

Working Paper Series
WP 2010-7

**We Know What You Choose!
External Validity of Discrete
Choice Models**

By

R. Karina Gallardo and Jaebong Chang

April 2010

We Know What You Choose! External Validity of Discrete Choice Models

R. Karina Gallardo and Jaebong Chang

April 22, 2010

Abstract:

For over the last thirty years the multinomial logit model has been the standard in choice modeling. Development in econometrics and computational algorithms has led to the increasing tendency to opt for more flexible models able to depict more realistically choice behavior. This study compares three discrete choice models, the standard multinomial logit, the error components logit, and the random parameters logit. Data were obtained from two choice experiments conducted to investigate consumers' preferences for fresh pears receiving several postharvest treatments. Model comparisons consisted of in-sample and holdout sample evaluations. Results show that product characteristics hence, datasets, influence model performance. We also found that the multinomial logit model outperformed in at least one of three evaluations in both datasets. Overall, findings signal the need for further studies controlling for context and dataset to have more conclusive cues for discrete choice models capabilities.

* Gallardo is an assistant scientist and extension specialist in the School of Economic Sciences, Tree Fruit Research and Extension Center, Washington State University, U.S., and Chang is a postdoctoral research associate in the Department of Agricultural Economics, Oklahoma State University, U.S.

Contact author: R. Karina Gallardo, School of Economic Sciences, Tree Fruit Research and Extension Center, karina_gallardo@wsu.edu, (509) 663-8181 x. 261.

I. Introduction

Modeling choice behavior has been utilized in fields as diverse as agricultural economics, environmental economics, marketing, and transportation, especially when eliciting consumers' willingness-to-pay (WTP) and preferences for goods and services or estimating welfare changes for alternative policies and market structures (e.g., see Boxall *et al.*, 2003; Brownstone *et al.*, 2000; Lusk and Schroeder, 2004). Such popularity has led to extensive development in the econometrics of discrete choice models and its applications. For over the last thirty years, the multinomial logit (MNL) model has been the standard in discrete choice models, which assumes independently and identically distributed (IID) stochastic components with type I extreme value in the random utility function, leading to the independence of irrelevant alternatives (IIA) assumption.

Such requirement provides ease of estimation in terms of calculation of probabilities and interpretation of results, yet it has provoked criticisms mainly on deficient representation of real behavior. Concerns about MNL limitations have resulted in a range of competing discrete choice models mostly generalizing the MNL by including preference heterogeneity (see the reviews in Louviere *et al.*, 2000; Train, 2003). Such advanced models aim to increase the accuracy of real world behavior by relaxing the IIA assumption and IID condition. Among such models, the random parameter (or mixed) logit (RPL) has been largely applied because of its flexible specification and behavioral richness by complete relaxation of the IID and IIA conditions (McFadden and Train, 2000; Greene and Hensher, 2003; Jones and Hensher, 2007). The RPL incorporates mean and standard deviation of random parameters estimates allowing for

observed and unobserved heterogeneity across individuals. Yet, RPL probability setting needs to be solved by simulations in most cases not ensuring a globally optimal solution.

Another extension of the MNL is the error component multinomial logit (EMNL) model. This model has been introduced about a decade ago (Brownstone *et al.*, 2000), but is seldom applied. Similar to random effects for panel data analysis, error components are simply included in the utility function to capture unobserved individual specific random effects. Although the model specification is straightforward compared with RPL, it is also necessary to integrate the model to estimate by simulations.

With the increasing popularity of MNL alternatives, a growing branch of literature in discrete choice modeling is focused on comparing results and predictability power across models. Performance of discrete choice econometric functions has been typically analyzed through in-sample statistics or out-of-sample criteria (Greene and Hensher, 2003; Jones and Hensher, 2007; Chang *et al.*, 2009; Shen, 2009). However, it is well known that as more complexity is added to a model the better will the model fit the data in-sample, while the contrary tends to be true out-of-sample. This suggests the need to incorporate both in-and out-of-sample criteria to proof validity of advanced models' results.

In this study, we utilize two choice experiment datasets to compare the performance of three discrete choice models - the MNL, the RPL, and the EMNL, measured in terms of WTP valuations, market share estimates and the prediction success index within sample. Moreover, this study compares the models' ability to predict holdout sample choices. Our results show that estimates for WTP and market share are

significantly different across models and overall predictive performance is different between in-sample and holdout sample tests.

II. Background

Several studies have compared the performance of generalized alternatives to the MNL by considering model-fit statistics, direct estimates for choice elasticity, and prediction accuracy. Among such generalizations, RPL has been most often tested because it has been referred as the most flexible model. Revelt and Train (1998) used the likelihood ratio index to compare RPL with MNL results to estimate households' choices between appliances under rebates and loans on high-efficiency appliance programs. They found that RPL index was higher implying a superior explanatory power when comparing with MNL. Brownstone *et al.* (2000) compared market share predictions for cars with three different fuel types. They found that RPL market share estimates were larger than MNL. Also they concluded that RPL yielded more reasonable market share predictions based on how market share proportions changed when a new alternative was introduced. Greene and Hensher (2003) compared and contrasted MNL, RPL, and latent class (LC) model to elicit preferences for three road types. They analyzed the mean WTP, choice elasticities, choice probabilities, and absolute shares in response to a change in the level of attributes across samples. They concluded that RPL and LC offered more robust results when compared to MNL, but observed that these results are conditioned to a single dataset under specific behavioral assumptions. More recently, Shen (2009) utilized two transport choice surveys to compare WTP for time savings, choice elasticities, predicted choice

probabilities, and prediction success indexes between LC and RPL. Results showed that LC generates superior statistical accuracy than RPL.

Overall previous findings imply that RPL performs better than MNL when looking at model statistics. Nonetheless Train (1998) claims no significant differences between RPL and MNL when comparing compensating variation associated with anglers' preferences for fishing sites, advocating for MNL robust results. He argued that model performance depends on the context of specific situations and data sources. Such findings prove the need to further investigate the ability of various model specifications under different conditions.

It is noteworthy that most past research compared MNL performance with other models by relying mostly on in-sample fit. Yet much of discrete choice modeling is used for predicting real behavior when decision making and out-of-sample tests shall be given equal attention. Only few studies have utilized out-of-sample data to compare across discrete choice models to make inferences on market behavior. Jones and Hensher (2007) investigated the prediction performance of the nested logit, RPL, and LC and found that these models had a high level of consistency on holdout sample outperforming MNL. Chang *et al.* (2009) compared the ability of MNL, the independent availability logit (AIL) and RPL to predict actual retail shopping behavior in three different products. They proved RPL's superior predictive performance and high level of external validity; interestingly they also found that MNL predictions were equally accurate to RPL under some circumstances. This suggests that more flexible models improve out-of-sample predictions, but this may not always hold true. Indeed, it is often observed that more parsimonious models outperform when forecasting (Chatfield, 1995).

Interestingly, the EMNL model has been rarely compared for its explanatory and predictive ability. Only Hensher *et al.* (2007) compared EMNL performance with MNL and showed that the former was superior over the latter, in terms of model-fit statistics, direct elasticities, and hold out sample. They concluded that overall the EMNL provided better explanatory power but “more limited additional predictive performance” than MNL.

To our knowledge none study has compared the statistical accuracy and external validity of EMNL with RPL compared to MNL. In this study we use two choice experiment data sets on consumers’ preferences for fresh pears quality that were conducted in two different time periods. Criteria to make comparisons include in-sample statistics such as WTP, market share estimates, and McFadden’s prediction success index; and to compare the predictive ability we conduct twenty repetitions of holdout sample choice tests.

III. Procedures and Methods

Choice Experiments

We utilized response datasets from two choice experiments on preferences for fresh pears. The experiments were part of sensory tests conducted in December 2008 and March 2009, at the Food Innovation Center, Oregon State University in Portland. The purpose of the sensory tests was to gather general information on consumer preferences for fresh pears and to elicit perceptions on eating quality resulting from different post-harvest treatments. Taste test participants were recruited using an online screening questionnaire sent to about 5,000 consumers in Portland. A planned sample size of 120 consumers was

selected and the criteria used for recruitment was based on the Pear Bureau Northwest depiction of fresh pears' consumers (Moffit, 2002).

During both sensory tests, participants were asked to taste pears under different treatments and to answer a questionnaire. Postharvest treatments differed across trials, given differences in time length in cold storage and fruit maturity. In December, there were four treatments consisting of fruit exposure to ethylene for one, two, and four days, and no ethylene. Whereas pears used in March were applied five treatments consisting of two days with ethylene, one day with ethylene, one day with ethylene plus one day in warm air, two days in warm air, and no ethylene.

Having tasted the pears, respondents were asked to answer choice experiment questions where they indicated which sample, linked to a randomly assigned price, was the most preferred. Choice experiment scenarios also included a "none" option. For the December trial, prices were obtained from grocery stores in Portland during the first week of December 2008 and ranged from \$1.49/lb to \$1.99/lb. To obtain different combinations of treatments and prices, a fractional factorial design was used and yielded thirty-two questions which were divided into four groups of eight questions each, randomly assigned to respondents. In March 2009, there were six options corresponding to ripening treatments and a "none" option. This time, prices were also obtained from Portland grocery stores during the first week of March 2009 and ranged from \$1.39 to \$2.19/lb. Twenty-five questions were generated by a fractional factorial design and divided randomly into two groups of 12 and 13 questions each. Figure 1 shows an example of choice experiment questions used in both trials.

The multinomial logit model

A random utility function for consumer i choosing option j is defined by

$$U_{ij} = \alpha_j + \beta P_{ij} + \epsilon_{ij}, \quad (1)$$

where α_j is the estimated constant parameter for ripening treatment j , β is the marginal utility of price, and P_{ij} is the price. Assuming the stochastic term (ϵ_{ij}) is IID with type I extreme value distribution yields the standard MNL model. The choice probability of an individual i choosing alternative j out of a set J is expressed as

$$P_{ij} = \frac{\exp(\alpha_j + \beta P_{ij})}{\sum_{k=1}^J \exp(\alpha_k + \beta P_{ik})}. \quad (2)$$

The random parameter logit model

The RPL posits preference heterogeneity by specifying individual-specific coefficients of the utility function to be continuously distributed. In this application, the alternative-specific constant terms are assumed as variant parameters across individuals and expressed as:

$$\alpha_{ij} = \bar{\alpha}_j + \sigma_j \omega_{ij} \quad (3)$$

where $\bar{\alpha}_j$ is the mean alternative-specific constant for alternative j , σ_j is the standard deviation of the distribution of α_{ij} around $\bar{\alpha}_j$, and ω_{ij} is a normally distributed random disturbance with mean zero and standard deviation one. The price coefficient β , however, is assumed as a fixed parameter.

The probability that individual i choose alternative j is the integral over all values of α_{ij} weighted by the density of α_{ij} ,

$$P_{ij} = \int \frac{\exp(\alpha_{ij} + \beta P_{ij})}{\sum_{k=1}^J \exp(\alpha_{ik} + \beta P_{ik})} f(\alpha_{ij}) d\alpha_{ij}, \quad (4)$$

where $f(\alpha_{ij})$ is the density function. Because the density function does not have a close form, it is calculated by approximation using simulation procedures, such as maximum simulated likelihood (Train, 2003).

While the RPL's open form probability distribution offers detail in understanding consumer heterogeneity in choice behavior, and flexibility in producing individual specific parameters, it implies computational burden and induces estimation issues associated with identification and normalization (Ben-Akiva and Bolduc, 1996; Ben-Akiva et al., 1997; McFadden and Train, 2000).

The error component multinomial logit model

The ECMNL describes the fact that the unobserved portion of utility is comprised by several components introducing more parsimonious distributions across random factors allowing flexible substitution patterns and correlation across alternatives (Revelt and Train, 1998). As in Brownstone *et al.* (2000), the ECMNL model captures alternative-specific unobserved variation by specifying the random parameters in equation (1) as

$$U_{ij} = \alpha_j + \beta P_{ij} + \theta_j \gamma_{ij} + \epsilon_{ij}, \quad (5)$$

where γ_{ij} is a alternative-specific random error component which is distributed normally with zero mean and standard deviation one and θ_j is the standard deviation of the error component. As with the RPL, because the random effects are included in the conditional choice probability of the ECMNL model it is necessary to estimate by utilizing the simulation approach.

IV. Results

Parameters estimates for the MNL, EMNL and RPL are presented in Table 1 and were computed using LIMDEP®, version 9.0. Random parameters estimates for both EMNL and RPL were assumed to follow a normal distribution, implying that parameters could be either positive or negative. Previous findings in Revelt and Train (1998), who used a log normal distribution, and Shen (2009), who used a triangular distribution, signaled no major improvements from using a different probability distribution than normal. As expected parameter estimates for all treatment options are positive, suggesting that consumers derive a utility from consuming pears regardless of the treatment received. Also note that all price coefficients are negative, indicating that an increase in price will have a negative effect on the willingness-to-pay for pears. It is noteworthy that log-likelihood values differ across datasets. For example, for the December dataset, EMNL outperforms RPL and MNL, whereas for the March dataset, RPL outperforms EMNL and MNL. In relation to parameter estimates for standard deviation in EMNL and RPL models, they resulted statistically significant at the 1% level. As suggested by Revelt and Train (1998) highly significant standard deviation coefficients prove heterogeneity across individual responses implying that models able to capture such heterogeneity, are more robust than models that do not.

Comparisons across models

Criteria to compare MNL, EMNL, and RPL models were twofold. First, in sample statistics willingness-to-pay (WTP), market share estimates, and prediction success indexes were calculated for both datasets. See table 2 for results. Second, prediction tests from twenty randomly selected holdout and estimation samples were conducted. Results for this second group of goodness of fit measures are reported in table 3.

WTP indicates the amount of money the individual would have to pay to be indifferent toward a pear receiving a treatment and no treatment at all. This statistic is equivalent to:

$$WTP = \frac{\alpha_j}{\beta} \quad (5)$$

where α_j and β are parameter estimates from equations (1), (3), and (5), depending whether model is MNL, EMNL, or RPL. Previous findings in Brownstone (2000) concluded that WTP estimated through EMNL or RPL typically remain with minor changes when compared to MNL. However, results from this study indicate that WTP values for MNL are consistently higher than EMNL and RPL. For example, the highest WTP in the December dataset was assigned to treatment 4 days ethylene, MNL value was \$2.53/lb, compared to \$2.01/lb and \$2.14/lb for EMNL and RPL, respectively. Similarly for the March dataset, the highest WTP was for treatment 1 day in ethylene plus 1 day in warm air, this time, the MNL WTP value was \$2.23/lb, compared to EMNL \$1.90/lb and RPL \$1.84.

Concerning, market share, this statistic represents the probability that consumers would choose pear under treatment j , having as all available alternatives to purchase pears under the five or six ripening treatments included in this study, at price level \$1.50/lb. MNL market share values resulted more disperse than EMNL and RPL values. For example, MNL market share for pear under treatment 4 days in ethylene for the December dataset is 50.7 percent compared to 99.7 percent and 94.5 percent when using parameters estimated estimating via EMNL and RPL, respectively. Correspondingly for the March dataset, MNL market share for treatment 1 day in ethylene plus 1 day in warm air was 37.7 percent compared to 71.5 percent and 79.9 percent when using EMNL and

RPL, respectively. Standard errors were estimated via parametric bootstrapping for WTP and market share estimates across three models. Standard errors across models do not show major variations.

With respect to prediction success indexes, this goodness of fit measure compares the proportion successfully predicted for an alternative compared to that which would be predicted by chance. The higher the value of this index, the greater the prediction capability of the model (Louviere *et al.*, 2000). Results are not consistent across datasets. For the data collected in December, EMNL is superior to MNL and RPL; whereas for the March dataset, MNL is superior to RPL and EMNL.

Results for prediction tests using a holdout sample are reported in table 4. We built our holdout sample testing on Haener *et al.* (2001). First, each dataset (December and March) were randomly divided into an estimation sample and a holdout sample. We estimated MNL, EMNL, and RPL parameters for the estimation sample. To assure reliability, we replicated this procedure twenty times. The prediction success for each replication was measured in terms of mean rank and percentage of correctly predicted choices. Unexpectedly, for the December dataset, the mean rank and average percentage of correct predictions favors the MNL model. Conversely, for the March dataset, RPL outperforms both MNL and EMNL.

In summary, there are three contrasting findings across datasets. First, likelihood values signal greater explanatory power to EMNL for the December dataset and to RPL for the March dataset. Second, prediction success indexes shows that for the December dataset EMNL outperforms, while for the March dataset MNL is superior to the other two models. Third, holdout samples tests reveal superior prediction ability for MNL in the

December dataset but for the March dataset it is RPL the model with the highest prediction ability. Results are different from previous findings in Revelt and Train (1998) and Jones and Hensher (2007) who found that RPL outperformed MNL.

An explanation for the differences across datasets is that product attributes influence model performance. Different treatments led to different eating quality characteristics that were perceived by consumers. See table 4 for a summary of participants' overall likings for pear quality characteristics. In the December trial, participants were more homogeneous in their preferences for each treatment than in March. Indeed, in December, 50 percent of respondents agreed in that their preferred sample was treatment 4 days ethylene. Whereas, a wider range of preferences is observed in March, 32 percent for 1 day ethylene plus 1 day warm air and 30 percent for 2 days warm air. We hypothesize that these differences in the distribution of preferences explains the differences in prediction ability across datasets. These claims agree with Train (1998) and Greene and Hensher (2003) who concluded that context, datasets and behavioral assumptions affect RPL superiority to MNL.

V. Conclusions

This study provides comparisons between three popular discrete choice modeling specifications, the MNL, EMNL, and RPL. We used two datasets, provided by two choice experiments that were conducted to measure preference for fresh pears under different ripening treatments. Criteria to compare across models included willingness-to-pay estimates, market share, within sample prediction index, and holdout samples mean rank and percentage of correctly predicted choices.

An increasing body of literature advocates for more flexible discrete choice models, claiming superior in-sample fit and greater out-of-sample predictability. Our results show that EMNL outperformed RPL and MNL when the products being tested exhibited heterogeneous quality characteristics quickly perceived by respondents. Whereas when differences were not easily perceived, RPL outperformed MNL and RPL. Interestingly, MNL outperformed for the holdout sample prediction when using the December dataset and exhibited a higher prediction success index than RPL and EMNL when using the March dataset. This result supports the claim in Chang *et al.* (2009) that more parsimonious models often exhibit a greater predictive ability. Overall, findings in this study raise similar issues to Train (1998) and Green and Hensher (2003) in that further studies controlling for context and dataset nature are needed since they are determinant for measuring the predictive performance of models more flexible than MNL.

References

- Ben Akiva, M., McFadden, D., Makoto, A., Bockenholt, U., Bolduc, D., Gopinath, D., Morikawa, T., Ramaswamy, V., Rao, V., Revelt, D., Steinberg, D. (1997) Modeling methods for discrete choice analysis. *Marketing Letters*, 8, 273-86.
- Ben Akiva, M. and Bolduc, D. (1996) Multinomial probit with a logit kernel and a general parametric specification of the covariance structure. Working paper, department d'Economique, Universite Laval, Quebec.
- Boxall, P., Englin, J., and Adamowicz, W. (2003) Valuing aboriginal artifacts: a combined revealed-stated preference approach. *Journal of Environmental Economics and Management*, 45, 213-30.
- Brownstone, D. (2000) Discrete Choice Modeling for Transportation. Paper prepared for the Ninth IATBR Travel Behavior Conference, Australia, July.
- Brownstone, D., Bunch, D., Train, K. (2000) Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B*, 34, 315-38.
- Chang, J. B., Lusk, J. L. and Norwood, F. B. (2009) How closely do hypothetical surveys and laboratory experiments predict field behavior? *American Journal of Agricultural Economics*, 91, 518-34.
- Chatfield, C. (1995) Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society*, 158, 419-66.
- Greene, W., and Hensher, D. (2003) A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B*, 37, 681-98.

- Haener, M. K., Boxall, P. C. and Adamowicz, W. L. (2001) Modeling recreation site choice: Do hypothetical choices reflect actual behavior? *American Journal of Agricultural Economics*, 83, 629-42.
- Hensher, D. A., Jones, S. and Greene, W. H. (2007) An error component logit analysis of corporate bankruptcy and insolvency risk in Australia, *The Economic Record*, 83, 86-103.
- Jones, S. and Hensher, D. A. (2007) Evaluating the behavioral performance of alternative logit models: An application to corporate takeovers research, *Journal of Business Finance & Accounting*, 34, 1193-220.
- Lusk, J. L. and Schroeder, T. C. (2004) Are choice experiments incentive compatible? A test with quality differentiated beef steaks, *American Journal of Agricultural Economics*, 86, 467-82.
- Louviere, J. J., Hensher, D. A. and Swait, J. D. (2000) *Stated Choice Methods: Analysis and Applications*, Cambridge University Press, Cambridge.
- McFadden, D. and Train, K. (2000) Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15, 447-70.
- Moffitt, K. (2002) Conditioned fruit: is it what consumers are looking for? Paper presented at the Washington Tree Fruit Postharvest Conference, Yakima, Washington, March 12-13.
- Revelt, D. and Train, K. (1998) Mixed logit with repeated choices: Households' Choices of Appliance Efficiency Level. *The Review of Economics and Statistics*, 80,
- Shen, J. (2009) Latent class model or mixed logit model? A comparison by transport mode choice data, *Applied Economics*, 41, 2915-24.

Train, K.(1998) Recreation demand models with taste differences over people. *Land Economics*, 74, 230-39.

Train, K. (2003) *Discrete choice methods with simulation*. Cambridge University Press.

Figure 1. Example of choice experiment questions used in the experiments conducted in December and March.

Table 1. Parameter estimates by experimental sample and model

Variable	DECEMBER 2008			MARCH 2009		
	MNL	EMNL	RPL	MNL	EMNL	RPL
<i>Treatments and price</i>						
<i>Utility function parameters</i>						
1-day ethylene	6.35* (0.51) ^[a]	14.43* (2.97)	22.16* (2.30)	-	-	-
2-days ethylene	6.73* (0.52)	23.42* (1.96)	27.29* (2.44)	-	-	-
4-days ethylene	7.82* (0.52)	29.85* (2.05)	30.36* (2.61)	-	-	-
No conditioning	6.92* (0.52)	23.24* (1.87)	25.91* (2.33)	-	-	-
1-day warm air	-	-	-	5.56* (0.33)	14.68* (1.01)	17.15* (1.50)
1-day ethylene	-	-	-	5.44* (0.34)	12.63* (0.98)	16.12* (2.24)
2-days warm air	-	-	-	5.82* (0.33)	15.35* (1.00)	18.24* (1.41)
1-day ethylene + 1-day warm air	-	-	-	6.44* (0.33)	16.84* (1.03)	20.01* (1.51)
No conditioning	-	-	-	5.33* (0.32)	13.69* (1.07)	14.50* (1.41)
Price	-3.09* (0.28)	-14.87* (1.06)	-14.18* (1.27)	-2.89* (0.18)	-8.87* (0.56)	-10.90* (0.82)
<i>Standard deviation</i>						
1-day ethylene	-	14.17* (1.95)	6.99* (1.14)	-	-	-
2-days ethylene	-	11.64* (1.09)	7.29* (0.75)	-	-	-
4-days ethylene	-	4.46* (0.48)	8.65* (1.16)	-	-	-
No conditioning	-	8.15* (0.96)	7.04* (0.77)	-	-	-
Random effects	-	0.68* (0.47)	-	-	-	-
1-day warm air	-	-	-	-	5.86* (0.48)	5.65* (0.71)
1-day ethylene	-	-	-	-	4.24* (0.36)	6.60* (1.08)
2-days warm air	-	-	-	-	3.87* (0.54)	3.95* (0.55)
1-day ethylene + 1-day warm air	-	-	-	-	5.19* (0.42)	6.03* (0.71)
No conditioning	-	-	-	-	10.18* (0.72)	6.81* (0.98)
Random effects	-	-	-	-	2.56* (0.33)	-
Log likelihood	-1049.84	-494.51	-497.59	-1039.58	-547.26	-522.34
No. of observations	4120	4120	4120	4248	4248	4248

^[a] Numbers in parenthesis are standard errors.

Note: One (*) asterisk indicates significance at 1% level.

Table 2. Willingness-to-pay (\$/lb) and market share estimates by experimental data and model

	December 2008			March 2009		
	MNL	EMNL	RPL	MNL	EMNL	RPL
<i>Willingness-to-pay</i>						
1-day ethylene	\$2.06 (0.08) ^[a]	\$0.97 (0.07)	\$1.56 (0.08)	-	-	-
2-days ethylene	\$2.18 (0.08)	\$1.57 (0.04)	\$1.92 (0.04)	-	-	-
4-days ethylene	\$2.53 (0.10)	\$2.01 (0.03)	\$2.14 (0.04)	-	-	-
No conditioning	\$2.24 (0.08)	\$1.56 (0.06)	\$1.83 (0.04)	-	-	-
1-day warm air	-	-	-	\$1.92 (0.06)	\$1.66 (0.06)	\$1.57 (0.07)
1-day ethylene	-	-	-	\$1.88 (0.06)	\$1.42 (0.04)	\$1.48 (0.18)
2-days warm air	-	-	-	\$2.01 (0.06)	\$1.73 (0.03)	\$1.67 (0.05)
1-day ethylene + 1-day warm air	-	-	-	\$2.23 (0.06)	\$1.90 (0.04)	\$1.84 (0.05)
No conditioning	-	-	-	\$1.84 (0.06)	\$1.54 (0.04)	\$1.33 (0.07)
<i>Market share</i> ^[b]						
1-day ethylene	0.117 (0.012)	0.000 (0.000)	0.000 (0.008)	-	-	-
2-days ethylene	0.170 (0.013)	0.002 (0.001)	0.044 (0.016)	-	-	-
4-days ethylene	0.507 (0.019)	0.997 (0.003)	0.945 (0.015)	-	-	-
No conditioning	0.206 (0.014)	0.001 (0.003)	0.011 (0.011)	-	-	-
1-day warm air	-	-	-	0.157 (0.016)	0.082 (0.035)	0.046 (0.054)
1-day ethylene	-	-	-	0.139 (0.014)	0.011 (0.005)	0.016 (0.048)
2-days warm air	-	-	-	0.203 (0.016)	0.161 (0.041)	0.136 (0.026)
1-day ethylene + 1-day warm air	-	-	-	0.377 (0.021)	0.715 (0.067)	0.799 (0.038)
No conditioning	-	-	-	0.124 (0.013)	0.031 (0.010)	0.003 (0.045)

^[a] Numbers in parentheses are standard errors obtained via parametric bootstrapping.

^[b] Market share estimates are calculated by assuming price is at \$1.50/lb for all options.

Table 3. Results of overall prediction success index from within sample and prediction tests from twenty models over random hold-out samples

Data Set	Model		
	MNL	EMNL	RPL
<i>Overall prediction success index from within sample</i>			
December 2008	0.038	0.042	0.030
March 2009	0.093	0.053	0.090
<i>Prediction test from twenty models over random samples and hold-out samples</i>			
Mean rank			
December 2008	1.550	2.050	1.900
March 2009	1.750	2.200	1.700
Average percentage correctly predicted			
	36.920	35.120	36.240
December 2008	[22.36, 51.17] ^[a]	[21.96, 51.17]	[21.96, 48.96]
	31.250	30.230	31.660
March 2009	[21.15, 42.82]	[19.08, 42.54]	[21.15, 45.11]

^[a] Numbers in brackets are minimum and maximum average percentages of the number of correctly predicted choice sets over the hold-out samples.

Table 4. Sensory tests responses for pears under each treatment

Treatment	Percentage of respondents who ranked “best” each sample (%)	Consumer Ratings for Eating Quality Attributes ^[a]				
		Overall	Sweetness	Juiciness	Firmness	Texture
<i>December trial</i>						
4-days ethylene	50	7.46	6.83	7.57	6.62	6.88
No conditioning	23	6.42	5.92	6.43	5.89	5.82
2-days ethylene	16	6.13	5.06	4.97	6.17	5.94
1-day ethylene	11	5.58	4.34	3.67	5.65	5.23
<i>March trial</i>						
1-day ethylene + 1-day warm air	32	6.60	6.05	6.73	6.68	6.60
2-days warm air	30	6.58	5.99	6.82	6.61	6.67
1-day warm air	17	6.18	5.00	5.63	6.29	6.09
1-day ethylene	6	6.06	5.15	5.62	6.15	5.75
No conditioning	16	6.02	5.28	5.78	6.21	5.78

^[a] Ratings used a 1-9 hedonic scale, where 1=extremely dislike and 9=extremely like.