Store-Level Retail Fruit Demand: Lessons from Omitted Variables

Aaron J. Johnson, Marc McFetridge, and Catherine A. Durham

This study examines how marketing strategies of produce managers affect consumer expenditures for fresh fruit. A non-linear Almost Ideal Demand System was used to model the share equations for Gala apples, Fuji apples, Red Delicious apples, other sweet apples, tart apples, pears, bananas, and oranges. Seventy-nine weeks of data on weekly store sales were collected from two grocery stores in the Portland, Oregon metropolitan area. The objective of this paper is to discuss those variables that were examined but found to be insignificant in the demand model. Those variables include displays, traffic flow, in-store specials, Food Alliance labeling and signage, lagged prices and advertisements, damage-quality measure, incorrect use of inserts and advertisements, nutritional/health information, and the availability of a smaller product. Reasons for the exclusion of these variables and the lessons learned are discussed.

The importance of the produce department to overall store profitability has increased over the past decade (Schaffner 2002; Gentry 2001; Perosio, McLaughlin, Cuellar and Park 2001). With the increased importance of the produce department the marketing strategies used by produce managers have become even more important to store profitability. A study was developed by the authors to explore how the marketing strategies used by produce managers impact consumer purchasing behaviors. Classic demand models for fresh fruit typically examine how aggregated sales are influenced by own and substitute prices, seasonality, advertising, and income levels. This new study examines other factors that expand on classic demand models by incorporating factors that are controlled by the produce managers into the demand model. Including these factors should provide a more complete picture of all the influences on consumer purchasing behavior for fresh fruit at the retail level. The objective of this paper is to report on the variables that were excluded from the final model and the lessons learned from those variables. The following sections of this paper discuss the data and methodology, the variables found to be significant, the variables omitted from the demand model, and conclusions.

Data and Methodology

The data used for this study combined weekly purchases from two retail grocery stores within the same chain. The stores organized their produce sections somewhat differently and were located in different demographic areas of the Portland, Oregon metropolitan area. Weekly store visits entailed the collection of data on apples, bananas, pears, oranges and other hand fruit.[1] Data on prices, origin of production, eco-labeling, fruit sizes, display sizes (each product could be displayed in multiple locations), the size of point-of-purchase material and the corresponding sensory wording, and quality measures were collected for apples, bananas and pears. Display locations were mapped and pictures of the produce area were taken on a weekly basis (for further detail on the data collection process contact the authors). To model demand for apples, bananas, pears, oranges and other hand fruit, the Almost Ideal Demand System (AIDS) was used (Deaton and Muellbauer 1980). Within the weekly data, three varieties of apples (Gala, large Fuji, and large Red Delicious) were available every week, which allowed for an estimation of separate share equations for each variety. The other apple varieties that appeared from week to week were aggregated into either an "other sweet apples"[2] share equation or a "tart apples"[3] share equation. Pears, regular bananas, oranges, and other hand fruits[4] also had

The authors are assistant professor, faculty research assistant, and assistant professor at Oregon State University.

[1] "Other hand fruit" is the type of fruit people can eat with little preparation, similar to apples; e.g., oranges and kiwis.

[2] Other sweet apples included Golden Delicious, Cameo, Jonagold, Small Red Delicious, Small Fuji, Pacific Rose, Honey Crisp, Sonata, and Queen apples.

[3] Tart apples included Braeburn, Pink Lady, McIntosh, Southern Rose, Pippin, and Granny Smith apples.

[4] Other hand fruit included kiwis, peaches, plums, bagged fruit, organic fruit, etc. This equation was excluded from the model to ensure that the data matrix would be non-singular.

share equations. The price is a weighted-average price for each of the aggregated share equations.

Variables Omitted from the Demand Model

This study found that prices, total display size, end displays, total point-of-purchase size, advertisements, a bagged-apple substitute, sensory wording, origin/labeling, quality measures, seasonality, total number of products available, and an expenditure index are all significant variables. Other variables were examined to determine their influence on consumer purchasing behaviors for fresh fruit. The results were inconclusive, and as a result the variables were omitted from the final model. The variables omitted from the final model were omitted for three specific reasons: being highly correlated with other demand variables, few non-zero observations (little variation in the variable), and insignificant estimated coefficients across share equations. The following section lists the variables omitted from the demand model based on the reasons for their omission, and provides a brief explanation of the variables.

Highly Correlated

Generally excluding highly correlated variables improves the power of the model without losing valuable information. Two store level variables, "outside displays" and "nutritional information," were considered but omitted from the final model because they were highly correlated with the store dummy variable. The outside-display variable was created to track whether or not a specific product was displayed outside the produce department, including displays outside of the main entrance of the store. The binary nutritional (health) information variable was assigned a one whenever Five-a-Day point-of-purchase promotional inserts, nutritional facts inserts, or other nutrition information material was observed.[5]

Few Non-Zero Observations

A number of store practices, for which information was collected and recorded as binary variables, were very infrequently observed. It is difficult to be certain if any significant coefficients estimated from variables under these circumstances are explaining the dependent variable or are registering some random occurrence. The variables excluded for having few non-zero observations are freestanding displays, entry displays, traffic flow, incorrect use of inserts and advertisements, in-store specials, and the availability of a smaller product.

Variables were created to track freestanding displays and entry displays. Freestanding displays are defined as having four sides that are accessible by consumers, and are usually in the form of bins placed on the floor. Entry displays are displays placed in the entrance of the produce department. These are the displays that consumers typically see as they first enter the produce department based on consumer movement through the store.

Traffic-flow variables were created to track how product placement in different consumer traffic-flow areas impacts the demand for fresh fruit. The displays in the produce department were classified as being in high, medium, and low traffic-flow areas. These classifications were created from discussions with the produce managers and their staff.

Binary variables were created to track the incorrect use of point-of-purchase inserts (either incorrect use of in-store special point-of-purchase signage or not using an advertisement point-of-purchase signage when a product was advertised), differences in prices between signage and flyer advertisement, and in-store specials.

A binary variable was created to determine how consumers' purchasing behaviors for fresh fruit are impacted by the availability of a smaller product. The PLU numbers that appear on individual products are different depending on the size of the product. However, if the large and small products appeared in the same display then the two sizes were treated as a single product. The small products were so infrequently available in a separate display when large products were also available that this variable was dropped.

Insignificant Estimated Coefficients

Determining the significance of estimated coefficients generated from the regression analysis was based on the probability-values. If these were greater than 0.1, the coefficient was interpreted to be not significantly different from zero. If an estimated

[5] The Five-a-Day point-of-purchase promotional inserts are used to promote eating five servings of fresh fruits or vegetables each day.

coefficient was determined to not be significantly different from zero, the corresponding variable does not appear to have an impact on the dependent variable. The variables that were insignificant in each of the estimated share equations and thus excluded from the final model are multiple displays, the Food Alliance label, lagged price and advertisements, and damage quality measure.

Variations of multiple-display variables were explored. The simplest of these were binary variables created to track outside displays or secondary indoor displays. Also, individual variables for the size of a primary, secondary, and tertiary display were created. The estimated coefficients for the rarely observed tertiary display-size variable were insignificant. Examining the display-size estimated coefficients for all displays revealed that the estimated coefficients of each of the display variables did not significantly differ from one another. Thus a single total display-size variable appeared adequate, and the variables for a primary, secondary, and tertiary display sizes were excluded from the model.

Ecolabel variables were created to examine whether the use of a regional ecolabel sticker and signage was influencing consumer behavior. Mixed results from the ecolabel variables led to a survey that revealed that 21% of the survey respondents in the stores where data was being collected were aware of the ecolabel, but only one third of those recognized which products are actually certified (Johnson et al. 2002).

Some limited testing was conducted on variables that might influence purchases the following week. The possibility of prior advertising affecting current sales was examined through a binary variable for inclusion of the product in the previous week's advertising flyer. The coefficients estimated for the lagged advertisements were insignificant across equations, so the lagged variables were excluded from the model.

The damage quality measure was developed to track whether the fruit within the displays exhibited any punctures or holes in the skin. A four-point quality scale was developed to document the percentage of damage that was present in the displays of fresh fruit (4 = little damage, 1 = large amounts of damage). The estimated coefficients for damage were insignificant across all share equations leading to its exclusion from the model.

Conclusions

The study discussed here provides insight into how different variables impact consumers' purchasing behaviors for fresh fruit at the retail level. The variables found to be significant include prices, total display size, end displays, total point-of-purchase size, advertisements, the availability of a bagged-apple substitute, sensory wording, origin/labeling, quality measures, seasonality, total number of products available, and consumer expenditures. This paper, however, focuses on those variables that were examined but excluded from the final model. These variables were excluded for being highly correlated with a store variable, having too few non-zero observations, or because the estimated coefficients were insignificant across share equations. Two specific lessons about conducting real-time market research are taken from this experience: the need for better experimental design through the use of appropriate statistical tools, and the need to standardize factors that are not core to the study.

This study was intended to examine how the marketing strategies used by produce managers affect consumer purchasing behaviors with minimal demands on the store or its produce managers. As a result, some variables of interest displayed little variation from week to week, as produce mangers seldom changed some arrangements within their department, and these variables were omitted in the final model. While economic research is usually based on aggregated, secondary data rather than experimental data, working directly at the data-collection level opens the possibility of designing for variation in variables of interest, which could greatly enhance the ability to analyze store strategies.

Though it would be necessary to obtain store cooperation, incorporating an experimental design into future projects could resolve the problem of limited variations. To achieve this efficiently and to prevent bias, the store factors of interest must be randomly changed over an extended period of time. A tool called a randomized block design would achieve this objective (Wackerly, Mendenhall, and Scheaffer 2002). The randomized block design would allow for the different strategies of interest, or treatments, to be evenly and appropriately distributed across time and stores to insure proper variation in the variables without introducing bias.

In addition to the use of a random block design, variables that are less central to the study's objectives ought to be standardized as much as possible across stores. This standardization would reduce the influence these factors may have on consumers' habits, allowing the store variables to account more for the differences in consumers between stores than differences between these practices. An example from this study that could have been easily standardized across stores is Five-A-Day nutritional brochures. As it was, the Five-a-Day signage only appeared in one store.

References

Deaton, A. and J. Muellbauer. 1980. "An Almost Ideal Demand System." *American Economic Review* 70:312–26.

Gentry, K. 2001. "New Washington Apple Chief Looks to Stem Economic Woes." *Fruit Grower News* 40(2):1,35.

Johnson, A. J., C. A. Durham, D. O. Andrade, and M. V. McFetridge. 2002. "Recognition of Eco-Labeled Products in Retail Environments." Poster at the Conference on Ecolabels and the Greening of the Food Market, Boston Massachusetts. November 7–8.

Perosio, D. J., E. W. McLaughlin, S. Cuellar, and K. Park. 2001. *Supply Chain Management in the Produce Industry*. Cornell University and Produce Marketing Association: Newark.

Schaffner, D. J. 2002. United States Produce Markets in Transition, Today and Tomorrow. *Journal of Food Distribution Research* 33(2):61–66.

Wackerly, D. D., W. Mendenhall. and R. L. Scheaffer. 2002. *Mathematical Statistics with Applications,* 6th edition. Wadsworth Group: Pacific Grove.