	0.1124	•
			(0.0011)	(0.0002)	(-0.0022)	(-0.0263)	
Taste	Actual	Forecast	Forecast After Omission				
Gr. Sm.	0.2667	0.2504	•	0.3363	0.2715	0.3093	0.2985
				(-0.0151)	(-0.0038)	(0.0098)	(-0.0175)
Fuji	0.2963	0.2875	0.3677	•	0.3207	0.3289	0.3580
			(-0.0158)		(0.0046)	(-0.0150)	(-0.0048)
Go. Del.	0.0815	0.0906	0.1258	0.1554	•	0.1140	0.1184
			(0.0049)	(0.0282)		(0.0056)	(0.0041)
Jona.	0.1630	0.1640	0.2531	0.2280	0.1841	•	0.2252
			(0.0343)	(-0.0022)	(0.0038)		(0.0182)
Red Del.	0.1926	0.2076	0.2537	0.2806	0.2239	0.2478	•
			(-0.0232)	(-0.0108)	(-0.0044)	(-0.0005)	

Each column sums to 100. The last five columns represent the omission of a different apple from the choice set. Numbers in parentheses are the difference between shares predicted by rank-ordered probit and shares under IIA.

much could only be cleared up vis-a-vis a taste test or the availability of more detailed information about apple characteristics.

Table 11 is likewise the counterpart to Table 8 after allowing the errors between the taste and visual regressions to be correlated. More people are predicted to accurately forecast from looking at the apples that they will prefer the taste of Granny Smith, Fuji, Golden Delicious, or Red Delicious than when ignoring the correlation between taste and visual rankings. On the other hand, more people who like the looks of Jonagold are predicted to actually like the taste of Fuji. More people who like the taste of Jonagold are predicted to like the looks of Jonagold and Granny Smith (the two tart apples). Also, more people who think they will like the Golden Delicious apple when they look at it, are predicted to actually prefer the taste of the Fuji, perhaps not recognizing this newer sweet variety. Recognizing that the

Table 11: Substitution patterns - Correlated errors between taste and visual

				Gr. Sm.	Fuji	Go. Del.	Jona.	Red Del.
Visual	Taste	Actual	Forecast	Forecast After Omission				
Granny Smith	Gr. Sm.	0.0963	0.0918	•	0.1160 (-0.0313)	0.1135 (-0.0016)	0.1687 (-0.0259)	0.1157 (-0.0094)
	Fuji	0.0296	0.0401	•	•	0.0585 (0.0082)	0.0886 (0.0035)	0.0554 (0.0007)
	Go. Del.	0.0222	0.0114	•	0.0192 (0.0009)	•	0.0281 (0.0040)	0.0176 (0.0021)
	Jona.	0.0593	0.0455	•	0.0646 (-0.0084)	0.0582 (0.0011)	•	0.0677 (0.0057)
	Red Del.	0.0296	0.0422	•	0.0591 (-0.0087)	0.0519 (-0.0010)	0.0887 (-0.0009)	•
Fuji	Gr. Sm.	0.0296	0.0177	•	•	0.0248 (0.0026)	0.0527 (0.0152)	0.0250 (0.0010)
	Fuji	0.0963	0.0504	0.0669 (-0.0156)	•	0.0691 (0.0060)	0.1274 (0.0206)	0.0693 (0.0007)
	Go. Del.	0.0000	0.0123	0.0170 (-0.0032)	•	•	0.0357 (0.0095)	0.0192 (0.0024)
	Jona.	0.0074	0.0276	0.0480 (0.0028)	•	0.0367 (0.0021)	•	0.0440 (0.0065)
	Red Del.	0.0222	0.0318	0.0444 (-0.0076)	•	0.0393 (-0.0005)	0.0828 (0.0154)	•
Golden Delicious	Gr. Sm.	0.0296	0.0308	•	0.0507 (0.0013)	•	0.0585 (-0.0068)	0.0384 (-0.0036)
	Fuji	0.0370	0.0424	0.0777 (0.0082)	•	•	0.0831 (-0.0069)	0.0537 (-0.0041)
	Go. Del.	0.0074	0.0167	0.0313 (0.0040)	0.0354 (0.0086)	•	0.0340 (-0.0014)	0.0228 (0.0001)
	Jona.	0.0222	0.0195	0.0589 (0.0269)	0.0384 (0.0071)	•	•	0.0280 (0.0014)
	Red Del.	0.0074	0.0191	0.0409 (0.0096)	0.0369 (0.0063)	•	0.0394 (-0.0010)	•
Jonagold	Gr. Sm.	0.0889	0.0932	•	0.1463 (-0.0032)	0.1120 (-0.0049)	•	0.1193 (-0.0076)
	Fuji	0.1259	0.1375	0.1966 (-0.0286)	•	0.1700 (-0.0024)	•	0.1796 (-0.0076)
	Go. Del.	0.0519	0.0414	0.0635 (-0.0042)	0.0857 (0.0194)	•	•	0.0587 (0.0024)
	Jona.	0.0593	0.0558	0.1145 (0.0230)	0.1014 (0.0118)	0.0695 (-0.0005)	•	0.0855 (0.0094)
	Red Del.	0.1037	0.0925	0.1349 (-0.0167)	0.1528 (0.0043)	0.1066 (-0.0094)	•	•
Red Delicious	Gr. Sm.	0.0222	0.0169	•	0.0233 (-0.0039)	0.0212 (0.0000)	0.0294 (-0.0065)	•
	Fuji	0.0074	0.0171	0.0265 (-0.0015)	•	0.0232 (0.0017)	0.0299 (-0.0064)	•
	Go. Del.	0.0000	0.0089	0.0139 (-0.0006)	0.0151 (0.0009)	•	0.0162 (-0.0026)	•
	Jona.	0.0148	0.0155	0.0318 (0.0063)	0.0235 (-0.0014)	0.0197 (0.0002)	•	•
	Red Del.	0.0296	0.0220	0.0334 (-0.0026)	0.0318 (-0.0035)	0.0260 (-0.0016)	0.0370 (-0.0097)	•

Each column sums to 100. The last five columns represent the omission of a different apple from the choice set. Numbers in parentheses are the difference between shares predicted by rank-ordered probit and shares under IIA.

errors in the taste and visual regressions may be correlated makes all of these effects stronger.

We can also look at the predicted effects of omitting one variety on both looks and taste ranking. If one takes out Granny Smith as a choice, more people who prefer the looks of Fuji or Golden Delicious are predicted to prefer the taste of Jonagold in comparison with the analysis ignoring correlations. This suggests that people who like the tart taste of Jonagold do not always realize that they will like it from looking at it. Likewise, if one takes out the Jonagold, more people will like the looks and taste of Granny Smith than did before taking into account the correlation.

5 Conclusions

This paper presents a new way of analyzing data from multiple rankings. It applies the rank-ordered probit model with a GHK simulator and Halton draws to data with visual and taste tests of apples allowing the errors in the two regressions to have arbitrary correlation. The rank-ordered probit is not difficult to implement and with the use of Halton draws its implementation is quite computationally efficient.

This manner of analyzing data could be quite useful for marketing. I find that, to increase customer satisfaction, placing tasters of apples next to each variety would be greatly beneficial, especially for newer varieties such as Fuji and Jonagold which may look more similar and be less commonly known. Because consumers tend to have bad visual reactions to Red Delicious apples, samples would be especially beneficial for Red Delicious growers. One could not have inferred this detailed information using a logit specification, or with uncorrelated errors between visual and taste rankings. Only with rank-ordered probit and allowing a correlation between the errors can one deduce such details.

Of course these results for marketing could be biased due to the fact that the experiment is based on an extremely simplified consumer choice experiment which ignores any interactions with prices or varieties (as the respondents didn't know which variety they were tasting). Another caveat is that our sample may include both inexperienced shoppers and individuals who do not typically buy apples. Giving away free samples would presumably be more useful for the population in our experiment than for experienced apple shoppers. In the future, it would be interesting to conduct these experiments with experienced buyers, and to ask them to rank apples based on looks alone, based on taste alone, and based on both (since purchase decisions probably depend on some weighted average of visual and taste rankings). In addition, future investigations should incorporate data on apple prices and study the effects of informing respondents as to the name of the varieties they are choosing between.

Jin et al. (2008) investigate why branding is so much less prevalent for produce compared to other consumer goods. They speculate that consumers are able to predict quality of produce based on intrinsic external attributes and thus branding may be less effective in conveying quality signals for fresh produce. I find that people actually have a rather poor ability to predict which apple they will like. As expected, people tend to prefer either the visual characteristics of multi-colored apples, or the visual characteristics of uni-colored apples. In terms of taste, they tend to prefer either sweeter apples or tarter apples. Red Delicious apples may have a worse reputation than they deserve. They are by far the least popular apple visually, but rank higher in terms of taste.

The information on correlation of unobserved characteristics could also help marketers decide which apples to place next to each other, and which may be considered substitutes such that the presence of one will decrease the share of consumers choosing the other. Moving away from the specific data in this paper, this technique could also be used to combine consumers' rankings over multiple attributes of other products such as televisions (screen size, ease of use, and picture quality) or shower cleaners (smell, skin irritation, and form (spray versus foam versus powder)) as well as to combine revealed and stated preference. In this way products could be designed which would better target specific segments of the population.

References

Adamowicz, W., J. Louviere, and M. Williams (1994). Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management 26*, 271–292.

Banović, M., K. G. Grunert, M. M. Barreira, and M. Aguiar Fontes (2009). Beef quality perception at the point of purchase: A study from Portugal. *Food Quality and Preference 20*(4), 335–342.

Beggs, S., S. Cardell, and J. Hausman (1981). Assessing the potential demand for electric cars. *Journal of Econometrics 16*, 1–19.

Bello Acebrón, L. and D. Calvo Dopico (2000). The importance of intrinsic and extrinsic cues to expected and experienced quality: An empirical application for beef. *Food Quality and Preference 11*(3), 229–238.

Ben-Akiva, M. and T. Morikawa (1990). Estimation of travel demand models from multiple data sources. In M. Koshi (Ed.), *Transportation and Traffic Theory*, pp. 461–476. Amsterdam: Elsevier.

Ben-Akiva, M., T. Morikawa, and F. Shiroishi (1992). Analysis of the reliability of preference ranking data. *Journal of Business Research 24*, 149–164.

Bhat, C. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B 35*, 677–693.

Bradlow, E. T. and P. S. Fader (2001). A Bayesian lifetime model for the "Hot 100" *Billboard* songs. *Journal of the American Statistical Association 96*(454), 368–381.

Bredahl, L. (2004). Cue utilisation and quality perception with regard to branded beef. *Food Quality and Preference 15*(1), 65–75.

Bredahl, L., K. G. Grunert, and C. Fertin (1998). Relating consumer perceptions of pork quality to physical product characteristics. *Food Quality and Preference 9*(4), 273–281.

Caparrós, A., J. L. Oviedo, and P. Campos (2008). Would you choose your preferred option? Comparing choice and recoded ranking experiments. *American Journal of Agricultural Economics 90*(3), 843–855.

Carson, R. T., J. J. Louviere, D. A. Anderson, P. Arabie, D. S. Bunch, D. A. Hensher, R. M. Johnson, W. F. Kuhfeld, D. Steinberg, J. Swait, H. Timmermans, and J. B. Wiley (1994). Experimental analysis of choice. *Marketing Letters 5*(4), 351–368.

Chan, W. and P. M. Bentler (1998). Covariance structure analysis of ordinal ipsative data. *Psychometrika 63*(4), 369–399.

Hajivassiliou, V., D. McFadden, and P. Ruud (1996). Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results. *Journal of Econometrics 72*, 85–134.

Hajivassiliou, V. A. and P. A. Ruud (1994). Classical estimation methods for LDV models using simuation. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 40, pp. 2383–2441. Elsevier Science.

Hausman, J. and P. Ruud (1987). Specifying and testing econometric models for rank-ordered data. *Journal of Econometrics 34*, 83–104.

Jin, Y. H., D. Zilberman, and A. Heiman (2008). Choosing brands: Fresh produce versus other products. *American Journal of Agricultural Economics 90*(2), 463–475.

Luce, R. D. and P. Suppes (1965). Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of Mathematical Psychology*, Volume 3, Chapter 19, pp. 249–410. John Wiley and Sons, Inc.

McCluskey, J. J., R. C. Mittelhammer, A. B. Marin, and K. S. Wright (2007). Effect of quality characteristics on consumers' willingness to pay for Gala apples. *Canadian Journal of Agricultural Economics 55*, 217–231.

Melton, B. E., W. E. Huffman, J. F. Shogren, and J. A. Fox (1996). Consumer preference for fresh food items with multiple quality attributes: Evidence from an experimental auction of pork chops. *American Journal of Agricultural Economics 78*(4), 916–923.

Nalley, L. L., D. Hudson, and G. Parkhurst (2006). Consistency of consumer valuation under different information sets: An experimental auction with sweet potatoes. *Journal of Food Distribution Research 37*(3), 56–67.

Poole, N. D., L. Martínez-Carrasco Martínez, and F. V. Giménez (2007). Quality perceptions under evolving information conditions: Implications for diet, health and consumer satisfaction. *Food Policy 32*, 175–188.

Riddel, M. and R. K. Schwer (2006). Winners, losers, and the nuclear-waste dilemma. *Environmental and Resource Economics 34*, 317–338.

Thybo, A. K., B. F. Kühn, and H. Martens (2004). Explaining Danish children's preferences for apples using instrumental, sensory and demographic/behavioural data. *Food Quality and Preference 15*(1), 53–63.

Train, K. (2000). Halton sequences for mixed logit. Department of Economics, University of California at Berkeley working paper no. E00-278.

Yao, G. and U. Böckenholt (1999). Bayesian estimation of Thurstonian ranking models based on the Gibbs sampler. *British Journal of Mathematical and Statistical Psychology 52*(1), 79–92.