ESSAYS ON CONSUMER DEMAND: MARKETING AND POLICY IMPLICATIONS FROM PRIMARY AND SECONDARY DATA

By

XUEYING MA

A dissertation submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
School of Economic Sciences

DECEMBER 2023

To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of XUEYING MA find it satisfactory and recommend that it be accepted.

R. Karina Gallardo, Ph.D., Chair

Felix Munoz-Garcia, Ph.D.

Wesley Blundell, Ph.D.

Jia Yan, Ph.D.

ACKNOWLEDGMENT

I want to express my sincere gratitude to my advisor, Dr. R. Karina Gallardo, for her invaluable guidance and unwavering support in the development of my dissertation. She provided me with the invaluable chance to engage in a precious research project, involving the entire process from survey design to the preparation for publication. Her patience and continuous mentorship have been instrumental in my growth and evolution as a researcher. I would like to thank my respected committee members; Dr. Felix Munoz-Garcia for his invaluable guidance and feedback on my research, and for his mentorship and support across a range of academic concerns, Dr. Wesley Blundell for his valuable feedback and instruction on my research and for providing me with a solid foundation in research methodology, Dr. Jia Yan for his invaluable feedback on my research and for imparting knowledge in econometrics and machine learning. I would like to thank Dr. Jill McCluskey for offering me the invaluable opportunity to contribute to a meaningful research project and for her support in various academic matters. I would also like to thank Dr. Elizabeth Canales for her expert guidance in discrete choice experiments and the use of Nlogit.

I really appreciate the financial support provided by the USDA National Institute of Food and Agriculture through the Specialty Crop Research Initiative project titled "VacciniumCAP: Leveraging Genetic and Genomic Resources to Enable Development of Blueberry and Cranberry Cultivars with Improved Fruit Quality Attributes" (Award 2019-21181-30015) under the supervision of Dr. Massimo Iorizzo. I would like to thank Jayme Smith of the Washington State Tree Fruit Association for sharing their data. Lastly, I would like to acknowledge the Mercer Fellowship for their support in funding my research.

I would like to express my profound gratitude to my mother, Hongjie Xing, and my father, Hongjun Ma, for their sacrifices in creating opportunities for me and supporting me in achieving my current position. I would also like to thank my partner, Fangjiao Ma, for his consistent motivation to help me grow and for providing me with dependable emotional support, which has been a wellspring of strength during challenging moments. I would like to acknowledge my grandmother Lingxiao Ma, and my aunt, Hongya Xing, for their encouragement in pursuing my dreams. I am thankful to Xiaoyi Chen for offering much-needed encouragement during my difficult times. I want to convey my appreciation to Jui Sen for her exceptional support, both in my academic and personal life.

Lastly, I would like to thank my cherished furry companions who have been a constant source of comfort, joy, and serenity throughout my academic journey. To my rabbit, Jingle, who has been the best therapist a tired, stressed-out PhD candidate could ask for. To my cat, Felix, whose curiosity and playfulness always managed to lighten my mood. I also want to extend my thanks to a previous furry companion. My cherished guinea pig, Fudan (may he rest in peace), holds a special place in my heart for the joy and tranquility he brought into my life, particularly during the most challenging moments when I was in my hometown, Pingdingshan, China, preparing to study abroad, which ultimately led me to my journey here. I am eternally grateful for the love, serenity, and comfort they have provided me throughout my academic journey and in life.

ESSAYS ON CONSUMER DEMAND: MARKETING AND POLICY IMPLICATIONS FROM PRIMARY AND SECONDARY DATA

Abstract

by Xueying Ma, Ph.D.
Washington State University
December 2023

Chair: R. Karina Gallardo

This dissertation focuses on analyzing consumer demand for agricultural products and explores its marketing and policy implications by utilizing both primary and secondary data. The first chapter delves into estimating consumers' willingness to pay for attributes such as sugar content, CRISPR technology, and cranberry flavor intensity in two cranberry products, each presented with different health-related information treatments. Respondents expressed a preference for reduced sugar content over regular sugar products, for conventional over CRISPR breeding methods, and for full/intense cranberry flavor over weak/bland flavor. Interestingly, information emphasizing cranberries' health benefits and sugar intake recommendations amplified the reject to reduced sugar content, surpassing the reject to CRISPR.

In the second paper, a Basket-Based Choice Experiment was employed to identify sensory and hedonic quality descriptors of fresh blueberries that could potentially increase the likelihood of purchase. The findings revealed that blueberry packages with a "Stay Fresh" descriptor had a lower price elasticity compared to packages without descriptors or those with "Sweety" and "Crunchy" descriptors. This suggests that consumers are less responsive to price fluctuations when blueberry packages include language indicating an extended shelf life. Moreover, the study

indicated that strawberries, blackberries, raspberries, and blueberries are more likely to be purchased together rather than considered substitutes. Additionally, the research identified specific demographic and behavioral factors associated with a higher likelihood of choosing blueberries from a selection of commonly consumed fruits.

In the third paper, weekly shipment data from nine apple varieties in Washington state were analyzed to assess the predictive capabilities of various time series models. While some traditional time series models, specifically the Seasonal Autoregressive Integrated Moving Average, excelled in predictive accuracy and capturing variability, this study recommends the use of a machine learning model, specifically Facebook Prophet, due to its computational efficiency and strong predictive accuracy. Additionally, the research explored the impact of introducing a new apple variety, Cosmic Crisp®, on the shipments of existing apple varieties using an interrupted time series analysis. The introduction of Cosmic Crisp® is associated differently across different apple cultivars, with some exhibiting no significant changes, while others experiencing a decrease in subsequent shipment levels. Importantly, there is no association between the overall apple shipments in Washington State and the introduction of Cosmic Crisp®. Furthermore, our results indicate that the shipments of Cosmic Crisp® are associated with an increase in the supply of specific apple varieties as well as the overall apple supply.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENT iii

ABSTRACT v

LIST OF TABLES ix

LIST OF FIGURES xi

CHAPTERS

CHAPTER ONE: WOULD CONSUMERS ACCEPT CRISPR FRUIT CROPS IF THE BENEFIT HAS HEALTH IMPLICATIONS? AN APPLICATION TO CRANBERRY PRODUCTS. 1

1.1 Introduction 2

1.2 Literature review 7

1.3 Methods 9

1.4 Results 20

1.5 Conclusions and Implications 26

REFERENCES 30

CHAPTER TWO: QUALITY-RELATED DESCRIPTORS TO INCREASE FRESH BLUEBERRIES PURCHASE - EVIDENCE FROM A BASKET-BASED CHOICE EXPERIMENT 46

2.1 Introduction 47

2.2 Literature Review 49

2.3 Experimental design 52

2.4. Empirical approach 55

2.5 Results 58

2.6 Conclusion and implications 66

REFERENCES 68

CHAPTER THREE: THE IMPACT OF INTRODUCING A NEW APPLE VARIETY ON WASHINGTON STATE'S APPLE SHIPMENTS - EVIDENCE FROM A TIME SERIES ANALYSIS 100

3.1 Introduction 101

3.2 Literature review 104

3.3 Methodology 108

3.4 Data 116

3.5 Empirical analysis 118

3.6 Discussion and conclusion 124

REFERENCES 127

APPENDIX 146

APPENDIX 1.A: DRIED CRANBERRY SURVEY: NARRATIVE PROVIDED TO RESPONDENTS BEFORE COMPLETING THE DISCRETE CHOICE SCENARIOS UNDER TREATMENT 4, EXAMPLE OF A DISCRETE CHOICE SCENARIO, AND CERTAINTY SCALE – TREATMENT 4. 147

APPENDIX 1.B: CRANBERRY JUICE SURVEY: NARRATIVE PROVIDED TO RESPONDENTS BEFORE COMPLETING THE DISCRETE CHOICE SCENARIOS UNDER TREATMENT 4, EXAMPLES OF A DISCRETE CHOICE SCENARIO, AND CERTAINTY SCALE UNDER TREATMENT 4. 149

LIST OF TABLES

Page

Table 1.1: List of attributes and attribute levels for sets of discrete choice experiment scenarios for dried cranberries, and cranberry juice. 37

Table 1.2: Description of base cranberry product option and hypothetical option used in compensating surplus analysis. 38

Table 1.3: Summary statistics of respondents' sociodemographic characteristics for the two surveys (dried cranberries and cranberry juice) and four treatments of information 39

Table 1.4: Coefficient estimates for the dried cranberry model, considering information effects using the GMNL–II model in WTP space. 40

Table 1.5: Coefficient estimates for the cranberry juice model, considering information effects using the GMNL–II model in WTP space. 41

Table 1.6: Measures of goodness of fit as part of the selection criteria to identify the number of classes in the latent class model, for the three survey versions: Dried cranberries and cranberry juice. 42

Table 1.7: Parameter estimates for the latent class model to represent preference heterogeneity for reduced sugar content – dried cranberries. 43

Table 1.8: Parameter estimates for the latent class model to represent preference heterogeneity for reduced sugar content –cranberry juice. 44

Table 2.1: Price levels used in the basket-based choice experiment. 81

Table 2.2: Socio-demographic characteristics of the pooled sample and across treatment.......... 82

Table 2.3: Summary statistics of sociodemographics across all treatment samples. 83

Table 2.4: Frequency distribution of responses to fresh fruit purchase habit questions. 88

Table 2.5: Weighted average of ratings of importance of different issues related to fresh blueberry consumption. 91

Table 2.6: Comparison of model specifications. 93

Table 2.7: Summary statistics of pooled sample and across treatment groups. 94

Table 2.8: Baseline utility estimates from the multivariate logit model. 95

Table 2.9: Baseline utility estimates from the MVL model – Sample of respondents presented the control treatment. ... 96

Table 2.10: Cross-utility effect estimates from multivariate logit model change in utility of purchasing. ... 97

Table 2.11: Own and cross price elasticities of blueberries at mean demographics and prices implied by multivariate logit model. ... 98

Table 2.12: Own and cross elasticities of fruits – Sample of respondents presented the control treatment. ... 99

Table 3.1: Descriptive statistics of the weekly shipments of apples from Washington. ... 136

Table 3.2: Augmented Dickey Fuller results. ... 137

Table 3.3: Time series model predictive accuracy comparison. ... 138

Table 3.4: Investigating the impact of Cosmic Crisp® presence on other apple varieties' shipments: A SARIMAX analysis. ... 139

Table 3.5: Investigating the impact of Cosmic Crisp® shipments on other apple varieties' shipments: A SARIMAX analysis. ... 140

Table 3.6: Investigating the impact of Cosmic Crisp® presence on other apple varieties' FOB prices: A SARIMAX analysis. ... 141

LIST OF FIGURES

Page

Figure 1.1: Trade-off between the WTP for regular sugar content and the WTP for CRISPR. ... 45

Figure 2.1: Example of a fruit-basket choice scenario ... 77

Figure 2.2: Blueberry clamshell labels used in each treatment. ... 78

Figure 2.3: Percent of times each fruit was chosen, across all treatments. ... 79

Figure 2.4: Histogram of individual own-price elasticity of blueberries. ... 80

Figure 3.1: Weekly Shipments of Apples from Washington, August 25, 2008, to February 27, 2023. ... 142

Figure 3.2: Predicted shipments vs. actual shipments ... 143

Figure 3.3: Comparison of Actual Shipments to SARIMAX Model Predictions in the Absence of Intervention. ... 144

Figure 3.4: Comparison of Actual FOB Prices to SARIMAX Model Predictions in the Absence of Intervention. ... 145

Dedication

To my parents, my partner, and my furry companions.

I am truly blessed to have you all in my life.

CHAPTER ONE: WOULD CONSUMERS ACCEPT CRISPR FRUIT CROPS IF THE BENEFIT HAS HEALTH IMPLICATIONS? AN APPLICATION TO CRANBERRY PRODUCTS.

Abstract

Cranberry products are perceived as healthy due to their high antioxidant content yet adding sugars to increase their palatability deters consumption. Plant breeding technologies such as gene editing, specifically the Clustered Regularly Interspaced Palindromic Repeats (CRISPR), offer a plausible alternative to develop cranberries with desired traits (e.g., lower acidity, increased sweetness). We estimated consumers' willingness to pay for sugar content, CRISPR, and cranberry flavor intensity for two cranberry products under different health-related information treatments. Respondents stated a discount for regular sugar content favoring reduced sugar products, for CRISPR compared to conventional breeding, and for weak/bland compared to full/intense cranberry flavor. Compensated valuation analysis of products with different attribute levels indicates that consumers were willing to pay a premium for cranberry products with reduced sugar content, CRISPR-bred, and full/intense cranberry flavor relative to products with regular sugar content, conventionally bred, and weak/bland flavor. Information treatments highlighting cranberries' health benefits and recommendations to limit sugar intake increased consumers' discounts for regular sugar content, surpassing the discount for CRISPR. This research underscores the importance of the conditions under which breeding technologies might gain public acceptance.

Keywords: Gene Editing, Consumer Preference, Cranberries, Health related Information, Willingness to Pay.

1.1 Introduction

Gene editing is a relatively new breeding technology with increasing applications since its development in the 2010s. This technology targets and controls a specific genome portion without inserting foreign DNA into the host organism (Doudna and Charpentier 2014). Among gene editing technologies, Clustered Regularly Interspaced Palindromic Repeats–Cas9 (hereafter CRISPR) is the most commonly used method due to its reduced cost, enhanced efficiency, and relative ease of use (Critchley et al., 2018). Since its introduction, the application of this technology has been tested in several areas of biological research and model systems, including human disease discovery and treatments, food processing, and crop improvement (Hall, 2016; Haspel, 2018).

Scientific research on CRISPR applications in agriculture is abundant. Findings have shown improvements in crop quality attributes, agronomic traits, and climate stress tolerance in multiple crops (Menz et al., 2020; Zhang et al., 2018). In this study, we investigate whether consumers would perceive gene editing as another iteration of genetic modifications with unpredictable consequences to human health and the environment, similar to their views on genetic engineering[1] applications to agriculture.

New plant breeding technologies, specifically genetic engineering, have scientifically proven contributions to agricultural crops. After three decades of scientific research and commercial applications of genetic engineering in agriculture, there is no proof of an increased risk to either

[1] In this study, we use genetic engineering to refer to the use of recombinant DNA technologies to alter the genetic sequence of an organism, and used to create a transgenic organism, that contains genome consisting of DNA sequences from a different species (Entine et al., 2021). In the survey conducted in this study, we used the terminology GMO- genetically modified organisms, because is the terminology most known to the public. When reporting results from the survey we used the term GMO otherwise we refer to genetic engineering.

human health or the environment compared to conventionally bred crops (Qaim, 2020). Despite the scientific evidence, food manufacturers and retailers shared the expectation that consumers would respond negatively to genetic engineering applications in foods (Kalaitzandonakes et al., 2018). The lack of wide public acceptance of genetic engineering applications in agriculture hindered the full realization of potential benefits (Alston and Pardey, 2021). However, the literature also shows that genetically engineered products exhibit the largest market share (60%) for specific food categories, such as salads and cooking oils (Kalaitzandonakes et al., 2018).

Cranberries offer an interesting case for investigating the trade-off between health-related attributes in agricultural products demanded by consumers and the application of new breeding technologies. Cranberries are high in acidity and contain low amounts of natural sugars. Therefore, the industry must add regular sugar or sugar naturally occurring in other fruits to improve the palatability of processed cranberry products; thus giving the perception that the total sugar content is higher compared to other fruit juices, when this is not necessarily the case. The high anthocyanin and proanthocyanidin content of cranberries has been proven to positively affect human health, juxtaposing with added sugar's perceived negative health effects. A plausible solution is to develop cranberry cultivars using gene editing technology or traditional breeding that either have lower acid content or high natural sugars and retain their anthocyanin and proanthocyanidin content. There is also an increased interest in the development of genetic markers for cranberries associated with sugars and phytochemicals to facilitate breeding with other species of the genus *Vaccinium,* such as wild cranberries, lingonberries, deer berries, even certain blueberries that are cross-compatible with cranberries and possess an array of phytochemicals and other traits with high agronomical value. The aim is to develop cranberry varieties with enhanced quality for processing and human nutrition and improved palatability.

The cranberry products included in this study were selected for their importance in the cranberry industry, as measured by their sales volume. In the United States, 95% of the cranberries grown are processed and 5% are sold fresh (Agricultural Marketing Resource Center, 2023). The major volume in the U.S. domestic market is for juices and sweetened dried cranberries. According to the largest cooperative of cranberry growers in the United States, on average 70-75% of the fruit is processed into juice and sweetened dried cranberries, 12-18% is processed into juice only, 4-6% is processed into sauce, and 3-5% is sold as fresh cranberries (R. Serres, personal communication, 31 May 2023).

The "unhealthy" perception of "Added sugars" is amplified by the recent U.S. Food and Drug Administration (FDA) labeling rule requiring products to explicitly report "Added Sugars" on the Nutrition Facts Panel (NFP) in addition to the "Total Sugar" content. Previous studies have found that consumers often misinterpret the information on "Added Sugar" and "Total Sugar" on the NFP of packaged products (Kim et al., 2021b; Laquatra et al., 2015; Tierney et al., 2017; Khandpur et al., 2020). Studies analyzing the new "Added Sugar" labeling mandate have found no effects of the labeling on purchase behavior (Neuhofer et al., 2020). Other studies found that individuals' self-perceived healthy lifestyles positively influenced the labeling effects on purchase behavior (Kim et al., 2021a; Fang et al., 2019).

Literature states that consumers' perception of genetic engineering is influenced by information available, prior knowledge, perceived risks and benefits, and individual characteristics (Hu et al., 2022; Uddin et al., 2022). Focusing on the effects of information, studies have analyzed different aspects, such as narrative style when presenting the information (Yang and Hobbs, 2020; McFadden et al., 2021); and the effects of trust on the sources of the information (Paudel et al., 2023). Studies centering on the effects of information explaining genetic engineering found that positive information increases acceptance of these methods (Lusk et al., 2004). And when both positive and negative information on genetic engineering is presented, the negative outweighs the effects of the positive information (Lee et al., 2018). Kilders and Caputo (2021) analyzed the effects of information centering on the results of applying gene editing technology, in this case, to breed cows with no horns, improving animal welfare. They found that when information highlighting the enhancement of animal welfare acceptance of both conventional and gene-edited dehorned cows increased. Also, that information increased the preference distributions, implying that information had a heterogeneous impact on preferences.

This study has four objectives. First, we estimate consumers' willingness to pay (WTP) for CRISPR versus conventional breeding in dried cranberries and cranberry juice. Second, given the importance of information on accepting a novel breeding technology, we center on the effects of information -highlighting the health benefits of consuming cranberries and the effects of sugar intake on diets- on consumers' WTP for CRISPR-bred cranberries. Third, we conduct a welfare analysis on the potential impact of using CRISPR. Finally, we assess differences in WTP for cranberry product attributes across respondent segments.

The contribution of this study is to advance knowledge on the public's acceptance of CRISPR, a relevant topic considering its exponential growth in the agri-food industry. The scientific community and agricultural stakeholders should know which crop improvements would enhance public acceptability or mitigate the rejection of CRISPR. To our knowledge, there are no studies investigating the application of gene editing to improve palatability or enhance the healthfulness of a product. That is, whether consumers would be more receptive to CRISPR when its application results in a cranberry product that is perceived to be healthier as it exhibits reduced total sugars or no added sugars. Given the potential of CRISPR technology, applying CRISPR to cranberries is plausible. This study also aims to fill the gap in understanding how various information treatments regarding recommendations to limit added sugars and the health benefits of cranberries could affect the WTP for total sugars and the trade-offs between reduced sugar content and CRISPR.

About CRISPR labeling regulations in the U.S., there is a mix of guidelines from multiple agencies, suggesting that the extent of the gene-edited crop regulations will happen on a case-by-case basis (Parrott, 2022). The U.S. Department of Agriculture-Animal and Plant Health Inspection Service (USDA-APHIS) approved the release of gene-edited organisms without further regulation only if it does not pose any plant or animal pest risk, beyond this, edited organisms are subject to regulatory status review (Entine et al., 2021). The U.S. Department of Agriculture-Agricultural Marketing Service (USDA-AMS) released the "National Bioengineered Food Disclosure Standard", stipulating that foods containing gene-edited ingredients would not be subject to disclosure; only if the ingredients do not come from crops involving novel DNA combinations that were created by other methods different from conventional breeding or found in nature (Entine et al., 2021). The Environmental Protection Agency (EPA) has shown the intention to regulate gene-edited plants that have a pesticidal property for pest resistance. The FDA released the "Plant and Animal Biotechnology Innovation Plan" to clarify their policies regarding food safety evaluations of foods containing ingredients from gene-edited crops (Entine et al., 2021).

Given that different labeling mandate scenarios for CRISPR foods are possible, we consider informative for the scientific community and the agricultural industry to explore consumers' acceptability of CRISPR for two reasons. First, one issue differentiating CRISPR from genetic engineering is that it was developed by academia, and all information about the technology and its applications are being made public. Increased transparency around CRISPR, community involvement, and applications that benefit the public interest could help gain public acceptance and dissipate some of the concerns raised about genetic engineering (Hall, 2016; Haspel, 2018). Second, more applications in the food and fiber sector are likely to be available in the marketplace. CRISPR is more affordable and accessible to a wider variety of institutions and companies and is not exclusive to large multinational companies (Haspel, 2018; Dewey, 2018).

1.2 Literature review

There is abundant literature on consumers' WTP for genetically engineered crops with a consistent finding: consumers are willing to pay price premiums to avoid foods that use ingredients from genetically engineered plants and animals (Lusk et al., 2005; Dannenberg, 2009). When comparing discounts across different food products, consistently, consumers applied a larger discount for genetically engineered fresh foods than for genetically engineered processed foods (Lusk et al., 2015). When studying the effects of labeling, a study found that the presence of “genetically engineered” labels boosted the demand for unlabeled apples, strawberries, and potatoes (Yeh et al., 2019).

Previous studies analyzing consumers' WTP for foods from gene-edited crops have found that individuals were willing to discount less for gene-edited foods than genetically engineered foods, with some exceptions (Hu et al., 2022). However, both gene-edited and genetically engineered plants and animals experienced a discount compared with their conventionally bred counterparts (Shew et al., 2018; An et al., 2019; Muringai et al., 2020; Yang and Hobbs, 2020; Marette et al., 2021; Kilders and Caputo, 2021). Shew et al. (2018) found that respondents were more willing to consume gene-edited compared to genetically engineered rice. However, their sample of respondents stated a discount for genetically engineered and gene-edited rice compared to the conventionally bred product. An et al. (2019) found that their respondents were willing to pay a price premium for gene-edited relative to genetically engineered canola oil. Yang and Hobbs (2020) and Marette, Disdier and Beghin (2021) concluded that their respondents were willing to discount both gene-edited and genetically engineered-apples. However, the discount for gene-edited was smaller than the discount for genetically engineered apples. Muringai et al. (2020) found that respondents stated a discount for frozen French fries produced using genetically engineered and gene-edited compared to conventionally bred potatoes. Still, the discount for gene-edited was smaller than that for genetically engineered potatoes.

On the effects of information, Kilders and Caputo (2021) found that information about the potential to use CRISPR to enhance animal welfare positively affected the WTP for milk from cows that have been gene-edited to prevent painful dehorning. Hu et al. (2022) found no differences between the WTP of CRISPR and genetically engineered orange juice. However, with information on how each technology works, the WTP for both technologies increased.

McFadden et al. (2021) found that both positive -linking gene editing with conventional breeding- and negative -linking gene editing with genetic engineering- messaging strategies led to similar discounts for gene editing and genetic engineering. Paudel et al. (2023) found that survey respondents exhibited greater preference and WTP for gene-edited foods developed by domestic startups and universities than multinational firms. They also found that communicating gene-edited crops' health and environmental benefits enhanced respondents' acceptance.

Overall, studies show that the extent of the acceptance of gene editing over genetic engineering depends, in part, on the nature of the innovation and, thus, the benefit perceived by the consumer. As agriculture faces a changing production environment, increasing global consumer demand, and consumer demand for healthier products, the future food supply will largely depend upon the development and application of technologies such as genome editing to ensure global food demand is met (Voytas and Gao, 2014; Qaim, 2020; Nes, Schaefer and P. Scheitrum, 2022).

1.3 Methods

1.3.1 Data collection – Experimental design

The data were collected via two online surveys —one for dried cranberries and another for cranberry juice— administered by the Qualtrics Research Services™ consumer research panel. We asked Qualtrics to gather participants over 18 years of age through random selection, to match the demographic profile of gender, age, and income as closely as possible to the general population in the United States. The surveys were pretested during a soft launch in November 2020, and data collection took place from December 2020 through March 2021.

The survey consisted of eight versions originating from the combination of four information treatments and two products (dried cranberries and cranberry juice[2]). Qualtrics provided 250 nationwide respondents for each survey version, resulting in 1,000 responses for the dried cranberry survey and 1,000 for the cranberry juice survey, totaling 2,000 respondents. The screening criteria for the respondents varied based on the product being surveyed. The dried cranberry survey aimed to gather responses from both regular and non-regular consumers, so individuals who purchased or consumed dried cranberries at least once a year were selected. For the cranberry juice survey, the study screened for individuals who were familiar with the different cranberry juice categories (100% cranberry juice, cranberry cocktail, and cranberry juice blend), as knowledge and experience with the product are necessary to mimic a real-life purchasing situation as closely as possible to provide accurate and unbiased estimates (Louviere, 2006). The Institutional Review Board approved the survey (Mississippi State University IRB–20–305).

1.3.2 Choice experiment design

This study used a discrete choice experiment (DCE) to elicit consumers' WTP for attributes of dried cranberries and cranberry juice. We used a cheap talk script (Champ et al., 2009) and a certainty scale to mitigate hypothetical bias (Hensher et al., 2012). The cheap talk script was presented to respondents in the explanation that preceded the DCE scenarios. An example of the

[2] The cranberry juice survey version included two sets of discrete choice experiments (DCE), one centering on an "unlabeled/generic" juice, where the DCE presented three alternatives: Option A, Option B, and Option C (no-buy option). These alternatives did not distinguish the type of juice. The DCE centered on "labeled" juice categories, where the DCE presented four alternatives: 100% juice, cocktail, juice blend, and the no-buy option. These later did distinguish between the three types of juice in the market. Findings from the unlabeled set are generally consistent with those for the labeled juice. To streamline the information presented in this manuscript, the "unlabeled" juice results are not reported but are available from the authors upon request.

complete set of descriptions presented to respondents is provided in Appendix 1.A for dried cranberries and Appendix 1.B for cranberry juice.

After each DCE scenario a follow–up certainty question was presented following Hensher et al. (2012) who proved that including questions to assess the extent to which a respondent is certain of actually choosing the DCE alternative mitigates the proneness to hypothetical bias. This study used the 1–10 certainty scale, where 1= “Very uncertain” and 10= “Very certain”. An example of the certainty scale used is also included in Appendices 1.A – 1.B.

A detailed description of the attributes and attribute levels included in both DCE versions (dried cranberries and cranberry juice) is presented in Table 1.1. The study included three attributes with two levels: total sugars, cranberry flavor, and cranberry breeding technologies. A description of each attribute was included in the survey, before the DCE; and can be found in Appendix 1.A for dried cranberries and Appendix 1.B for cranberry juice. The total sugar levels for the dried cranberries were presented as “Regular” and “50% Less Sugar” which are equivalent to 29 g. and 14 g. of sugars per serving, respectively. For cranberry juice, “Original” is equivalent to 25 g., and “50% Less Sugar” is equivalent to 12 g. of sugars. These levels were aligned with commercial products in the market. Instead of added sugar, total sugar was included because the study aimed to investigate if there were tradeoffs between total sugar content and the possibility of reducing it by applying CRISPR. By doing so, it is possible that the flavor intensity of cranberries is affected. Given that literature suggests flavor attributes are crucial for consumers’ acceptance, we included flavor with two levels, full/intense and bland/weak. Price was included with three levels for each product. For dried cranberries: $1.99, $2.99, and $3.99

per 6-oz bag, and for juice: $2.49, $2.99, and $3.99 per 64-fl oz bottle. These prices were consistent with prices of similar commercial products when the survey took place.

In the dried cranberries survey, respondents evaluated six hypothetical purchase scenarios. Each scenario consisted of two 6–oz bags of dried cranberries with varying attribute levels options (A and B) and a no–buy option. An example of a dried cranberries choice scenario is presented in Appendix 1.A. For the cranberry juice version, respondents were presented with six scenario choices with three cranberry juice options — "100% Juice," "Cocktail," and "Blend" — and a no–buy alternative. The narrative that preceded the juice DCE had a definition of each juice type following guidance from industry stakeholders. An example of the cranberry juice scenario choice is presented in Appendix 1.B.

1.3.3 Information treatments

Because the public is often exposed to different types of food information that could affect food preferences (Dutriaux et al., 2021), we tested how emphasizing different kinds of information scripts would impact the WTP for total sugars, flavor intensity, and plant breeding technology. We included four information treatments. Treatment 1 was the control, with no–information.

Treatment 2 presented a script on the health benefits of cranberries: *Cranberries are considered a superfood due to their high nutrient and anthocyanin content. Anthocyanins are substances that can prevent or slow damage to cells caused by free radicals. The anthocyanin properties of cranberries provide multiple health benefits, including the support of cardiovascular health and reduction of the risk of some cancers.* We hypothesize that the treatment 2 information would result in a higher WTP (decreased price discount) for regular sugar content, CRISPR breeding, and bland and weak cranberry flavor, compared to the control: H_{01}: $WTP^{treatment1} \leq WTP^{treatment2}$, H_{a1}: $WTP^{treatment1} > WTP^{treatment2}$.

Treatment 3 presented a script with the recommended sugar intake limit and the benefit of limiting sugar consumption: *The FDA defines "Added Sugars" as sugars that are added during the processing of foods. Added sugars increase calories without contributing important nutrients. The Dietary Guidelines for Americans recommend limiting the daily amount of added sugars consumed to no more than 10% of total calories per day (which is equivalent to 200 calories or 50 grams per day). Diets lower in sugar–sweetened foods are associated with a reduced risk of developing cardiovascular disease*[3]. We anticipate that the inclusion of the information in treatment 3 would lead to a lower WTP (increased price discount) compared to the control, for regular sugar content, CRISPR breeding, and bland and weak cranberry flavor: H_{02}: $WTP^{treatment1} \geq WTP^{treatment3}$, H_{a2}: $WTP^{treatment1} < WTP^{treatment3}$.

Treatment 4 included both sets of information provided in treatments 2 and 3. In this case, we expect that the health benefits information will counterbalance the impact of the dietary

[3] Note here that the study focuses on consumers' preference for total sugar content in cranberry products and the health information treatment explains the recommendation to limit added sugars, not total sugars. We chose this path for a couple of reasons. First, we based our health information treatment on recommendations found in the Dietary Guidelines for Americans (U.S. Department of Agriculture, Dietary Guidelines for Americans, 2020), which provide recommendations to limit calories from added sugars and avoid foods and beverages with added sugars, but do not include recommendations for total sugars. Second, information on added sugars is important given the new FDA's labeling rule requiring products to explicitly report added sugars on the NFP in addition to the total sugar content. Most of the sugars in dried cranberries and juice cocktails come from added sugars, as cranberries have minimal naturally occurring sugars. Thus, while consumers tend to focus on total sugar content (Tierney et al., 2017; Rampersaud et al., 2014), with CRISPR there is a potential to develop varieties low in acid which would result in lower sugar content in the form of less sugars added to improve palatability. However, we acknowledge that based on how the information treatment was presented, we cannot disentangle how respondents reacted to this information as we could capture mixed total sugar and added sugar avoidance reactions.

recommendation to limit added sugar consumption, resulting in the same willingness to pay as in the control (H_{03}: $WTP^{treatment1}=WTP^{treatment4}$, H_{a3}: $WTP^{treatment1}\neq WTP^{treatment4}$).

The text describing each information treatment was presented right before the DCE. An example of Treatment 4, which includes both treatments 2 and 3 scripts, is shown right before the DCE exhibits in Appendices 1.A for dried cranberries and 1.B for cranberry juice. A between-subjects design was used for all survey versions. Respondents were randomly assigned to each information treatment.

The survey included questions about respondents' sociodemographic characteristics such as gender, age, racial–ethnicity, education, income, number of people in the household, presence of children, self–reported health status, and diet-related chronic disease diagnoses. The survey also included questions to gauge preferences for cranberry product attributes, food purchase habits, and respondents' use of NFP labels, use of information on the NFP label, and a heat map for respondents to identify the piece of information on the NFP most important to them. In addition, we included questions to measure whether respondents correctly interpreted the added sugar line on the NFP. We also asked questions to assess perceptions on using new technologies for food production and processing, plant breeding technologies (specifically genetic engineering versus CRISPR), and the level of trust on different information sources related to food.

1.3.4 Empirical approach

The current study's empirical approach stems from the demand theory by Lancaster (1966) and the random utility model by McFadden (1974). The demand theory states that consumers derive utility from the attributes inherent to a good rather than the good itself. At the same time, the random utility model postulates that consumers' utility can be explained by a deterministic component given by the good's attributes and a random component given by unobserved factors.

This study estimated the models in WTP space using the Generalized Multinomial Logit Model (GMNL) proposed by Fiebig et al. (2010). The GMNL models allow for scale heterogeneity and preference heterogeneity. Scale heterogeneity is defined as the variance in the degree of randomness between respondents in the decision–making process. Following Fiebig et al. (2010), The general specification of the GMNL model is as follows,

$$U_{ni} = [\sigma_n \beta + \gamma \eta_n + (1-\gamma)\sigma_n \eta_n] x_{ni} + \varepsilon_{ni}$$

$$\beta_n = \sigma_n \beta + [\gamma + (1-\gamma)\sigma_n]\eta_n \tag{1.1}$$

where σ_n is the individual–specific scale of the idiosyncratic error term that captures scale heterogeneity and is log-normally distributed with mean $\bar{\sigma}$ and standard deviation τ. η_n is a vector of individual specific taste deviate from the mean based on observed attributes, and it captures residual preference heterogeneity; and γ is a parameter between 0 and 1 that controls how the variance of residual taste heterogeneity η_n varies with the scale heterogeneity σ_n.

We estimated the different model formulations encompassed by the GMNL: the Type II, where $\gamma = 0$ (GMNL–II) and Type I where $\gamma = 1$ (GMNL–I) to the Random Parameter Logit (RPL) model. In the GMNL–I model, the standard deviation of residual taste heterogeneity is independent of the scale, whereas in the GMNL–II model, it is proportional to the scale. The RPL model is a special case of the GMNL model where the scale of the error term, σ_n, is normalized to 1 (Fiebig et al., 2010). All models were estimated using the "gmnl" package in R 4.0.5 (Sarrias and Daziano, 2017). After comparing goodness of fit indicators [Akaike

Information Criterion (AIC), Bayesian Information Criterion (BIC), and the likelihood functions] of the models estimated, we found that the GMNL–II model outperforms the GMNL–I and RPL models. Thus, we report the results of the GMNL–II model.

Following the general form of the GMNL and the attributes of the cranberry products in our study, we can write the utility respondent *n* derives from choosing alternative *i* as:

$$U_{ni} = \beta_{ASC_n} ASC_n + \sigma_n(-Price + \beta_{RS} RegularSugar + \beta_{BF} BlandFlavor + \beta_{CRISPR} CRISPR + L\eta_n) + \varepsilon_{ni} \quad (1.2)$$

where $Price$ is a continuous variable that takes any of the three values in the experimental design, and whose coefficient takes a fixed value of –1. The coefficient of *Price* is normalized to -1 so that the attribute coefficients can be directly interpreted as WTP values (Sarrias and Daziano, 2017). $RegularSugar$ is a binary variable indicating the product has a regular sugar content, $BlandFlavor$ is a binary variable indicating the product has a bland/weak cranberry flavor intensity (flavor was described as the overall combination of sensations and its influence by taste, aroma, look and texture), $CRISPR$ is a binary variable indicating that the product is made from CRISPR cranberries, β_{ASC_n} is the alternative–specific constant (ASC), *L* is the lower triangular matrix of the Cholesky decomposition, and η_n follows a standard normal distribution.

We allowed for scale heterogeneity in the scale parameter σ_n across individuals based on their stated choice certainty level (Kunwar, Bohara and Thacher 2020), such that:

$$\sigma_n = exp\ (\delta certain_n + \tau\nu_n) \quad (1.3)$$

where δ is the parameter of the observed heterogeneity in the scale term, τ is the coefficient on the unobserved scale heterogeneity, $\nu_n \sim N(0,1)$, and $certain_n$ is an indicator variable equal to 1 if individual n's level of certainty after responding each choice scenario is greater or equal to 7 and 0 otherwise[4].

Marginal rate of substitution

To investigate the tradeoffs between having a product with regular sugar content and acceptance of CRISPR technology, we estimated the marginal rate of substitution (MRS) between regular sugar content and CRISPR:

$$MRS_{RS,CRISPR,n} = \frac{\beta_{RS,n}}{\beta_{CRISPR,n}} \tag{1.4}$$

This ratio captures the relative discount for products with regular sugar content versus CRISPR-bred cranberries. This study approximated the variance of the trade-off using the delta method, which is often employed to approximate the variance of the ratio of two random variables. The Delta method can be viewed as a generalized central limit theorem that asymptotically approximates normal random variables using the Taylor series (Casella and Berger, 2021). Based on the following equation, the variance of the trade-off can be estimated using the delta method:

$$Var(\frac{X}{Y}) \approx \left(\frac{\mu_X}{\mu_Y}\right)^2 (\frac{Var(X)}{\mu^2{}_X} + \frac{Var(Y)}{\mu^2{}_Y} - 2\frac{Cov(X,Y)}{\mu_X\mu_Y}) \tag{1.5}$$

A coefficient closer to 1 denotes that respondents were indifferent between having a product with regular sugar content and the breeding method CRISPR, a coefficient > 1 indicates that the aversion to regular sugar is larger than the aversion to CRISPR, and a coefficient < 1 suggests

[4] 14.93%, 15.27%, 15.47%, and 11.67% of responses in treatment 1–4 have a certainty level less than 7 for the dried cranberries survey, respectively; 13.96%, 17.60%, 16.18%, and 19.02% of responses in treatment 1– 4 have a certainty scale less than 7 for the cranberry juice survey.

that the aversion to CRISPR is larger than the aversion to regular sugar content in the cranberry products included in this study.

Compensating surplus

To further understand respondents' preferences for cranberry products with different combination of attributes levels, we computed the compensating surplus (CS). This represents the welfare change for consumers when going from a base option to an improved hypothetical scenario. Following Britwum and Yiannaka (2019) and Espinosa-Godedet al. (2010), CS is defined as:

$$Compensating\ surplus = -\left(\frac{1}{\beta_{price}}\right)(V_1 - V_2) \tag{1.6}$$

where V_1 is the conditional indirect utility of the base option, V_2 is associated with the hypothetical option which represents the alternative with the change. The base option and hypothetical option are described in Table 1.2. The base option for both cranberry products (dried cranberries and cranberry juice) is defined as cranberry products manufactured from conventionally bred cranberries, with regular sugar content and a weak cranberry flavor. CRISPR-bred cranberries with reduced sugar content and full cranberry flavor are the hypothetical alternative. We solely used data from the control treatment (no additional information) group in order to rule out any potential impacts of the information treatment. In our GMNL-II certainty model, the parameter of *price* is fixed at -1, thus the economic surplus becomes:

$$Compensating\ surplus = V_1 - V_2 = \Delta\widehat{V_\iota}. \tag{1.7}$$

Latent class model

A latent class analysis was performed to investigate the factors triggering the heterogeneity in WTP estimates for the cranberry products' attributes. The model assumes unobservable characteristics are captured by class membership variables or respondents' socioeconomic characteristics, cranberry and food purchase habits, knowledge, and perceptions of plant breeding methods (genetic engineering and CRISPR)[5].

The latent class model captures the heterogeneous preferences by identifying segments within the sample of survey respondents, namely classes. Accordingly, individuals were grouped into several latent classes or unobservable subgroups. Preferences across classes are heterogeneous, but choices within each class are homogeneous. The mathematical formulation of the latent class model can be found in Greene and Hensher (2003).

To identify the number of classes, this study used a set of indicators including measures of goodness of fits such as AIC, BIC, and likelihood function; the best–fitting model is the one with the smaller AIC and BIC. Other criteria include the interpretability of results and classification diagnosis. The latter ensures that selected classes are not an expanded version of the other

---

[5] These variables were selected by running three ordinary least square regressions having the WTP for gene editing, regular sugar, and flavor as dependent variables and all responses to questions asked in the survey. The variables selected were the ones that were consistently statistically significant for all three regressions. Specific variables included: a binary variable equaling 1 if the income was higher or equal the sample average at $87.500/year; binary variable equaling 1 if the respondent indicated that the added sugar information on the NFP was important or crucial; binary variable equaling 1 if the respondent interpreted correctly the total sugar and the added sugar information on the NFP; binary variable equaling 1 if the respondent attributed their highest attention to the total sugar content on the NFP on a heat map question; binary variable equaling 1 if the respondent indicated that they liked extremely an intense cranberry flavor; binary variable equaling 1 if the respondent indicated that health was important/crucial when buying cranberry products; binary variable equaling 1 if the ingredient list was important/crucial when buying cranberry products; binary variable equaling 1 if they consider that CRISPR and GMO are different and they know the difference; binary variable equaling 1 if they consider that CRISPR and GMO are different but they don't know the difference; binary variable equaling 1 if they consider there are no differences between CRISPR and GMO; binary variable equaling 1 if they are willing to purchase CRISPR food if the breeding method information is the only information known; binary variable equaling 1 if they are willing to purchase CRISPR food if this increases insect resistance and herbicide tolerance; binary variable equaling 1 if they are willing to purchase CRISPR food if this reduces the environmental impact of food production, binary variable equaling 1 if they are willing to purchase CRISPR food if this increases nutrient content in food, binary variable equaling 1 if they are willing to purchase CRISPR food if this reduces the need to add sugars in food processing.

(Nylund-Gibson and Choi, 2018). The latent class models were estimated in R 4.0.5 using the package “gmnl” developed by (Sarrias and Daziano, 2017).

1.4 Results

Table 1.3 presents the sociodemographic characteristics of the two survey versions, dried cranberries, and cranberry juice, across four information treatment groups. Almost all groups of respondents were comparable to the general U.S. population regarding gender and income (U.S. Census Bureau, 2020). The proportion of respondents with at least a four–year college degree in our sample was higher than the U.S. population, which is consistent with the profile of those who are more responsive to surveys in general (Curtin et al., 2000). Moreover, compared to the general U.S. population, our sample contained a higher proportion of respondents with at least one child.

To ensure the sample of respondents across different information treatments was comparable, we used a pairwise t-test to examine statistical differences in salient sociodemographic characteristics across the treatment groups. We found that respondents in the two survey versions and across treatments were reasonably similar regarding gender, age, education, and income (Table 1.3). Differences were observed in the cranberry juice survey sample, where the treatment four subsamples exhibited a higher proportion of respondents with larger family sizes (≥ 3 members) and at least one child in their households compared to the group responding to treatments 1−3 (Table 1.3).

1.4.1 Willingness–to–pay results

All WTP models reported were estimated with unscaled random alternative specific constants (ASCs), correlated parameters, and choice certainty on the scale parameter.

Dried cranberries

Across all treatments, respondents stated their willingness to discount the price of dried cranberries with regular compared to reduced sugar content (Table 1.4). The price discount ranged from $2.33 to $3.85. The information on the health benefits of cranberries -treatment 2- did not impact the WTP (fail to reject the null hypothesis). Conversely, the price discount for regular sugar increased under treatment 3 -information on the recommendation to limit sugar consumption- and 4 -health benefits and dietary effects of reducing sugar intake- (reject both null hypotheses). This coincides in part with McFadden et al. (2021), who concluded that the information with negative connotations is more impactful than positive ones.

Respondents also consistently stated a discount for CRISPR compared to conventional-bred cranberries, ranging from $1.43 to $2.12 across information treatments. This finding is consistent with previous literature in which consumers favor conventional breeding over gene editing (An et al., 2019; Marette et al., 2021; Muringai et al., 2020; Shew et al., 2018; Yang and Hobbs, 2020). These results differ from Hu et al. (2022), who found that respondents stated similar WTP for juice from gene-edited and conventionally bred oranges in the absence of information. Also, there were no statistically significant differences between the discount for CRISPR under the control and the different information treatments. This finding differs from studies concluding that information affected the WTP for CRISPR-bred foods (Paudel et al., 2023; Kilders and Caputo, 2021; Hu et al., 2022).

Importantly, the magnitude of the discount for regular sugar was more significant than the magnitude of the discount for CRISPR. This was further emphasized in the marginal rate of substitution (Figure 1.1) since we observed that the aversion towards products with regular sugar content was larger than that towards foods bred using CRISPR. This result is promising for the scientific community employing this new breeding technology in agriculture, as it implies that the aversion towards this new breeding method could be mitigated by offering consumers a product with reduced sugars.

Considering the magnitude of the WTP estimates, respondents placed flavor intensity as more important than the total sugar content when no information was provided and when both sets of information (health benefits and dietary effects of reducing sugar intake) were provided. This coincides with literature stating that consumers usually prioritize taste over health when purchasing foods (Malone and Lusk, 2017). In addition, our results indicate that when consumers see sugar-related health information, they are willing to trade off a weaker flavor for lower sugar content. However, if no information about the need to limit sugar consumption is presented or when information that may counteract the health-related sugar message is presented (e.g., benefits from consuming cranberries), consumers are unwilling to trade off a weaker flavor for reduced sugar content.

The opt-out ASCs were negative across all information treatments indicating respondents prefer the cranberry product alternatives over the no–buy option. The standard deviations of the random parameters and the standard deviation of the scale parameter, τ, were all statistically significant, indicating preference heterogeneity across respondents, and demonstrating the importance of considering variations in preferences. Also consistent with findings in Kilders and

Caputo (2021) with more information -comparing treatment control with treatment 4- the standard deviation of the mean WTP for reduced sugar and CRISPR increased, indicating that more information increased heterogeneity in responses. However, this is not consistent across the type of information. For example comparing control with treatment 3 (effects of sugars on diet) the standard deviation of the WTP for reduced sugar and CRISPR decreases, indicating that this information leads to less heterogeneity in responses.

The parameter estimate for the certainty scale variable was statistically significant, although the results were inconsistent across information treatments. The negative sign associated with certainty meant that respondents who were certain about their choices made more stochastic choices. The literature offers no concluding findings on what should be the sign of this parameter. Beck et al. (2013) and Kunwar et al. (2020) found that respondents who marked they were certain to make more deterministic choices. Conversely, Rahman and Bohara (2023) reported a positive sign for respondents were both certain and uncertain about their choices. These inconsistencies may be attributed to differences in the sample of respondents.

Cranberry juice

Similar to dried cranberries, respondents stated a price discount for regular sugar content ranging from \$1.23 to \$2.04 (Table 1.5). Only, under treatment 4, when presenting both sets of information -cranberry health benefits and dietary effects of limiting sugar intake-, the price discount significantly increased from \$1.23 to \$1.61 (reject the null hypothesis). Consistent with results from the dried cranberry survey, respondents stated a price discount for CRISPR that ranged from \$1.05 to \$2.33. Compared to the control treatment, the price discount for CRISPR was statistically larger when presenting information on the dietary effects of sugar intake (treatment 3) and both health benefits and dietary effects of sugar intake (treatment 4). This implies that accessing information increases expectations for cranberry products, increasing the aversion to the new CRISPR technology.

Consistent with findings from the dried cranberry survey, the cranberry juice survey respondents assigned higher importance to flavor intensity compared to regular sugar content and breeding method—across all treatments. Recall that flavor in this survey was described as the overall combination of sensations influenced by the taste, aroma, look, and texture. Because of the dilutions, the preference for an intense cranberry flavor is more evident for juices than dried cranberries. No clear pattern was observed in the effect of information on the discount for flavor intensity.

Similar to the dried cranberry survey models, the standard deviations of the parameters were statistically significant, denoting heterogeneity across respondents. The standard deviation of the scale parameter, τ, was statistically significant, and the parameter estimate for the certainty scale variable was negative and statistically significant. Here, with some exceptions, the additional information also increases the magnitude of the standard deviation of the WTP for reduced sugar content and CRISPR, leading us to conclude that more information increased heterogeneity in the WTP for these two attributes. This result coincides with Kilders and Caputo (2021).

The opt-out ASCs for each juice label were positive and statistically significant, implying that respondents preferred each juice alternative over the no–buy option. In addition, the 100% juice was chosen over the cocktail and blend options.

Consistently, the standard deviations of the parameters were statistically significant, denoting heterogeneity across respondents, the standard deviation of the scale parameter, τ, was statistically significant, and the parameter estimate for the certainty scale variable was negative and statistically significant. Similar to dried cranberries, the additional information also increases the magnitude of the standard deviation of the WTP for reduced sugar content and CRISPR, leading us to conclude that more information increased heterogeneity in the WTP for these two attributes.

Compensating surplus results

We estimate the compensating surplus for products with different attribute levels and find that respondents were willing to pay an overall price premium for a cranberry product made from CRISPR-bred cranberries, with reduced sugar content, and full/intense cranberry flavor relative to a product made with conventionally-bred berries, regular sugar content and weak/bland flavor. Interestingly a higher premium was observed for dried cranberries ($3.90), compared to cranberry juice ($2.47). This implies that while the CRISPR attribute alone is disfavored by respondents, when it (CRISPR) is presented as part of a bundle of desired attributes such as reduced sugars and full/intense flavor, respondents were willing to pay a price premium for the desired bundle. In other words, respondents were willing to buy the desired bundle as long as its price did not exceed the baseline prices (representing a product made with conventionally-bred berries, regular sugar content, and weak/bland flavor) of $3.90 for dried cranberries and $2.47 for cranberry juice. These insights suggest that when breeding methods such as CRISPR result in products with preferred product attributes, consumers may be willing to accept these products if the benefits offered offset consumer’s discount for CRISPR.

1.4.2 Latent class model results

To avoid confounding with information treatment effects, we only used the observations from the control treatment (no additional information) group in the latent class analyses. We opted for three classes across all regressions, as these models exhibit the lower values for the AIC and the BIC, ensuring the interpretability of results and the number of statistically significant parameter estimates in each class. Table 1.6 presents the measures of goodness of fit used as part of the criteria to select the number of classes. The three latent classes identified varied in the acceptance/rejection of the different attributes of dried and cranberry juice (Table 1.7 and Table 1.8). Concerning the acceptance of CRISPR, we found that for dried cranberries, a group was willing to pay a price premium for CRISPR compared to conventional breeding. This group stated they would purchase CRIPSR food if this reduced the need to add sugars in food processing. Also, this group was the least to correctly interpret the difference between total and added sugars and paid the least attention to total sugar content on the NFP.

For cranberry juice, one observes three segments of respondents: strong CRISPR rejection (class 1), mild CRISPR rejection (class 2) and the indifferent group (class 3). The indifferent group would display a larger proportion of respondents (compared to those who strongly reject CRISPR) with income ≥$87,500/year, larger proportion of respondents who know that CRISPR and GMO are different and they know the difference. The latter result is aligned with McFadden et al. (2021), in that there is some connection between the association of CRISPR to Genetic Modification and the acceptance of CRISPR. Interestingly, the group that shows a mild rejection to CRISPR had a larger proportion of respondents who indicated that they would be willing to purchase CRISPR food is this reduces the need to add sugars in food processing.

1.5 Conclusions and Implications

Given the potential for abundant CRISPR applications to improve crops (and beyond), this study investigated respondents' WTP for this technology, considering that the benefit will be a product with reduced sugar content. Specifically, we examined respondents' WTP for regular sugar content (vs. reduced sugar content) and a product produced with cranberries developed using gene editing CRISPR (vs. conventional breeding). We examined cranberry products (dried cranberries and cranberry juice) because, despite their health benefits, cranberry products could be high in sugars— added by the industry to make them palatable. CRISPR could be used to develop cultivars with desired traits in terms of decreased acidity or increased natural sugar content). In general, across the three cranberry products evaluated, respondents stated willingness to discount the price for cranberries bred using CRISPR compared to conventional breeding, which is consistent with most literature (An et al., 2019; Marette et al., 2021; Muringai wt al., 2020; Shew et al., 2018; Yang and Hobbs, 2020), with some exceptions (Hu et al., 2022).

Participants were also willing to discount the price for cranberry products with regular sugar content compared to reduced sugar and for products with weak/bland flavor compared to full/intense flavor. The overall results were consistent even after presenting information scripts either emphasizing the health benefits of cranberries, the dietary effects of limiting sugar intake, or both. These findings differ from the literature, concluding that additional information impacts the WTP for CRISPR-bred foods (Hu et al., 2022; Kilders and Caputo, 2021; Paudel et al., 2023).

When analyzing the entire product, compensating surplus analyses indicate that consumers would be willing to pay a price premium for cranberry products that exhibit a reduced sugar content, are CRISPR-bred, and display a full/intense flavor relative to products with conventionally bred fruit but with less preferred attributes (i.e., regular sugar content and weak flavor). Respondents were heterogeneous in their preferences for CRISPR-bred cranberries. Consistently across the three cranberry products, those willing to pay a price premium for CRISPR-bred or those indifferent between CRISPR and conventional-bred cranberries stated they would purchase CRISPR food if this reduces the need to add sugars in food processing. This emphasizes the need to increase public awareness of the benefits of applying CRISPR to unaware population segments and those who believe that gene editing is another iteration of genetic modification.

Our results contribute to the scientific community interested in knowing how receptive consumers would be to new plant breeding technologies. The literature shows that consumers would be more acceptant if these technologies directly benefited them. This study shows that respondents were more reluctant to have a product with regular sugar content than a product using CRISPR-bred cranberries, as evidenced by the marginal rates of substitution between regular sugar content and CRISPR. Further, we show that respondents would be willing to pay a price premium for all three cranberry processed products if they exhibit a reduced sugar content, a full/intense cranberry flavor, and are CRISPR-bred. This study contributes to the food industry and policymakers’ understanding of food choice drivers and could help inform the design of strategies and policies that will lessen consumers' pessimistic perceptions about novel breeding technologies, particularly when these technologies could lead to healthier food alternatives.

As a final point, a limitation of this study is the discrepancy in our goal to estimate WTP for reduced total sugars and the information treatment that mentions added sugars. We based the decision to mention added sugars in the information treatment, following the Dietary Guidelines for Americans (U.S. Department of Agriculture, Dietary Guidelines for Americans, 2020), that recommend limiting calories from added sugars and avoid foods and beverages with added sugars, but do not include recommendations for total sugars. Also, information on added sugars is more relevant considering the new FDA's labeling rule requiring products to explicitly report added sugars on the NFP in addition to the total sugar content. Moreover, literature suggests that consumers tend to focus on total sugar content more than added sugars (Tierney et al., 2017; Rampersaud et al., 2014). CRISPR offers the feasibility to develop cranberry varieties low in acid which would result in lower total sugar content reducing the need to add sugars. However, we acknowledge that based on how the information treatment was presented, we are unable to disentangle how respondents reacted to this information as we could be capturing mixed total sugar and added sugar avoidance reactions. Future research should consider assessing the dynamics of total sugar and added sugar labeling and the effect of health-related information on consumers perceptions.

REFERENCES

Agricultural Marketing Resource Center (2023) Cranberries. Retrieved from https://www.agmrc.org/commodities-products/fruits/cranberries. Accessed October 17, 2023

Alston, J. M., & Pardey, P. G. (2021). The economics of agricultural innovation. Handbook of agricultural economics, 5, 3895-3980.

An, H., P. Lloyd-Smith & W.L. Adamowicz (2019) Strategic Behavior in Stated Preferences and the Demand for Gene-edited Canola. In *Agricultural and Applied Economics Association Annual Meeting*. Atlanta, Georgia, United States.

Beck, M.J., J.M. Rose & D.A. Hensher (2013) Consistently inconsistent: The role of certainty, acceptability and scale in choice. *Transportation Research Part E: Logistics and Transportation Review,* 56, 81-93.

Britwum, K. & A. Yiannaka (2019) Consumer willingness to pay for food safety interventions: The role of message framing and issue involvement. *Food Policy,* 86, 101726.

Champ, P.A., R. Moore & R.C. Bishop (2009) A Comparison of Approaches to Mitigate Hypothetical Bias. *Agricultural and Resource Economics Review,* 38(2), 166-180.

Critchley, C., D. Nicol, G. Bruce, J. Walshe, T. Treleaven & B. Tuch (2018) Predicting Public Attitudes Toward Gene Editing of Germlines: The Impact of Moral and Hereditary Concern in Human and Animal Applications. *Front Genet,* 9, 704.

Curtin, R., S. Presser & E. Singer (2000) The effects of response rate changes on the index of consumer sentiment. *Public Opin Q,* 64(4), 413-28.

Dannenberg, A. (2009) The dispersion and development of consumer preferences for genetically modified food — A meta-analysis. *Ecological Economics,* 68(8-9), 2182-2192.

Dewey, C. (2018) The Future of Food. Scientists Have Found a Fast and Cheap way to Edit your Food's DNA.

Doudna, J.A. & E. Charpentier (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213), 1258096.

Dutriaux, L., E.K. Papies, J. Fallon, L. Garcia-Marques & L.W. Barsalou (2021) Incidental exposure to hedonic and healthy food features affects food preferences one day later. *Cogn Res Princ Implic,* 6(1), 78.

Entine, J., M.S.S. Felipe, J.-H. Groenewald, D.L. Kershen, M. Lema, A. McHughen, A.L. Nepomuceno, R. Ohsawa, R.L. Ordonio, W.A. Parrott, H. Quemada, C. Ramage, I. Slamet-Loedin, S.J. Smyth & D. Wray-Cahen (2021) Regulatory approaches for genome edited agricultural plants in select countries and jurisdictions around the world. *Transgenic Research,* 30(4), 551-584.

Espinosa-Goded, M., J. Barreiro-Hurlé & E. Ruto (2010) What Do Farmers Want From Agri-Environmental Scheme Design? A Choice Experiment Approach. *Journal of Agricultural Economics,* 61(2), 259-273.

Fang, D., R.M. Nayga, H.A. Snell, G.H. West & C. Bazzani (2019) Evaluating USA's New Nutrition and Supplement Facts Label: Evidence from a Non-hypothetical Choice Experiment. *Journal of Consumer Policy,* 42(4), 545-562.

Fiebig, D.G., M.P. Keane, J. Louviere & N. Wasi (2010) The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. *Marketing Science*, 29(3), 393-421.

Greene, W.H. & D.A. Hensher (2003) A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological,* 37(8), 681-698.

Hall, S.S. (2016) New Gene-Editing Techniques Could Transform Food Crops-or Die on the Vine.

Haspel, T. (2018) The Public Doesn't Trust GMO's. Will it Trust CRISPR?

Hensher, D.A., J.M. Rose & M.J. Beck (2012) Are there specific design elements of choice experiments and types of people that influence choice response certainty? *Journal of Choice Modelling,* 5(1), 77-97.

Hu, Y., L.A. House & Z. Gao (2022) How do consumers respond to labels for crispr (gene-editing)? *Food Policy,* 112, 102366.

Kalaitzandonakes, N., J. Lusk & A. Magnier (2018) The price of non-genetically modified (non-GM) food. *Food Policy,* 78(C), 38-50.

Khandpur, N., E.B. Rimm & A.J. Moran (2020) The Influence of the New US Nutrition Facts Label on Consumer Perceptions and Understanding of Added Sugars: A Randomized Controlled Experiment. *J Acad Nutr Diet,* 120(2), 197-209.

Kilders, V. & V. Caputo (2021) Is Animal Welfare Promoting Hornless Cattle? Assessing Consumer's Valuation for Milk from Gene-edited Cows under Different Information Regimes. *Journal of Agricultural Economics,* 72(3), 735-759.

Kim, E.J., B. Ellison, B. McFadden & M.P. Prescott (2021a) Consumers' decisions to access or avoid added sugars information on the updated Nutrition Facts label. *PLoS One,* 16(3), e0249355.

Kim, E.J., B. Ellison, M.P. Prescott & R.M. Nayga, Jr. (2021b) Consumer Comprehension of the Nutrition Facts Label: A Comparison of the Original and Updated Labels. *Am J Health Promot,* 35(5), 648-657.

Kunwar, S.B., A.K. Bohara & J. Thacher (2020) Public preference for river restoration in the Danda Basin, Nepal: A choice experiment study. *Ecological Economics,* 175, 106690.

Lancaster, K.J. (1966) A New Approach to Consumer Theory. *Journal of Political Economy,* 74(2), 132-157.

Laquatra, I., K. Sollid, M. Smith Edge, J. Pelzel & J. Turner (2015) Including "Added Sugars" on the Nutrition Facts Panel: How Consumers Perceive the Proposed Change. *J Acad Nutr Diet,* 115(11), 1758-63.

Lee, J.Y., M.P. Popp, E.J. Wolfe, R.M. Nayga, Jr., J.S. Popp, P. Chen & H.S. Seo (2018) Information and order of information effects on consumers' acceptance and valuation for genetically modified edamame soybean. *PLoS One,* 13(10), e0206300.

Lusk, J., B. McFadden & B. Rickard (2015) Which biotech foods are most acceptable to the public? *Biotechnology Journal,* 10(1), 13-16.

Lusk, J.L., L.O. House, C. Valli, S.R. Jaeger, M. Moore, J.L. Morrow & W.B. Traill (2004) Effect of information about benefits of biotechnology on consumer acceptance of genetically modified food: evidence from experimental auctions in the United States, England, and France. *European Review of Agricultural Economics,* 31(2), 179-204.

Lusk, J.L., M. Jamal, L. Kurlander, M. Roucan & L. Taulman (2005) A Meta-Analysis of Genetically Modified Food Valuation Studies. *Journal of Agricultural and Resource Economics,* 30.

Malone, T. & J. Lusk (2017) Taste trumps health and safety: Incorporating consumer perceptions into a discrete choice experiment for meat. *Journal of Agricultural and Applied Economics,* 49, 1-19.

Marette, S., A.-C. Disdier & J.C. Beghin (2021) A comparison of EU and US consumers' willingness to pay for gene-edited food: Evidence from apples. *Appetite,* 159, 105064.

McFadden, B.R., B.N. Anderton, K.A. Davidson & J.C. Bernard (2021) The effect of scientific information and narrative on preferences for possible gene-edited solutions for citrus greening. *Applied Economic Perspectives and Policy,* 43(4), 1595-1620.

McFadden, D. (1974) Conditional Logit Analysis of Qualitative Choice Behavior, in P. Zarembka (ed.) Frontiers in Econometrics, 105o142. Academic Press, New York.

Menz, J., D. Modrzejewski, F. Hartung, R. Wilhelm & T. Sprink (2020) Genome Edited Crops Touch the Market: A View on the Global Development and Regulatory Environment. *Frontiers in Plant Science,* 11(1525).

Muringai, V., X. Fan & E. Goddard (2020) Canadian consumer acceptance of gene-edited versus genetically modified potatoes: A choice experiment approach. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie,* 68(1), 47-63.

Nes, K., K.A. Schaefer & D. P. Scheitrum (2022) Global Food Trade and the Costs of Non-Adoption of Genetic Engineering. *American Journal of Agricultural Economics,* 104(1), 70-91.

Neuhofer, Z., B.R. McFadden, A. Rihn, X. Wei, H. Khachatryan & L. House (2020) Can the updated nutrition facts label decrease sugar-sweetened beverage consumption? *Econ Hum Biol,* 37, 100867.

Nylund-Gibson, K. & A.Y. Choi (2018) Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science,* 4, 440-461.

Parrott, W.A. (2022) Gene-Edited Assisted Plant Breeding and the Regulatory Policy Regulating Its Use in the U.S. and Around the World. American Society for Horticultural Science.

Paudel, B., D. Kolady, D. Just & M. Ishaq (2023) Effect of information and innovator reputation on consumers' willingness to pay for genome-edited foods. *Food Quality and Preference,* 107, 104825.

Qaim, M. (2020) Role of New Plant Breeding Technologies for Food Security and Sustainable Agricultural Development. *Applied Economic Perspectives and Policy,* 42(2), 129-150.

Rahman, M.M. & A.K. Bohara (2023) Assessing the preference and spatial dependence of a solid waste management system in Nepal: A choice experiment approach. *Journal of Environmental Management,* 327, 116805.

Rampersaud, G. C., Kim, H., Gao, Z., & House, L. A (2014) Knowledge, perceptions, and behaviors of adults concerning nonalcoholic beverages suggest some lack of comprehension related to sugars. Nutrition Research, 34(2), 134-142.

Sarrias, M. & R. Daziano (2017) Multinomial Logit Models with Continuous and Discrete Individual Heterogeneity in R: The gmnl Package. *Journal of Statistical Software,* 79(2), 1 - 46.

Shew, A.M., L.L. Nalley, H.A. Snell, R.M. Nayga & B.L. Dixon (2018) CRISPR versus GMOs: Public acceptance and valuation. *Global Food Security,* 19, 71-80.

Tierney, M., A.M. Gallagher, E.S. Giotis & K. Pentieva (2017) An Online Survey on Consumer Knowledge and Understanding of Added Sugars. *Nutrients,* 9(1).

U.S. Census Bureau (2020) Quick Facts United States. Retrieved from https://www.census.gov/quickfacts/fact/table/US/SEX255221. Accessed October 17, 2023.

U.S. Department of Agriculture (2020) Dietary Guidelines for Americans, 2020-2025. Retrieved from https://www.dietaryguidelines.gov/sites/default/files/2020-12/Dietary_Guidelines_for_Americans_2020-2025.pdf . Accessed October 17, 2023.

Uddin, A., R.K. Gallardo, B. Rickard, J. Alston & O. Sambucci (2022) Consumer acceptance of new plant-breeding technologies: An application to the use of gene editing in fresh table grapes. *PLOS ONE*, 17(12), e0270792.

Voytas, D.F. & C. Gao (2014) Precision Genome Engineering and Agriculture: Opportunities and Regulatory Challenges. *PLOS Biology*, 12(6), e1001877.

Yang, Y. & J.E. Hobbs (2020) The Power of Stories: Narratives and Information Framing Effects in Science Communication. *American Journal of Agricultural Economics*, 102(4), 1271-1296.

Yeh, D.A., M.I. Gomez & H.M. Kaiser (2019) Signaling impacts of GMO labeling on fruit and vegetable demand. *PLoS One*, 14(10), e0223910.

Zhang, Y., K. Massel, I.D. Godwin & C. Gao (2018) Applications and potential of genome editing in crop improvement. *Genome Biol*, 19(1), 210.

Table 1.1: List of attributes and attribute levels for sets of discrete choice experiment scenarios for dried cranberries, and cranberry juice.

Attributes	Alternative possibilities available for each attribute	
	Dried cranberry survey version	Cranberry juice survey version
Total sugars (per serving size 1 cup)	• 29 g • 14 g	• 25 g • 12 g
Cranberry flavor	• Full/intense • Bland/weak	• Full/intense • Bland/weak
Cranberry breeding technology	• Conventional breeding • Gene editing	• Conventional breeding • Gene editing
Price	• $1.99/6–oz bag • $2.99/6–oz bag • $3.99/6–oz bag	• $2.49/64–fl oz bottle • $2.99/64–fl oz bottle • $3.49/64–fl oz bottle

Table 1.2: Description of base cranberry product option and hypothetical option used in compensating surplus analysis.

	Dried cranberries	Cranberry juice
Base option	A 6–oz bag of dried cranberries made from conventional bred cranberries, regular sugar content, weak/bland cranberry flavor, priced at $2.99	A 64 fl-oz bottle of cranberry juice made from conventional bred cranberries, regular sugar content, weak/bland cranberry flavor, priced at $2.99
Hypothetical option	A 6–oz bag of dried cranberries made from CRISPR bred cranberries, reduced sugar content, full/intense cranberry flavor, priced at $2.99	A 64 fl-oz bottle of cranberry juice made from CRISPR bred cranberries, reduced sugar content, full/intense cranberry flavor, priced at $3.99

Table 1.3: Summary statistics of respondents' sociodemographic characteristics for the two surveys (dried cranberries and cranberry juice) and four treatments of information

	Description	U.S. Census 2020[1]	Dried cranberries				Cranberry juice				Pairwise comparison between treatments (t–stat)					
			Treatment				Treatment				Dried cranberries Treatments			Cranberry juice Treatments		
			1	2	3	4	1	2	3	4	1–2	1–3	1–4	1–2	1–3	1–4
Gender	Female	0.51	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.00	0.00	0.00	0.00	0.00	0.00
Age	24 years or less	0.32	0.11	0.14	0.15	0.15	0.13	0.12	0.14	0.13						
	25-34 years	0.14	0.30	0.26	0.25	0.26	0.27	0.29	0.27	0.27						
	35-44 years	0.13	0.25	0.28	0.26	0.28	0.28	0.32	0.26	0.28						
	45-54 years	0.13	0.20	0.17	0.18	0.17	0.16	0.12	0.18	0.16						
	55-64 years	0.13	0.05	0.04	0.06	0.06	0.05	0.06	0.05	0.08						
	65+ years	0.16	0.10	0.11	0.09	0.09	0.10	0.10	0.10	0.07						
	Mean		40.69	40.24	40.02	39.82	40.14	39.86	40.26	39.99	0.35	0.51	0.67	0.21	0.09	0.11
College	1 if college degree	0.32	0.58	0.56	0.56	0.60	0.49	0.53	0.56	0.48	0.36	0.36	0.54	0.89	1.43	0.18
Income	1 if <$25,000/year	0.16	0.19	0.13	0.14	0.20	0.17	0.23	0.20	0.21	1.82**	1.44	0.11	1.67*	0.91	1.13
	2 if $25,000–$34,999/year	0.09	0.11	0.10	0.14	0.06	0.11	0.09	0.11	0.10	0.59	0.81	1.90**	0.74	0.14	0.59
	3 if $35,000–$49,999/year	0.12	0.10	0.12	0.11	0.10	0.11	0.07	0.10	0.09	1.00	0.44	0.00	1.41	0.44	0.75
	4 if $50,000–$74,999/year	0.17	0.13	0.12	0.12	0.15	0.14	0.16	0.14	0.16	0.27	0.41	0.65	0.63	0.13	0.50
	5 if $75,000–$99,999/year	0.13	0.12	0.16	0.11	0.14	0.10	0.10	0.12	0.10	1.43	0.14	0.80	0.00	0.73	0.30
	6 if $100,000–$149,999/year	0.16	0.20	0.22	0.17	0.20	0.20	0.20	0.20	0.22	0.33	1.03	0.22	0.11	0.00	0.55
	7 if $150,000–$199,999/year	0.07	0.07	0.06	0.11	0.08	0.08	0.07	0.09	0.06	0.35	1.41	0.17	0.51	0.48	0.88
	8 if $200,000/year or more	0.08	0.08	0.09	0.11	0.08	0.10	0.08	0.05	0.07	0.32	1.07	0.16	0.63	1.88**	1.14
	Mean		4.16	4.42	4.39	4.28	4.26	4.07	4.07	4.07	1.32	1.12	0.60	0.96	0.97	0.95
Household size	1 if ≥3 members		0.56	0.54	0.6	0.61	0.50	0.52	0.54	0.60	0.36	0.9	1.18	0.27	0.89	2.16**
Children	1 if ≥1 child under 18	0.31	0.47	0.46	0.48	0.52	0.43	0.46	0.5	0.53	0.27	0.18	1.16	0.72	1.52	2.24**

Source: United States Census Bureau, 2020.
Single, double, and triple asterisks (*, **, ***) indicate statistical significance at the 10%, 5%, and 1% levels.

Table 1.4: Coefficient estimates for the dried cranberry model, considering information effects using the GMNL–II model in WTP space.

Variables	Coefficient estimates				Pairwise comparison between information treatments (t–stat)[1]		
	Information treatments						
	1	2	3	4	1–2	1–3	1–4
	Mean willingness to pay ($/6–oz bag)						
Sugar content: Regular vs. reduced	-2.33***	-3.54***	-3.56***	-3.85***	1.17	2.24**	4.02***
	(0.38)	(0.54)	(0.45)	(0.66)			
Breeding method: CRISPR vs. conventional breeding	-1.43***	-1.88***	-1.40***	-2.12***	0.94	0.02	1.44
	(0.28)	(0.34)	(0.29)	(0.47)			
Cranberry flavor: Bland/weak vs. full/intense	-3.00***	-3.12***	-2.79***	-4.27***	-0.69	-0.80	1.53
	(0.46)	(0.51)	(0.40)	(0.75)			
Opt-out	-4.78***	-5.27***	-5.41***	-4.83***	—	—	—
	(0.28)	(0.32)	(0.30)	(0.29)			
	Standard deviation						
Sugar content: Regular vs. reduced	4.61***	5.54***	4.38***	5.87***	—	—	—
	(0.61)	(0.80)	(0.58)	(0.95)			
Breeding method: CRISPR vs. conventional breeding	2.73***	2.23***	2.62***	3.62***	—	—	—
	(0.45)	(0.61)	(0.49)	(0.79)			
Cranberry flavor: Bland/weak vs. full/intense	3.97***	3.18***	3.20***	3.97***	—	—	—
	(0.59)	(0.59)	(0.51)	(0.82)			
Opt-out	1.04***	0.17	1.29***	2.01***	—	—	—
	(0.34)	(0.76)	(0.33)	(0.35)			
Scale heterogeneity (τ)	0.99***	1.29***	0.92***	0.94***	—	—	—
	(0.10)	(0.11)	(0.33)	(0.10)			
Certain	-0.44***	-0.40***	-0.51***	-0.79***	—	—	—
	(0.09)	(0.08)	(0.07)	(0.12)			
N. of observations	1500	1500	1500	1500			
Log likelihood	-1262.68	-1212.37	-1217.48	-1204.58			
Akaike information criterion	2557.35	2456.74	2466.95	2441.16			
Bayesian information criterion	2642.37	2541.75	2551.96	2526.17			

[1] The t–tests were based on the following hypotheses: H_{01}: $WTP^{treatment1} \geq WTP^{treatment2}$; H_{02}: $WTP^{treatment1} \leq WTP^{treatment3}$; H_{03}: $WTP^{treatment1} = WTP^{treatment4}$. The t-test uses WTP values that were bootstrapped from the normal distribution based on estimates from the GMNL–II model.
Single, double, and triple asterisks (*, **, ***) indicate statistical significance at the 10%, 5%, and 1% levels. Standard errors are in parentheses.

Table 1.5: Coefficient estimates for the cranberry juice model, considering information effects using the GMNL–II model in WTP space.

Variables	Coefficient estimates				Pairwise t–test comparison between information treatments (t–stat)[1]		
	Information treatments						
	1	2	3	4	H01[1]: 1–2	H02: 1–3	H03: 1–4
	Mean willingness to pay ($/64–fl. oz bottle)						
Sugar content: Regular vs. reduced	-1.23***	-2.04***	-1.44***	-1.61***	1.57	0.11	2.31**
	(0.19)	(0.29)	(0.21)	(0.28)			
Breeding method: CRISPR vs. conventional breeding	-1.05***	-1.61***	-1.46***	-2.33***	3.14	4.26***	7.04***
	(0.19)	(0.24)	(0.19)	(0.35)			
Cranberry flavor: Bland/weak vs. full/intense	-2.29***	-2.43***	-1.98***	-3.02***	-0.36	-2.04	0.90
	(0.29)	(0.30)	(0.23)	(0.41)			
100% Juice	6.95***	8.78***	7.94***	8.15***	—	—	—
	(0.41)	(0.56)	(0.42)	(0.51)			
Cocktail	5.43***	6.92***	6.29***	6.67***			
	(0.38)	(0.50)	(0.37)	(0.47)			
Blend	5.45***	7.30***	6.31***	6.49***			
	(0.39)	(0.52)	(0.38)	(0.47)			
	Standard deviation						
Sugar content: Regular vs. reduced	2.48***	3.23***	2.97***	3.32***	—	—	—
	(0.37)	(0.44)	(0.35)	(0.51)			
Breeding method: CRISPR vs. conventional breeding	0.75***	2.21***	1.79***	2.51***	—	—	—
	(0.26)	(0.33)	(0.26)	(0.40)			
Cranberry flavor: Bland/weak vs. full/intense	2.90***	2.71***	1.59***	2.63***	—	—	—
	(0.39)	(0.39)	(0.27)	(0.49)			
100% Juice	0.66***	2.59***	0.68**	0.85**	—	—	—
	(0.24)	(0.40)	(0.31)	(0.34)			
Cocktail	1.77***	2.40***	1.65***	0.26			
	(0.21)	(0.26)	(0.24)	(0.47)			
Blend	0.62	0.70***	1.17***	1.01***			
	(0.47)	(0.26)	(0.37)	(0.23)			
Scale heterogeneity (τ)	0.18	0.72***	0.65***	0.95***	—	—	—
	(0.12)	(0.08)	(0.06)	(0.10)			
Certain	-0.34***	-0.29***	-0.11*	-0.47***	—	—	—
	(0.09)	(0.08)	(0.06)	(0.08)			
N. of observations	1500	1500	1500	1500			
Log likelihood	-1505.65	-1485.88	-1485.30	-1453.32			
Akaike information criterion	3069.31	3029.75	3028.60	2964.65			
Bayesian information criterion	3223.39	3183.83	3182.68	3118.73			

[1] The t–tests were based on the following hypotheses: H_{01}: $WTP^{treatment1} \geq WTP^{treatment2}$; H_{02}: $WTP^{treatment1} \leq WTP^{treatment3}$; H_{03}: $WTP^{treatment1} = WTP^{treatment4}$. The t-test uses WTP values that were bootstrapped from the normal distribution based on estimates from the GMNL–II model.
Single, double, and triple asterisks (*, **, ***) indicate statistical significance at the 10%, 5%, and 1% levels.
Standard errors are in parentheses.

Table 1.6: Measures of goodness of fit as part of the selection criteria to identify the number of classes in the latent class model, for the three survey versions: Dried cranberries and cranberry juice.

Classes	N. of observations	Likelihood function	AIC	BIC
Model selection criteria for the latent class model – Dried cranberries				
2	1500	-1339.12	2730.24	2868.30
3	1500	-1284.90	2663.80	2913.52
4	1500	-1242.81	2621.63	2982.92
Model selection criteria for the latent class model – Cranberry juice				
2	1500	-1692.57	3445.13	3604.53
3	1500	-1600.09	3306.18	3587.78
4	1500	-1540.95	3233.91	3637.71

Table 1.7: Parameter estimates for the latent class model to represent preference heterogeneity for reduced sugar content – dried cranberries.

Variable	Latent class model parameter estimates		
	Class 1	Class 2	Class 3
Share	41 %	23%	35%
Price	-0.16	0.04	-0.94***
	(0.14)	(0.06)	(0.13)
Breeding method: CRISPR vs. conventional breeding	-1.37***	-0.45***	0.43***
	(0.27)	(0.09)	(0.16)
Sugar content: Regular vs. reduced	-0.73***	-0.44***	-0.99***
	(0.24)	(0.09)	(0.19)
Cranberry flavor: Bland/weak vs. full/intense	-1.84***	-0.73***	0.17
	(0.30)	(0.09)	(0.16)
Opt-out	-0.81*	-2.84***	-3.91***
	(0.47)	(0.27)	(0.44)
Household income ≥$87,500/year	BASE	0.91***	-1.02***
		(0.20)	(0.31)
Added sugar info on NFP is important/crucial		-0.65***	-0.30
		(0.21)	(0.35)
Interpret correctly total sugars and added sugar on NFP		-0.13	-0.82***
		(0.19)	(0.25)
Highest attention is to total sugars content on NFP		0.91***	-3.27***
		(0.22)	(0.91)
Like intense cranberry flavor		0.33*	0.08
		(0.20)	(0.27)
Health motives important when buying cranberry products		0.92***	0.54*
		(0.25)	(0.31)
Ingredients important/crucial when buying cranberry products		-0.46**	0.32
		(0.22)	(0.37)
CRISPR and GMO are different and know the difference		1.27***	0.25
		(0.31)	(0.43)
CRISPR and GMO are different but don't know the difference		1.20***	-0.52
		(0.22)	(0.37)
No difference between CRISPR and GMO		1.19***	0.31
		(0.32)	(0.38)
Willing to purchase CRISPR food if breeding method is only information known		1.18***	0.22
		(0.43)	(0.41)
Willing to purchase CRISPR food if this increases insect resistance and herbicide tolerance		-0.93**	1.83***
		(0.40)	(0.39)
Willing to purchase CRISPR food if this reduces environmental impact of food production		0.31	-0.13
		(0.29)	(0.42)
Willing to purchase CRISPR food if this increases nutrient content in food		2.12***	-2.46***
		(0.40)	(0.75)
Willing to purchase CRISPR food if this reduces the need to add sugars in food processing		-1.37***	3.40***
		(0.35)	(0.81)
Constant		-0.56**	-0.15
		(0.27)	(0.27)
N. of observations	1500		
Log likelihood	-1284.90		
Akaike information criterion	2663.80		
Bayesian information criterion	2913.52		

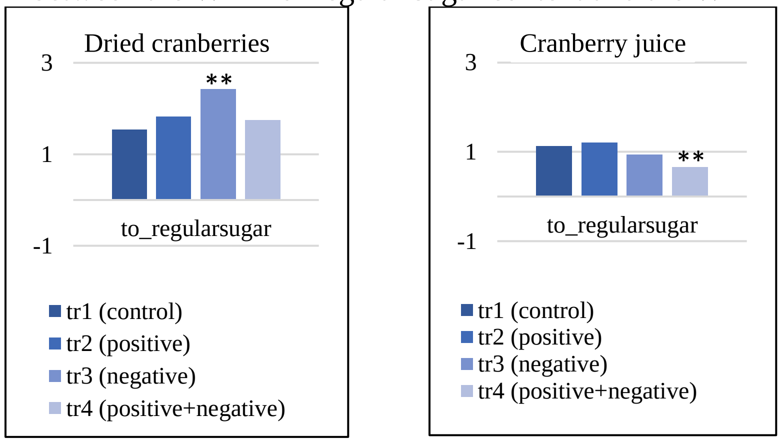
Single, double, and triple asterisks (*, **, ***) indicate statistical significance at the 10%, 5%, and 1% levels.
Standard errors are in parentheses.

Table 1.8: Parameter estimates for the latent class model to represent preference heterogeneity for reduced sugar content –cranberry juice.

Variable	Latent class model parameter estimates		
	Class 1	Class 2	Class 3
Share	24%	56%	19%
Price	-0.97***	-0.38***	0.10
	(0.26)	(0.12)	(0.26)
Breeding method: CRISPR vs. conventional breeding	-0.76***	-0.35***	-0.05
	(0.20)	(0.11)	(0.30)
Sugar content: Regular vs. reduced	-1.22***	-0.17	-0.95***
	(0.25)	(0.12)	(0.29)
Cranberry flavor: Bland/weak vs. full/intense	-1.51***	-1.27***	0.77**
	(0.27)	(0.14)	(0.31)
100% Juice	3.91***	1.09***	5.00***
	(0.83)	(0.42)	(1.44)
Cocktail	3.34***	3.79***	2.32
	(0.90)	(0.44)	(1.43)
Blend	2.90***	4.13***	2.09
	(0.85)	(0.41)	(1.48)
Household income ≥$87,500/year	BASE	0.37**	1.18***
		(0.18)	(0.19)
Added sugar info on NFP is important/crucial		-0.66***	-0.36*
		(0.19)	(0.21)
Interpret correctly total sugar and added sugar on NFP		-0.07	-0.28
		(0.17)	(0.18)
Highest attention is to total sugar content on NFP		0.14	0.15
		(0.22)	(0.24)
Like intense cranberry flavor		0.31	-0.45**
		(0.19)	(0.22)
Health motives important when buying cranberry products		0.18	0.32
		(0.20)	(0.21)
Ingredients important/crucial when buying cranberry products		-0.87***	-0.36
		(0.20)	(0.22)
CRISPR and GMO are different and know the difference		1.37***	1.21***
		(0.38)	(0.40)
CRISPR and GMO are different but don't know the difference		-0.20	-0.05
		(0.19)	(0.20)
No difference between CRISPR and GMO		0.22	0.06
		(0.28)	(0.30)
Willing to purchase CRISPR food if breeding method is only information known		-0.50	-1.01***
		(0.31)	(0.34)
Willing to purchase CRISPR food if this increases insect resistance and herbicide tolerance		-0.18	0.44
		(0.29)	(0.30)
Willing to purchase CRISPR food if this reduces environmental impact of food production		0.06	0.26
		(0.31)	(0.36)
Willing to purchase CRISPR food if this increases nutrient content in food		0.83***	0.46
		(0.32)	(0.32)
Willing to purchase CRISPR food if this reduces the need to add sugars in food processing		1.11***	0.01
		(0.33)	(0.37)
Constant		0.83***	-0.23
		(0.22)	(0.23)
N. of observations	1500		
Log likelihood	-1600.09		
AIC	3306.18		
BIC	3587.78		

Single, double, and triple asterisks (*, **, ***) indicate statistical significance at the 10%, 5%, and 1% levels.
Standard errors are in parentheses.

Figure 1.1: Trade-off between the WTP for regular sugar content and the WTP for CRISPR.



Notes:
The pairwise t-tests were based on the following hypotheses:
H_{04}: $\text{Tradeoff}^{\text{treatment1}} \leq \text{Tradeoff}^{\text{treatment2}}$;
H_{05}: $\text{Tradeoff}^{\text{treatment1}} \geq \text{Tradeoff}^{\text{treatment3}}$;
H_{06}: $\text{Tradeoff}^{\text{treatment1}} = \text{Tradeoff}^{\text{treatment4}}$.
The t-test uses tradeoff values that were bootstrapped from the normal distribution based on estimates from the GMNL–II model. Single and double asterisks (*, **) indicate statistical significance at the 10% and 5% levels.

Figure 1.1 shows that for dried cranberries under all treatments, the coefficient is significantly larger than one, implying that the aversion to regular sugar content is larger than the aversion to CRISPR. The effect of information on the coefficient is not consistent across products. Only treatment 3 yields a statistically significant different result from the control. For the cranberry juice, the coefficients in treatment 1, 2, and 3 are not significantly different from 1, meaning that respondents are indifferent between the regular sugar content and CRISPR. However, in treatment 4, the coefficient is significantly lower than the control, indicating that when both pieces of information were provided, the aversion towards CRISPR is larger than the aversion to regular sugar content.

CHAPTER TWO: QUALITY-RELATED DESCRIPTORS TO INCREASE FRESH BLUEBERRIES PURCHASE - EVIDENCE FROM A BASKET-BASED CHOICE EXPERIMENT.

Abstract

Given the increased incidence of diet-related health issues, identifying strategies to increase the consumption of fruit and vegetables is essential for the agri-food chain. Fruits like blueberries contain high phenolic phytochemicals, which act as antioxidants and are responsible for many health benefits. This study used a Basket-Based Choice Experiment (BBCE) to identify which sensory and hedonic quality descriptors of fresh blueberries will likely increase the likelihood of purchasing. The study revealed that fresh blueberries containing a “Stay Fresh” descriptor on the package had a smaller price elasticity relative to packages with no descriptors or “Sweety” and “Crunchy” descriptors. This suggests consumers are less sensitive to price changes when the package of blueberries have trigger words that communicate a longer shelf-life. In addition, the study found that strawberries, blackberries, raspberries, and blueberries, are more likely to be purchased together rather than being substitutes. Finally, the study found that males, older individuals, employed individuals, those with a college degree, physically fit individuals, white individuals from the Northeast region, those who placed a high value on nutrition, and those who had a higher weekly budget for fresh fruits were more likely to choose blueberries from a basket of commonly consumed fruits. These findings can help blueberry growers, retailers, and marketers develop strategies to increase the demand for fresh blueberries.

Keywords: Consumer Choice, Blueberries, Price Elasticity, Complementarity, Quality Descriptor.

2.1 Introduction

To reduce the prevalence of overweight and obesity[6], the World Health Organization (WHO) recommends increasing the consumption of fruits and vegetables, legumes, whole grains, and nuts (World Health Organization, 2021). Fruits containing high phenolic phytochemicals, such as blueberries, blackberries, pomegranate, cranberries, plums, and apples, contribute to the anti-obesity effects of fruit consumption (Wolfe et al., 2008; Sharma et al., 2016). Recent studies demonstrate the benefits of blueberries and its anthocyanins components in reducing the risk of cardiovascular disease (CVD), type 2 diabetes mellitus (T2DM), hypertension, and cognitive decline in older adults (Basu et al., 2010; Cassidy et al., 2011; Devore et al., 2012; Jennings et al., 2012; Wedick et al., 2012; Cassidy et al., 2013; Kalt et al., 2020; Istek and Gurbuz, 2017). Given the host of health benefits associated with consuming fruits and vegetables, the food industry and policymakers must identify marketing strategies promoting healthy foods, such as blueberries.

The U.S. per capita consumption of fresh blueberries has doubled over the past decade, from a yearly average of 1.2 in 2012 to 2.3 pounds in 2021 (USDA, 2023a). Reasons for the increase in blueberry consumption include the recognition of health benefits, as well as their improved quality, year-round availability, and convenient packaging (Cook, 2011). Other reasons include the decrease in the real price of blueberries and prices of blueberry substitutes, the increased individual income in the United States, and the promotion efforts of the U.S. Highbush Blueberry Council (USHBC) (Kaiser, 2015). USHBC (2018) confirmed the need to increase per

---

[6] Obesity and overweight represent worldwide problems. The World Health Organization (WHO) reports that as of 2016, 39% of adults worldwide were overweight and 13% obese (World Health Organization, 2021). The United States has one of the highest prevalence of overweight and obesity among adults and children worldwide. From 1999-2000 to 2017-2018, the age-adjusted prevalence of obesity increased from 30.5% to 42.4%. Obesity costs the U.S. healthcare system $147 billion a year (Hales et al., 2020).

capita consumption, highlighting the importance of identifying population segments representing growth opportunities. In the early 2010s, reports by Fresh Trends (2013) and Brazelton (2013) suggest that a small percentage of individuals (about 15%) who purchased blueberries were frequent buyers. As a result, Gilbert et al. (2014) emphasizes the importance of converting seldom buys into regular purchases through product satisfaction. The authors claim that limitations in the industry's ability to supply berries with a consistent appearance, texture, and sensory profiles curtail the number of consistent buyers, hindering the growth of the U.S. blueberry industry.

Despite the need for the berry industry to center on "consumer satisfaction," there is limited research connecting the sensory quality characteristics of blueberries with the possibility of raising per-capita consumption. Specifically, the impact of blueberry word descriptors indicating sensory or hedonic quality attributes on the likelihood of purchasing blueberries when they are presented alongside a selection of other fresh fruits commonly consumed in the United States. Findings from this study could prove useful to stakeholders to provide consistent high-quality blueberries that meet consumers' expectations and potentially increase per capita consumption. Also, this information could provide valuable insights for shaping policies intended to encourage the consumption of healthy fruits.

This study aims to estimate blueberries' price elasticity under different word descriptors signaling specific sensory and hedonic quality attributes such as "Sweety," "Crunchy," and "Stay Fresh." The end-goal is to reveal what specific sensory and hedonic quality attributes would increase the probability of consumers purchasing fresh blueberries by triggering the expectancy of higher quality blueberries and increasing their desirability. Second, the study identifies how those specific sensory quality factors affect consumers' sensitivity to changes in blueberries' price (i.e., own price elasticity). Third, we estimate the cross and own price elasticities of popular fresh fruits in the United States and identify complementarity and substitution patterns. Fourth, the study seeks to establish the profile of individuals who regularly purchase blueberries.

The choice of the words “Crunchy” and “Sweety” was based on previous literature that found that the intensity of these sensory quality attributes are strongly associated with consumers' acceptance of blueberries (Saftner et al., 2008; Blaker et al., 2014; Gilbert et al., 2014; Mennella et al., 2017; Sater et al., 2021; Yue and Wang, 2016; Donahue and Work, 1998; Qu et al., 2017). The label “Stay Fresh” was chosen to convey the idea of preferences for an extended shelf-life. Prior studies investigating preferences for shelf-life, utilized a combination of dates to assess consumer preferences for the same fruit or vegetable (Baselice et al., 2017; D'Amato et al., 2023; Zheng et al., 2016). However, our study had a different purpose, it aimed to examine whether a label with descriptors signaling quality attributes (e.g., extended shelf-life, crunchiness, or sweetness) influenced the decision to purchase blueberries when presented alongside a larger selection of the fruits most commonly consumed in the United States.

2.2 Literature Review

Previous research has shown that food labels and descriptions on the outside package of food products impact consumers' a priori expectations regarding taste and quality (Blackmore et al., 2021; Piqueras-Fiszman and Spence, 2015; Papies et al., 2020a; Liem et al., 2012). For instance, including sensory and hedonic descriptors on food labels enhances eating simulations and food attractiveness, positively impacting consumers' preferences (Turnwald and Crum, 2019; Papies et al., 2020b). Woods et al. (2011) demonstrated that labels suggesting extra sweetness increased the perception of the intensity of sweetness. Additionally, flavor descriptors can affect desire and consumer behavior (Hazebroek and Croijmans, 2023).

A substantial body of literature explores consumers' evaluation of blueberries and their attributes. These studies conclude that consumers prefer locally produced fresh blueberries (Shi et al., 2011; Shi et al., 2013; Shi et al., 2015; Hu et al., 2009; Girgenti et al., 2016; Qu et al., 2017). The size of the blueberry and eating quality attributes, including the intensity of blueberry flavor, sweetness, freshness, juiciness, tartness, and texture, are strongly associated with consumers' acceptance of blueberries (Saftner et al., 2008; Blaker et al., 2014; Gilbert et al., 2014; Mennella et al., 2017; Sater et al., 2021; Yue and Wang, 2016; Donahue and Work, 1998; Qu et al., 2017). In particular, firmness has been identified as one of the most critical factors positively impacting consumers' preferences (Sater et al., 2021; USDA, 2020; Ehlenfeldt and Martin, 2002). Regarding consumers' preferences for processed blueberry products, consumers favor organic and sugar-free features (Hu et al., 2009). In tandem with consumers' preferences, blueberry producers consider improved fruit quality, firmness, flavor, and shelf-life (Gallardo et al., 2018).

Previous studies estimated consumers' willingness to pay (WTP) for quality characteristics associated with blueberries and blueberry products (Hu et al., 2009; Shi et al., 2011; Shi et al., 2013; Shi et al., 2015; Stevens et al., 2015). These studies found that consumers are willing to pay more for organic blueberries, locally produced, pollinated by native bees, and sugar-free. To elicit WTP, researchers commonly used discrete choice experiments (DCE) (Hu et al., 2009; Shi et al., 2011; Stevens et al., 2015) and experimental auctions (Shi et al., 2013; Shi et al., 2015). Similar approaches have been used in consumer WTP studies for other berries (Hoke et al., 2017; Darby et al., 2006).

Unlike previous research, this study evaluates how the effects of labels denoting sensory quality attributes impact consumers' likelihood of purchasing blueberries among a variety of fresh fruits. We apply the basket choice models, that has been used in a limited capacity in the agricultural economics literature (Kwak et al., 2015; Richards et al., 2018; Caputo and Lusk, 2022). In particular this study follows Caputo and Lusk (2022) who introduced the basket-based choice experiment (BBCE). Unlike traditional DCEs, which offer participants a single food item with different attribute levels, a BBCE presents participants with various food items, allowing them to select a food item or a combination of food items to construct a bundle, resembling an actual grocery shopping experience more closely. The use of BBCE also allows us to identify substitution and complementary patterns and to identify potential changes in price elasticities due to the inclusion on labels of word descriptors that signal hedonic or sensory quality attributes.

Additionally, the BBCE approach includes the selection of shopping bundles multiple times, enabling the analysis of the substitution and complementarity relationships between products included. Similar studies but using different methods have been used in the literature. Song and Chintagunta (2006) used a bundle-specific utility that included a category-specific and a brand-specific component. Their analysis used store-level scanner data, showing that softeners and detergents are complementary and brands influence cross-price elasticities. Kwak et al. (2015) utilized an assortment choice model to investigate consumers' yogurt choices during a single shopping trip. They found that brand impacts consumers' yogurt choices, and when the perceived product quality is higher, consumers prefer less variety. Richards et al. (2018) used a shopping basket model to explore the effect of complementary goods on retail competition based on household-scanner data collected on purchase occasions. They discovered that the presence of complementary goods decreases retail competition.

Caputo and Lusk (2022) conducted a BBCE study on 21 foods, which enabled them to examine the substitution and complementarity patterns among foods and their own and cross-price elasticities. The researchers applied their results to assess how changes in the price of food items can affect the nutritional intake of individuals, as well as the welfare consequences of food policies. Their analysis revealed that changes in food prices have the potential to influence dietary habits by impacting nutrient intake. Neill and Lahne (2022) conducted an experiment involving six vegetable choices using a basket and expenditure-based choice experiment (BEBCE) design combined with a sensory experiment. They found that most vegetables are considered complements or independent of each other rather than substitutes. These studies' outcomes provide insights into consumer behavior and inform policy decisions.

2.3 Experimental design

Data were collected through an online survey using the Qualtrics platform and the Qualtrics Consumer Research Panel. The survey was pretested in a soft launch in February 2023, with the full implementation during the second and third weeks of the same month. Researchers requested Qualtrics to select a sample close to the U.S. population based on three demographics: age, income, and region. The survey had four different versions based on labeling treatments applied to blueberries only and to be explained below, with Qualtrics providing 801, 802, 805, and 800 participants for each survey version, totaling 3,208 responses across the United States. The survey inclusion criteria required respondents to be 18 years or older, be the primary grocery shopper in their household, and to had consumed blueberries at least once during the previous year. Every survey version included the BBCE. Additionally, respondents were asked about their blueberry preferences and consumption patterns, attention to labels, and sociodemographic characteristics. The Institutional Review Board (IRB) approved the survey at Washington State University, IRB number 19812-001.

2.3.1. The food basket choice experiment

For our study, we selected the top fresh fruits most popular in the United States in 2021 (International Fresh Produce Association, 2022). We also added blackberries to the list as they are often seen as a substitute for blueberries (Cook, 2011; Sobekova et al., 2013; Arnade and Kuchler, 2015). We removed lemons and cherries from our study because lemons are typically eaten in combination with other foods rather than being consumed alone, and cherries are seasonal and not commonly available throughout the year. Despite the controversy surrounding whether avocados are considered fruits (USDA, 2019), they are typically displayed with other fruits at grocery stores; therefore, we decided to incorporate them into our study to imitate an actual shopping scenario. Thus, our chosen 14 fresh fruits to be included in the choice basket were apples, avocados, bananas, blackberries, blueberries, cantaloupe, grapes, oranges, peaches, pears, pineapple, raspberries, strawberries, and watermelon.

Every survey participant was presented with six choice scenarios (choice baskets) in which they were presented with fourteen different fresh fruits at a posted price. The same fruit options were presented in each scenario, as scenarios only differ in the prices shown to participants. In each scenario, participants were asked to choose the fruit or combination of fruits they would most likely buy in a shopping experience. They could opt out if none of the fruits at the listed prices appealed to them. Figure 2.1 Panel A illustrates a screenshot of a fruit basket choice scenario presented to the survey participants. Participants could add any fruit to their virtual shopping cart on the right by clicking the "+" icon. To assist respondents in setting realistic expenditure levels, we requested that they state their weekly spending on fresh fruits before conducting the choice experiment. Neill and Lahne (2022) suggested that one displays the budget respondents indicated for fresh fruit in each choice question, reminding them of their cognitive budget constraint. Showing their regular fruit budget encouraged respondents to view the experiment as real rather than hypothetical.

Additionally, to enhance the real experience of the decision-making process, the total cost of the respondent's choices was presented as "Total Bill" in their shopping cart, following Caputo and Lusk (2022). If they changed their mind or accidentally selected the wrong fruit, they could remove it by clicking on the "X" icon. Participants could use the "Clear Cart" button to remove all fruits selected. The shopping cart displayed the selected fruits' total cost on the screen's right side. After selecting their desired fruits, respondents could click the "Finish" button to complete their purchases. If they chose not to purchase fruits in a scenario, they could click the "No Buy" option to proceed to the following scenario with an empty cart. A screenshot of a scenario where fruits have been added to the shopping cart is presented in Figure 2.1 Panel B.

Each choice question had the same format, with the only variation across questions being the prices of the fruits. Each fruit option's price varied at three levels: low, medium, and high. The prices used in each choice were selected through research of online grocery store prices across the United States during the last week of January 2023 (Table 2.1). These prices were also compared and validated using prices reported by the Agricultural Marketing Service, weekly advertised fruit retail prices (USDA, 2023b). Out of the 3^{14} possible price combinations, an orthogonal fractional factorial design selected a subset of 54 fruit choice scenarios. The 54 questions were organized into nine blocks consisting of six scenarios each. Each participant was randomly assigned to one of the nine blocks.

To minimize the impact of hypothetical bias, we employed a cheap talk script (Champ et al., 2009) in all treatments. Furthermore, we implemented a random ordering of choice sets to decrease the likelihood of learning effects and ordering bias (Caputo et al., 2017).

2.3.2. Treatments

The goal of the study is to measure the impact of blueberry word descriptors indicating sensory or hedonic quality attributes on the likelihood of purchasing blueberries. The exercise used “Crunchy,” “Stay Fresh,” and “Sweety” on the front of the blueberry packages in a between-subjects design yielding four treatments. Hereafter, when we refer to “labeling treatment,” each is associated with a different survey version. To avoid any effect from only having the blueberries exhibit a labeling treatment, all the fruits (including blueberries) in the control and the other three treatments displayed the logo “Farmers’ Best.” All the fruit options, including the blueberries, remained the same within each treatment, and only prices varied randomly. In sum, the four treatments are as follows. Treatment 1 was the control, in which all fruits, including the blueberries, exhibited the logo “Farmers’ Best” on the clamshell. Treatment 2 presented the blueberry clamshell with the word “Crunchy,” suggesting the blueberries exhibit a crisp texture. Treatment 3 showed the blueberry clamshell with the phrase “Stay Fresh,” meaning the blueberries exhibit long-lasting durability in the refrigerator. And treatment 4 presented the blueberry clamshell with the word “Sweety” to indicate that the blueberries taste sweet. Figure 2.2 shows the labeling treatments. We expect the treatment groups to exhibit lower own-price elasticities, implying that consumers will display less sensitivity towards price changes for fresh blueberries labeled “Crunchy,” “Stay fresh,” or “Sweety” relative to those without labeling (the control group).

2.4. Empirical approach

We used the Multivariate Logistic (MVL) choice models to model basket-based choices (Song and Chintagunta, 2006; Kwak et al., 2015; Richards et al., 2018; Caputo and Lusk, 2022). This approach treats every possible bundle as a distinct choice alternative, resulting in

2^{14}=16,384 possible bundles. The empirical approach is based on the random utility model introduced by McFadden (1974). Following Richards et al. (2018) and Caputo and Lusk (2022), the utility respondent *i* derives from choosing bundle *b* is,

$$U_{ib} = V_{ib} + \varepsilon_{ib} \tag{2.1}$$

where ε_{ib} is distributed generalized extreme value following Train (2009), and V_{ib} can be written as,

$$V_{ib} = \sum_{j=1}^{J} \alpha_{ij} x_j + 0.5 \sum_{j=1}^{J} \sum_{k \neq j}^{J} \gamma_{jk} x_j x_k \tag{2.2}$$

x_j equals 1 if fruit *j* is being placed in basket *b* and 0 otherwise. α_{ij} is the baseline utility for fruit *j* derived by individual *i*, and it can be defined as $\alpha_{ij} = \alpha_{0,j} + \beta_j p_j + X_i \delta_i$, where $\alpha_{0,j}$ is the constant of fruit *j* or the opt-out option, p_j is the price of fruit *j*, X_i is the vector of characteristics of individual *i*, and β_j and δ_i are parameters. Parameter γ_{jk} represents the relationships between the item *j* and *k* in terms of their effect on utility. When $\gamma_{jk} > 0$, items *j* and *k* are complements in utility. When $\gamma_{jk} < 0$, items *j* and *k* are substitutes in utility. The utility of consuming item *j* is invariant in the presence of item *k* when $\gamma_{jk} = 0$. We followed previous studies (Besag, 1974; Cressie, 1993; Russell and Petersen, 2000; Kwak et al., 2015; Caputo and Lusk, 2022) and implemented the following restrictions $\gamma_{jk} = \gamma_{kj}$ and $\gamma_{jj} = 0$.

When ε_{ib} is represented by the generalized extreme value distribution following Train (2009), the probability of individual *i* choosing the observed basket *b* among the 16,384 possible bundles is,

$$Prob(individual\ i\ chooses\ basket\ b) = \frac{e^{V_{ib}}}{1+e^{V_{ib}}} \tag{2.3}$$

The composite conditional likelihood[7] function can be expressed by the multivariate logit form in the following form (Besag, 1974; Caputo and Lusk, 2022),

$$Prob(individual\ i\ places\ fruit\ j\ in\ basket) = \frac{e^{z_{ij}}}{1+e^{z_{ij}}} \tag{2.4}$$

$$z_{ij} = \alpha_{ij} + \sum_{k\neq j}^{J} \gamma_{jk} y_{ik} \tag{2.5}$$

where z_{ij} is the utility of individual *i* who places fruit *j* in the basket, y_{ij} equals to 1 when the individual *i* places fruit *k* in the basket, and 0 otherwise.

2.4.1 Own and cross-price elasticities

Previous research suggests that the cross-utility effects γ_{ij} do not necessarily indicate whether two products are substitutes or complements based on price (Richards et al., 2018; Caputo and Lusk, 2022). Instead, it signifies how one fruit's consumption impacts the other fruits' utility. Therefore, to determine if products are substitutes or complements in response to price changes, we utilize the estimates from the MVL model to estimate the own and cross-price elasticities of fruit items. Specifically, the arc elasticity is determined by examining how the likelihood of selecting fruit *j* changes with the prices of fruit *j* and *k*. To estimate the elasticity of fruit *j* resulting from a 1% increase in the midpoint price of fruit *k*, p_k, we follow:

$$e_{jk} = \frac{Prob(fruit\ j\ is\ placed\ in\ basket\ at\ 1.01\times\ p_k) - Prob(fruit\ j\ is\ placed\ in\ basket\ at\ p_k)}{Prob(fruit\ j\ is\ placed\ in\ basket\ at\ p_k)} \tag{2.6}$$

The probability of fruit *j* being in a basket can be calculated by adding up the probability of choosing all the baskets that contain fruit *j*, contingent on individual characteristics and the price of each fruit.

[7] According to Bell et al. (2018) the composite conditional likelihood (CCL) method treats conditional probabilities as separate yet correlated. Although the CCL approach yields consistent estimators, it tends to produce less efficient standard errors.

2.5 Results

Table 2.2 presents the sociodemographic variables included in the model. This includes the pooled sample and each treatment group. Compared to the 2021 U.S. population, as the U.S. Census Bureau (U.S. Census Bureau, 2021) reported, our sample exhibited a higher percentage of individuals with a four-year college degree, a greater proportion of females, and a higher rate of white respondents. This pattern aligns with the demographics of individuals more responsive to surveys (Curtin et al., 2000). Furthermore, our sample had a larger proportion of individuals with at least one child while having a lower percentage of respondents with a higher income (≥ $75,000 per year) compared to the overall U.S. population. The summary statistics of the entire set of sociodemographic variables across all treatment samples can be found in Table 2.3.

2.5.1 Survey findings

First, we present the control treatment (no label) group responses to avoid labeling treatment effects. The average number of selected fruits was 5.25 (out of 14 fruits presented in each scenario). Around 5.86% of baskets were empty as respondents chose not to buy any fruits, while 1.43% of baskets were full as respondents chose to buy all 14 fruits. As depicted in Figure 2.3, bananas were the most commonly selected fruit, appearing on average across the four treatments in 74% of baskets. Apples followed appearing on average across treatment on 56% of baskets. Blueberries were the third most frequently selected fruit, on average, in 48.5% of baskets. The top five fruits also included strawberries and grapes, which aligns with the International Fresh Produce Association's report, where bananas, strawberries, grapes, apples, and watermelon are the most popular fruits consumed by Americans (International Fresh Produce Association, 2022). However, watermelon was not a popular option in our basket-based choice experiment. This could be attributed to our survey being conducted in February when watermelon is out of season, making it less popular. Respondents stated they consumed blueberries more frequently than respondents to the International Fresh Produce Association's survey, which ranked blueberries seventh in popularity among fresh fruits consumed in the United States. Results indicate that our respondents' weighted average per capita yearly purchase of fresh blueberries was 13.41 pounds[8], six times more than the U.S. per capita availability of fresh blueberries in 2021, at 2.3 pounds (USDA, 2023a). This could be the result of our study only including respondents who had consumed fresh blueberries the previous year, which excludes respondents who do not consume blueberries.

The frequency distribution of responses to fresh fruit purchase questions can be found in Table 2.4. The most prevalent form of blueberry consumed was fresh, followed by frozen and dried blueberries. The preferred method of consuming fresh blueberries was eating them raw and alone, adding them as toppings to granola or yogurt, and using them in smoothies and beverages. The most popular package size was a 6 oz package, followed by 4.4 oz and 12 oz packages, the weighted average quantity of fresh blueberries purchased during one shopping occasion was 0.80 pounds. Furthermore, our survey revealed the top five reasons why fresh blueberries were not

[8] Considering that our survey inclusion criterion was consuming fresh blueberries at least once during last year, our sample excludes individuals/households who do not consume blueberries, while including frequent and non-regular consumers. In our sample, 3.25% of respondents stated that they purchase fresh blueberries 4-6 times per week, 12.36% of them purchase fresh blueberries 2-3 times per week, 23.97% of them purchase fresh blueberries once a week, 25.72% of them purchase fresh blueberries once every 2 to 3 weeks, 14.36% of them purchase fresh blueberries once a month, 20.35% of them purchase fresh blueberries less than once a month. Based on the data, the weighted average frequency of purchasing fresh blueberries was 45.32 times per year. Regarding the quantity purchased, 23.22% of participants reported buying a single package of 4.4 oz blueberries, while 7.37% purchased two packages of the same size. Similarly, 23.85% of participants purchased one package of 6 oz blueberries, while 7.87% purchased two packages of the same size. Additionally, 11.61% of participants purchased one package of 11-12 oz blueberries, while 5.12% purchased two packages of the same size. For the larger size, 7.37% of participants purchased one package of 16-18 oz blueberries, while 2.87% purchased two packages of the same size. For even larger size, 1.75% of them purchase one package of 24 oz blueberries and 1.12% of them purchase one package of 32 oz blueberries. Less than 1% of participants reported purchasing three or more packages of each size. Based on the weighted data, the average amount of fresh blueberries purchased in one shopping occasion was 12.88 oz. Considering the average household size of 2.72 people in our sample, the weighted average annual purchase of fresh blueberries per person was 214.60 oz, or 13.41 pounds.

consumed frequently (less than once a month). These included price, short shelf-life at home, lack of freshness, lack of fresh blueberries available, and unwanted package size. Among these reasons, price was the primary concern, identified by 54.56% of the respondents. This corresponds with Yue and Wang's (2016) discovery that price played a significant role for U.S. consumers when selecting fresh blueberries. In contrast to our survey results, Girgenti et al. (2016) found that the price of a product was not a crucial factor for Italian consumers when choosing blueberries and raspberries, as these products were typically bought in small quantities. Table 2.5 summarizes the responses to questions about other aspects of blueberry consumption. When asked to rate the importance of blueberry quality attributes on a 1-5 scale (1=most important, 5=least important), the freshness was rated the highest, followed by free from defects, ripeness, phytonutrient content, and sweetness. When asked what quality characteristics should be improved to increase consumption, improved eating quality traits were rated the highest, followed by improved visual quality traits, staying fresh longer, an improved response to climate change, and nutritional traits. When asked about statements associated with blueberry consumption (Mezzetti and Predieri, 2022), the top three rated were natural product, eaten at breakfast, and a good source of vitamins and minerals. When asked about the importance of labels, the top three rated were pesticide-free, domestic product, and not genetically engineered.

2.5.2 Baseline utility estimates

We estimated two different MVL models, Model 1 using a single price effect β for all fruit varieties resulting in 422 parameters, and Model 2 assigning a specific price effect for each fruit variety β_j, resulting in 435 parameters. Table 2.6 presents the model fit statistics for both specifications. Our findings show that Model 2 generally outperformed Model 1 based on AIC and loglikelihood values. We also performed a likelihood-ratio test, where the likelihood-ratio test statistic was computed as -2($LLR_{Model\ 1}$-$LLR_{Model\ 2}$), with 13 degrees of freedom. The likelihood-ratio test statistics were 36, 29, 25, 49, and 101 for the control, “Crunchy”, “Stay fresh”, “Sweety”, and pooled sample. We rejected the null hypothesis in all cases, indicating that Model 2 was preferred. Therefore, we chose to use the estimation from Model 2 in this study[9].

Individual characteristics were selected based on their significance in a logistic regression model, where the choice of selecting blueberries was the dependent variable and the set of individual characteristics were the independent variables. Table 2.7 presents the summary statistics for the individual characteristics used in our model. All individual characteristics were represented as dummy variables except for the stated weekly budget for fresh fruits, where the mean value was $36.75.

The results from Model 2 for the pooled and all treatment samples are shown in Table 2.8. As expected and consistent with the law of demand, the price coefficient was negative for the pooled and the treatment samples implying that blueberries were less likely to be placed in baskets as prices increased. Also consistent for the pooled and treatment samples is that non-millennials and white respondents were more likely to choose blueberries. Other sociodemographics, such as gender, presence of children, employment, college education, conservative views, living in the Northeast, being physically fit, and considering health and nutrition important, are statistically significant factors but not consistently across treatment samples. In general, male respondents, employed, with a college degree, were physically fit, lived in the Northeast region, placed a high value on nutrition, and had a higher weekly budget on fresh fruits were more likely to choose blueberries. These findings align with the results of

[9] Model 1 and Model 2 yielded parameter estimates similar in magnitudes.

Gilbert et al. (2014), who reported that blueberry buyers tend to have an income exceeding $100,000 and reside in the northeastern region of the United States. Additionally, it aligns with the findings by Laaksonen et al. (2016), who found that elderly and health-conscious consumers were more inclined to be interested in berries. Reports by The Packer (2019) support these findings, as there is a higher probability of fresh blueberry purchases among white, older, and high-income individuals.

Other sociodemographic results that were statistically significant and negative, albeit not consistently across samples, indicate that conservative views, households with at least one child, and respondents diagnosed with diabetes or high cholesterol were less likely to include blueberries in their shopping carts. A possible explanation might be that those diagnosed with diet-related diseases might not necessarily respond to healthful eating styles or health information even after diagnosis (Mancino and Kinsey, 2004). On the presence of children, Mennella and Bobowksi (2015) found that kids dislike the bitterness of berries, possibly explaining why households with children are less likely to choose blueberries.

Other respondents' characteristics with significance and sign coefficients inconsistent across pool and treatment samples were liberal views and the importance of the label "non-GMO." Respondents who identified with liberal views chose blueberries in the "Stay fresh" group, while they were less likely to select blueberries in the "Sweety" group. Respondents for whom the label non-GMO was important were more likely to choose blueberries in the pooled and control sample but less likely to select blueberries in the "Sweety" group.

Table 2.9 shows the baseline utility estimates from the MVL model for the 14 fruits in the control group. Younger respondents were more prone to choose the empty cart option. Conversely, those employed, with a college degree, identified as liberal, resided in the South, and cared about nutrition were less likely to choose the no-buy option. Notably, males were more likely than females to add oranges and pears to their baskets, while females were more likely to select avocados, grapes, raspberries, and strawberries. Younger individuals were more likely to opt for strawberries and pineapple. Higher-income participants were less likely to pick oranges and were more likely to choose bananas, cantaloupe, and strawberries. Those with children tended to choose apples, avocados, grapes, pears, and strawberries but avoided peaches, pineapples, and raspberries. Employed individuals were less likely to choose apples and pears. And individuals with a college degree were more likely to select apples, blackberries, peaches, pears, and raspberries. These findings were consistent with Caputo and Lusk (2022), who found that females are more likely to choose strawberries than males, and high-income individuals more often select strawberries.

2.5.3 Cross-utility effect estimates

As shown in Table 2.10, most of the cross-effects show a positive correlation, which suggests that the fruits have complementary effects on respondents' utility. Across all treatments (except for the "Stay fresh"), negative coefficients were observed in the cross-utility effects between blueberries and pineapples, as well as blueberries and oranges, indicating that pineapples and oranges are not likely to be purchased with blueberries. Interestingly, berries (i.e., strawberries, blackberries, raspberries, and blueberries) are more likely to be bought together, inconsistent with previous studies on elasticity patterns between berries. For example, Sobekova et al. (2013) found that strawberries, blueberries, blackberries, and raspberries are substitutes for one another. Kaiser (2015) also found that strawberries are a substitute for blueberries. Our finding is consistent with Caputo and Lusk's (2022) findings that apples and bananas are frequently purchased together and proposed that this tendency for similar items to be bought together may be due to "variety seeking or differential preferences among members of a household."

2.5.4 Own and cross-price elasticities

Table 2.11 shows the own and cross-price elasticities of blueberries. The elasticities were computed based on the parameter estimates from Model 2, using the mean individual characteristics and medium price. The cross-price effects indicate a negative correlation, indicating that all products have complementary demands. This means that an increase in the price of other fresh fruits decreases the likelihood of purchasing blueberries. In contrast to the findings of Sobekova et al. (2013), our study suggests that all types of berries have complementary relationships.

The cross-price effect with the highest magnitude in all treatment groups is from strawberries, indicating that a 1% increase in the price of strawberries leads to a 0.20% decrease in the likelihood of buying blueberries. Blueberries with the words “Stay fresh” on the package showed an own-price elasticity of -0.544, implying that the demand for “Stay fresh” blueberries declined by 0.54% on average with a one percent increase in price. This value is the lowest compared to other sample treatments. The blueberries’ own price elasticity for blueberries with the word “Control” is -0.759, “Crunchy” is -0.768, “Sweety” is -0.736, and the pooled sample is -0.685. This finding is consistent with prior research, which has shown that sensory and quality labels, along with product descriptions, have an impact on consumers' perception of the product (Blackmore et al., 2021; Piqueras-Fiszman and Spence, 2015; Papies et al., 2020a). Our result suggests that blueberries with a longer shelf life could potentially reduce consumers’ sensitivity to price changes.

Across all treatments, the treatment that showed the highest frequency of blueberry selection was the “Stay fresh” label, with 52.19% of baskets including blueberries, which was higher than the other treatments. For instance, the control treatment had a selection rate of 50.75%, the “Crunchy” treatment had a selection rate of 43.37%, and the “Sweety” treatment had a selection rate of 47.67%. See Figure 2.3. This is consistent with previous studies that found blueberries’ freshness was the most significant factor in consumers' decision to purchase them (Yue and Wang, 2016; Girgenti et al.,2016; Qu et al., 2017). Even though previous studies have shown that consumers prefer sweet and crispy fresh blueberries (Gilbert et al., 2014; Yue and Wang, 2016), our results do not indicate that a “Sweety” or “Crunchy” logo would increase the probability of purchase.

We used individual characteristics to calculate the individual own-price elasticity of fresh blueberries. We conducted t-tests to determine if there was a significant difference in the own price elasticity of fresh blueberries across different treatment groups. Figure 2.4 shows the histograms of each group's price elasticity of fresh blueberries. Most individuals in the “Control” and “Crunchy” groups had a price elasticity for blueberries between -1.1 and -0.5. In the “Stay fresh” group, most individuals had a price elasticity between -0.7 and -0.4, while in the “Sweety” group, the majority had a price elasticity between -1.0 and -0.5. Our t-test results indicated that respondents had a smaller price elasticity for fresh blueberries labeled as "Stay fresh" ($t= -32.434$, p-value<0.001) and "Sweety" ($t= -3.773$, p-value<0.001) compared to fresh blueberries without any labeling. There was no significant difference in the price elasticity of "Crunchy" compared to the "Control" groups ($t= 0.330$, p-value$=0.629$). Additionally, the price elasticity for "Stay fresh" blueberries was smaller than that of "Sweety" blueberries ($t= -31.236$, p-value<0.001). These findings support the conclusion that consumers are less responsive to price changes for fresh blueberries labeled as "Stay fresh" or "Sweety."

In terms of own and cross elasticities of all 14 fruits, we found that in the Control treatment (Table 2.12), bananas were the most own price inelastic fruit (-0.15) while cantaloupe (-1.059), watermelon (-0.921), and blackberries (-0.960) were the most own price elastic fruits. The scope of the price elasticity of fruits is consistent with what was found in the previous study (Andreyeva et al., 2010). In the case of berries, our findings showed that strawberries and blueberries were less responsive to changes in other berries' prices. Overall, results suggest that the demand for each type of berry is sensitive to price changes, with blackberries being the most sensitive with a price elasticity of -0.96, followed by strawberries (-0.811), raspberries (-0.785) and blueberries (-0.759). These results are consistent with the findings of Sobekova et al. (2013), which similarly reported that blueberries have lower price elasticities than other berries.

2.6 Conclusion and implications

This study investigated quality-related attributes of blueberries that may contribute to increasing the likelihood of consumers purchasing fresh blueberries. We collected survey data and found that the top five reasons for infrequent consumption of fresh blueberries were price-related, the short shelf-life of the fruit at home, concerns about its freshness, unavailability of fresh blueberries in the market, and packaging size. Among these reasons, price emerged as the respondents' primary concern. Our analysis using Multivariate Logit (MVL) models and elasticity analysis confirmed that quality-related descriptors on the label of packaged blueberries impact consumers' behavior, in terms of their sensitivity to price changes and likelihood of purchase. Specifically, we found that blueberries labeled with descriptors indicating a longer shelf life (e.g., "Stay fresh") were associated with reduced sensitivity to price changes among respondents. Instead of being substitutes for each other, our results suggested that berries (blueberries, blackberries, raspberries, and strawberries) complement each other. These berries are likely to be purchased together at the grocery store.

The findings of this study provide useful insights for blueberry growers, retailers, and marketers that could help inform the development of strategies to increase the per capita consumption of fresh blueberries. For example, this information could help stakeholders understand how highlighting distinct and desirable sensory qualities of their blueberries, differentiate them in the market and draw in consumers seeking specific sensory attributes, potentially leading to higher sales. Moreover, these findings can serve as a quality control mechanism for stakeholders to monitor and maintain the desired sensory characteristics (indicated on the label) ensuring consumers receive a consistent high-quality experience with each purchase occasion. Ultimately, this information offers valuable insights for shaping policies aimed at promoting the consumption of healthy fruits.

REFERENCES

Andreyeva, T., Long, M. W., & Brownell, K. D. (2010). The impact of food prices on consumption: a systematic review of research on the price elasticity of demand for food. American journal of public health, 100(2), 216-222.

Arnade, C., & Kuchler, F. (2015). Measuring the impacts of off-season berry imports (No. 1477-2016-121093).

Baselice, A., Colantuoni, F., Lass, D. A., Nardone, G., & Stasi, A. (2017). Trends in EU consumers' attitude towards fresh-cut fruit and vegetables. Food Quality and Preference, 59, 87-96.

Basu, A., Du, M., Leyva, M. J., Sanchez, K., Betts, N. M., Wu, M., ... & Lyons, T. J. (2010). Blueberries decrease cardiovascular risk factors in obese men and women with metabolic syndrome. The Journal of nutrition, 140(9), 1582-1587.

Bel, K., Fok, D, and Paap, R. (2018). Parameter Estimation in Multivariate Logit Models with Many Binary Choices. Econometric Reviews 37:534-550.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society: Series B (Methodological), 36(2), 192-225.

Blaker, K. M., Plotto, A., Baldwin, E. A., & Olmstead, J. W. (2014). Correlation between sensory and instrumental measurements of standard and crisp-texture southern highbush blueberries (Vaccinium corymbosum L. interspecific hybrids). Journal of the Science of Food and Agriculture, 94(13), 2785-2793.

Blackmore, H., Hidrio, C., & Yeomans, M. R. (2021). A taste of things to come: The effect of extrinsic and intrinsic cues on perceived properties of beer mediated by expectations. Food Quality and Preference, 94, 104326.

Brazelton, C. 2013. 2012 World blueberry acreage and production report. North American Highbush Blueberry Council. Folsom, CA.

Caputo, V., & Lusk, J. L. (2022). The basket-based choice experiment: a method for food demand policy analysis. Food Policy, 109, 102252.

Caputo, V., Scarpa, R., & Nayga Jr, R. M. (2017). Cue versus independent food attributes: the effect of adding attributes in choice experiments. European Review of Agricultural Economics, 44(2), 211-230.

Cassidy, A., Mukamal, K. J., Liu, L., Franz, M., Eliassen, A. H., & Rimm, E. B. (2013). High anthocyanin intake is associated with a reduced risk of myocardial infarction in young and middle-aged women. Circulation, 127(2), 188-196.

Cassidy, A., O'Reilly, É. J., Kay, C., Sampson, L., Franz, M., Forman, J. P., ... & Rimm, E. B. (2011). Habitual intake of flavonoid subclasses and incident hypertension in adults. The American journal of clinical nutrition, 93(2), 338-347.

Champ, P. A., Moore, R., & Bishop, R. C. (2009). A comparison of approaches to mitigate hypothetical bias. Agricultural and Resource Economics Review, 38(2), 166-180.

Cook, R. L. (2011). Fundamental forces affecting the US fresh berry and lettuce/leafy green subsectors. Choices, 26(4).

Cressie, N. (2015). Statistics for spatial data. John Wiley & Sons.

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. Public opinion quarterly, 64(4), 413-428.

D'Amato, A., Goeschl, T., Lorè, L., & Zoli, M. (2023). True to type? EU-style date marking and the valuation of perishable food. Food Policy, 114, 102381.

Darby, K., Batte, M. T., Ernst, S. C., & Roe, B. E. (2006). Willingness to pay for locally produced foods: A customer intercept study of direct market and grocery store shoppers (No. 379-2016-21958).

Devore, E. E., Kang, J. H., Breteler, M. M., & Grodstein, F. (2012). Dietary intakes of berries and flavonoids in relation to cognitive decline. Annals of neurology, 72(1), 135-143.

Donahue, D. W., & Work, T. M. (1998). Sensory and textural evaluation of Maine wild blueberries for the fresh pack market. Journal of texture studies, 29(3), 305-312.

Ehlenfeldt, M. K., & Martin, R. B. (2002). A survey of fruit firmness in highbush blueberry and species-introgressed blueberry cultivars. HortScience, 37(2), 386-389.

Fresh Trends. (2013). The Packer. Vance Publishing, Lenexa, KS.

Gallardo, R. K., Zhang, Q., Dossett, M., Polashock, J. J., Rodriguez-Saona, C., Vorsa, N., ... & Iorizzo, M. (2018). Breeding trait priorities of the blueberry industry in the United States and Canada. HortScience, 53(7), 1021-1028.

Gilbert, J. L., Olmstead, J. W., Colquhoun, T. A., Levin, L. A., Clark, D. G., & Moskowitz, H. R. (2014). Consumer-assisted selection of blueberry fruit quality traits. HortScience, 49(7), 864-873.

Girgenti, V., Massaglia, S., Mosso, A., Peano, C., & Brun, F. (2016). Exploring perceptions of raspberries and blueberries by Italian consumers. Sustainability, 8(10), 1027.

Hales, C. M., Carroll, M. D., Fryar, C. D., & Ogden, C. L. (2020). Prevalence of obesity and severe obesity among adults: United States, 2017–2018. NCHS Data Brief, no 360. National Center for Health Statistics, (360).

Hazebroek, B. K., & Croijmans, I. (2023). Let's talk over coffee: Exploring the effect of coffee flavour descriptions on consumer imagery and behaviour. Food Quality and Preference, 105, 104757.

Hoke, O., Campbell, B., Brand, M., & Hau, T. (2017). Impact of information on northeastern US consumer willingness to pay for aronia berries. HortScience, 52(3), 395-400.

Hu, W., Woods, T., & Bastin, S. (2009). Consumer acceptance and willingness to pay for blueberry products with nonconventional attributes. Journal of Agricultural and Applied Economics, 41(1), 47-60.

International Fresh Produce Association, 2022. Top 20 Fruits and Vegetables Sold in the U.S. Retrieved from https://www.freshproduce.com/resources/consumer-trends/top-20/. Accessed October 17, 2023

Istek, N., & Gurbuz, O. (2017). Investigation of the impact of blueberries on metabolic factors influencing health. Journal of Functional Foods, 38, 298-307.

Jennings, A., Welch, A. A., Fairweather-Tait, S. J., Kay, C., Minihane, A. M., Chowienczyk, P., ... & Cassidy, A. (2012). Higher anthocyanin intake is associated with lower arterial stiffness and central blood pressure in women. The American Journal of clinical nutrition, 96(4), 781-788.

Kaiser, H. M. (2015). An economic analysis of domestic market impacts of the US Highbush Blueberry Council. Report prepared for the US Highbush Blueberry Council, July, 21. Retrieved from https://ushbc.blueberry.org/wp-content/uploads/sites/5/2021/03/CornellStudyReportFINAL.pdf . Accessed October 17, 2023.

Kalt, W., Cassidy, A., Howard, L. R., Krikorian, R., Stull, A. J., Tremblay, F., & Zamora-Ros, R. (2020). Recent research on the health benefits of blueberries and their anthocyanins. Advances in Nutrition, 11(2), 224-236.

Kwak, K., Duvvuri, S. D., & Russell, G. J. (2015). An analysis of assortment choice in grocery retailing. Journal of Retailing, 91(1), 19-33.

Laaksonen, O., Knaapila, A., Niva, T., Deegan, K. C., & Sandell, M. (2016). Sensory properties and consumer characteristics contributing to liking of berries. Food quality and preference, 53, 117-126.

Liem, D. G., Aydin, N. T., & Zandstra, E. H. (2012). Effects of health labels on expected and actual taste perception of soup. Food Quality and Preference, 25(2), 192-197.

Mancino, L., & Kinsey, J. (2004). Diet quality and calories consumed: the impact of being hungrier, busier and eating out (pp. 04-02). Food Industry Center, University of Minnesota.

Mennella, J. A., & Bobowski, N. K. (2015). The sweetness and bitterness of childhood: Insights from basic research on taste preferences. Physiology & behavior, 152, 502-507.

Mennella, J. A., Colquhoun, T. A., Bobowski, N. K., Olmstead, J. W., Bartoshuk, L., & Clark, D. (2017). Farm to sensory lab: Taste of blueberry fruit by children and adults. Journal of food science, 82(7), 1713-1719.

Mezzetti, B., & Predieri, S. 2022. Survey on market potential and consumers’ inclination for new resilient and added value berries. National Council of Research. Retrieved from https://survey.cnr.it/index.php?r=survey/index&sid=567822&lang=en. Accessed October 17, 2023

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

Neill, C.L. and Lahne, J., 2022. Matching reality: A basket and expenditure based choice experiment with sensory preferences. Journal of choice modeling, 44, p.100369.

Papies, E. K., Barsalou, L. W., & Rusz, D. (2020a). Understanding desire for food and drink: A grounded-cognition approach. Current Directions in Psychological Science, 29(2), 193-198.

Papies, E. K., Johannes, N., Daneva, T., Semyte, G., & Kauhanen, L. L. (2020b). Using consumption and reward simulations to increase the appeal of plant-based foods. Appetite, 155, 104812.

Piqueras-Fiszman, B., & Spence, C. (2015). Sensory expectations based on product-extrinsic food cues: An interdisciplinary review of the empirical evidence and theoretical accounts. Food Quality and Preference, 40, 165-179.

Qu, S., Lamm, A. J., & Rumble, J. N. (2017). Marketing Power Berries: An Importance-Performance Analysis of Blueberry. Journal of Applied Communications, 101(3).

Richards, T. J., Hamilton, S. F., & Yonezawa, K. (2018). Retail market power in a shopping basket model of supermarket competition. Journal of Retailing, 94(3), 328-342.

Russell, G. J., & Petersen, A. (2000). Analysis of cross category dependence in market basket selection. Journal of Retailing, 76(3), 367-392.

Saftner, R., Polashock, J., Ehlenfeldt, M., & Vinyard, B. (2008). Instrumental and sensory quality characteristics of blueberry fruit from twelve cultivars. Postharvest Biology and Technology, 49(1), 19-26.

Sater, H., Ferrão, L. F. V., Olmstead, J., Munoz, P. R., Bai, J., Hopf, A., & Plotto, A. (2021). Exploring environmental and storage factors affecting sensory, physical and chemical attributes of six southern highbush blueberry cultivars. Scientia Horticulturae, 289, 110468.

Sharma, S. P., Chung, H. J., Kim, H. J., & Hong, S. T. (2016). Paradoxical effects of fruit on obesity. Nutrients, 8(10), 633.

Shi, L., House, L., & Gao, Z. (2011). Consumer WTP for blueberry attributes: A hierarchical Bayesian approach in the WTP space (No. 321-2016-10989).

Shi, L., House, L. A., & Gao, Z. (2013). Impact of purchase intentions on full and partial bids in BDM auctions: Willingness-to-pay for organic and local blueberries. Journal of Agricultural Economics, 64(3), 707-718.

Shi, L., Gao, S., House, L. A., & Heng, Y. (2015). A multi-store auction to measure willingness-to-pay for organic and local blueberries. Journal of Food Distribution Research, 46(856-2016-58191), 153-162.

Sobekova, K., Thomsen, M. R., & Ahrendsen, B. L. (2013). Market trends and consumer demand for fresh berries. APSTRACT: Applied Studies in Agribusiness and Commerce, 7(1033-2016-84241), 11-14.

Song, I., & Chintagunta, P. K. (2006). Measuring cross-category price effects with aggregate store data. Management Science, 52(10), 1594-1609.

Stevens, T., Hoshide, A. K., & Drummond, F. A. (2015). Willingness to pay for native pollination of blueberries: A conjoint analysis. Intl. J. Agric. Marketing, 2, 068-077.

The Packer, 2019. Infographic: Berry purchases by region. Retrieved from https://www.thepacker.com/markets/fruit/infographic-berry-purchases-region. Accessed October 17, 2023

Train, K. E. (2009). Discrete choice methods with simulation. Cambridge university press.

Turnwald, B. P., & Crum, A. J. (2019). Smart food policy for healthy food labeling: Leading with taste, not healthiness, to shift consumption and enjoyment of healthy foods. Preventive medicine, 119, 7-13.

U.S. Census Bureau, 2021. U.S.C. Bureau.

USDA. (2019). What foods are in the Vegetable Group? Retrieved from https://ask.usda.gov/s/article/What-foods-are-in-the-Vegetable-Group. Accessed October 17, 2023.

USDA. (2020). New to Market Product Report – Fresh Blueberries. Retrieved from https://apps.fas.usda.gov/newgainapi/api/Report/DownloadReportByFileName?fileName=New%20to%20Market%20Product%20Report%20-%20Fresh%20Blueberries_Beijing%20ATO_China%20-%20Peoples%20Republic%20of_06-11-2020. Accessed October 17, 2023

USDA. (2023a). Food Availability (Per Capita) Data System. Fruit (fresh). Retrieved from https://www.ers.usda.gov/data-products/food-availability-per-capita-data-system/. Accessed October 17, 2023.

USDA. (2023b). Agricultural Marketing Service. Weekly Advertised Fruit & Vegetable Retail Prices. Retrieved from https://www.marketnews.usda.gov/mnp/fv-report-retail?category=retail&portal=fv&startIndex=1&class=ALL. Accessed October 17, 2023.

USHBC. (2018). Annual Report. Retrieved from https://blueberry.org/ushbc/wp-content/uploads/sites/5/2020/12/Blueberry-18-19-Annual-Report.pdf. Accessed May 9, 2023

USHBC. (2022). The State of the U.S. Blueberry Consumer IPSOS Usage and Attitude Survey. Presented at the USHBS Fall Summit Keynote, October 6.

Wedick, N. M., Pan, A., Cassidy, A., Rimm, E. B., Sampson, L., Rosner, B., ... & van Dam, R. M. (2012). Dietary flavonoid intakes and risk of type 2 diabetes in US men and women. The American journal of clinical nutrition, 95(4), 925-933.

Wolfe, K. L., Kang, X., He, X., Dong, M., Zhang, Q., & Liu, R. H. (2008). Cellular antioxidant activity of common fruits. Journal of agricultural and food chemistry, 56(18), 8418-8426.

Woods, A. T., Lloyd, D. M., Kuenzel, J., Poliakoff, E., Dijksterhuis, G. B., & Thomas, A. (2011). Expected taste intensity affects response to sweet drinks in primary taste cortex. Neuroreport, 22(8), 365-369.

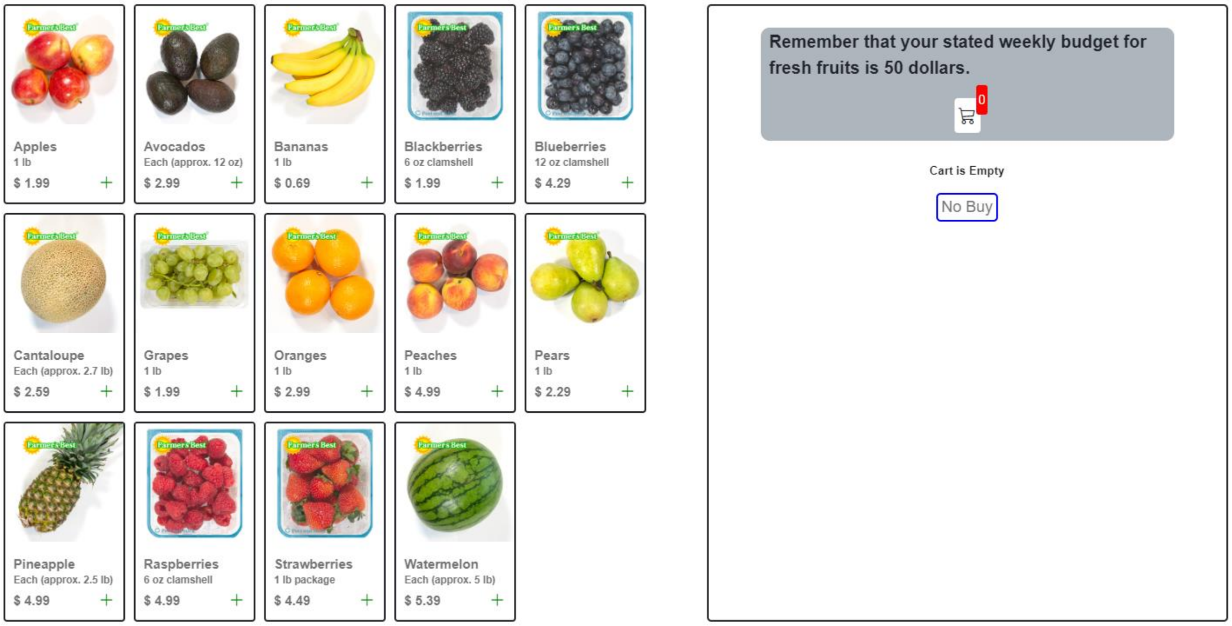
World Health Organization. 2021. Obesity and overweight. Key Facts. Newsroom Factsheets. Retrieved from https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight. Accessed October 17, 2023

Yue, C., & Wang, J. (2016, April). Consumer preferences for fresh blueberry attributes. In XI International Vaccinium Symposium 1180 (pp. 1-8).

Zheng, X., Yue, C., Gallardo, K., McCracken, V., Luby, J., & McFerson, J. (2016). What attributes are consumers looking for in sweet cherries? Evidence from choice experiments. Agricultural and Resource Economics Review, 45(1), 124-142.

Figure 2.1: Example of a fruit-basket choice scenario

Panel A. Example of a fruit-basket choice scenario - empty cart.



Panel B. Example of a fruit-basket scenario with the selected fruits placed in the shopping cart.

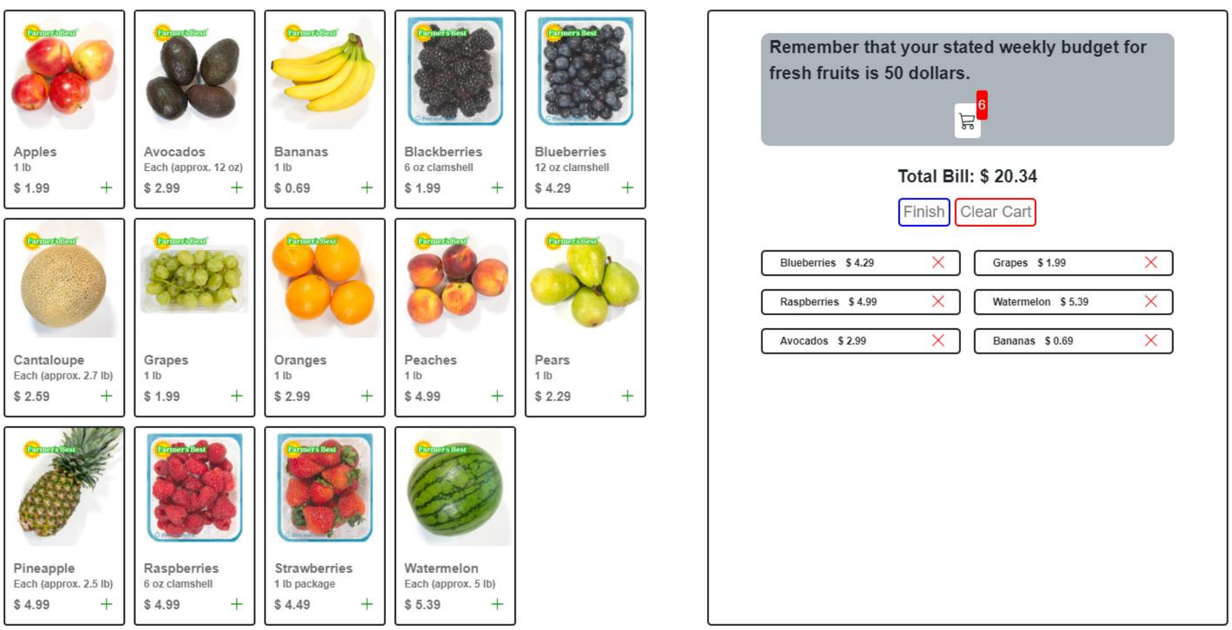
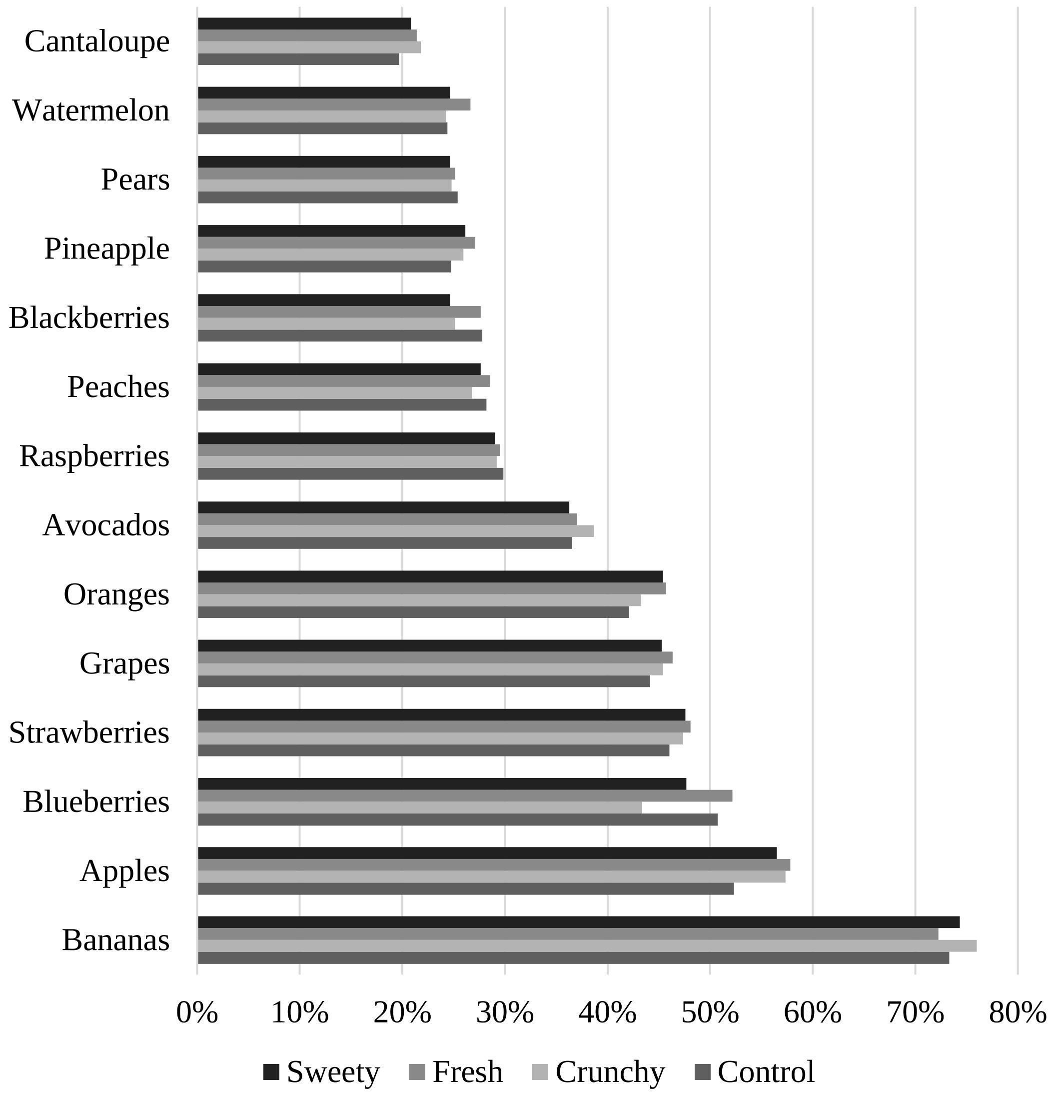


Figure 2.2: Blueberry clamshell labels used in each treatment.

Figure 2.3: Percent of times each fruit was chosen, across all treatments.



Note: The label treatment (“Sweety”, “Stay fresh”, “Crunchy”) was only applied to blueberries, the other fruits displayed the exact same generic label across all treatments.

Figure 2.4: Histogram of individual own-price elasticity of blueberries.

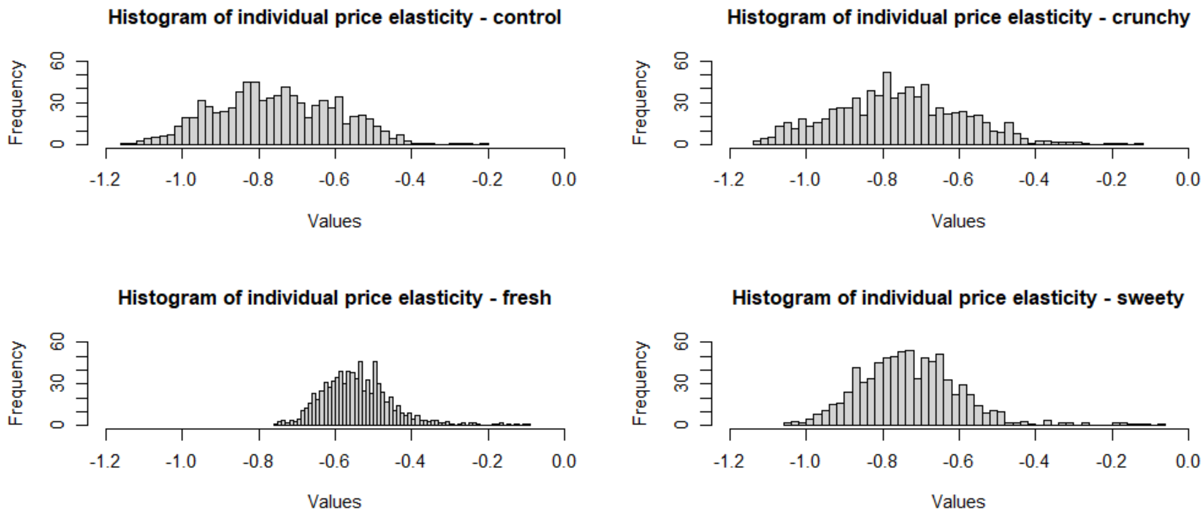


Table 2.1: Price levels used in the basket-based choice experiment.

Product	Unit	Prices		
		Low	Medium	High
Apples	Per pound	$0.99	$1.99	$2.99
Avocados	Each (approximate 12 oz)	$0.79	$1.89	$2.99
Bananas	Per pound	$0.49	$0.69	$0.89
Blackberries	6 oz package	$1.99	$3.59	$5.19
Blueberries	12 oz package	$2.59	$4.29	$5.99
Cantaloupe	Each (approximate 2.7 lb)	$2.59	$3.79	$4.99
Grapes	Per pound	$1.99	$3.49	$4.99
Oranges	Per pound	$0.79	$1.89	$2.99
Peaches	Per pound	$1.59	$3.29	$4.99
Pears	Per pound	$1.09	$2.29	$3.49
Pineapple	Each (approximate 2.5 lb)	$1.99	$3.49	$4.99
Raspberries	6 oz package	$1.79	$3.39	$4.99
Strawberries	1 lb package	$2.99	$4.49	$5.99
Watermelon	Each (approximate 5 lb)	$3.79	$5.39	$6.99

Table 2.2: Socio-demographic characteristics of the pooled sample and across treatment

	Description	U.S. Census 2021	Pooled sample All treatments N=3,208	Pairwise comparison between pooled sample and U.S. Census 2021 (t-value)	Treatment sample Treatments Control N=801	Crunchy N=802	Fresh N=805	Sweety N=800
Female	1 if female; 0 otherwise	0.51	0.64	15.21***	0.63	0.67	0.62	0.63
Millennial	1 if born in or after 1982; 0 otherwise		0.56		0.57	0.56	0.55	0.56
High income	1 if $75,000/year or more; 0 otherwise	0.47	0.45	2.12**	0.45	0.48	0.42	0.46
Children	1 if ≥1 child under 18; 0 otherwise	0.33	0.37	4.48***	0.37	0.38	0.37	0.35
Employed	1 if employed; 0 otherwise		0.62		0.62	0.62	0.62	0.62
College	1 if minimum of 4-year college degree; 0 otherwise	0.32	0.54	24.69***	0.54	0.54	0.53	0.54
White	1 if white; 0 otherwise	0.61	0.74	17.26***	0.73	0.74	0.73	0.77
N			3208		801	802	805	800

[1] Single, double, and triple asterisks (*, **, ***) indicate statistical significance at the 10%, 5%, and 1% levels.

Table 2.3: Summary statistics of sociodemographics across all treatment samples.

		Treatments				
	Description	Control N=801	Crunchy N=802	Fresh N=805	Sweety N=800	Pooled N=3,208
Gender	Female	63.09	67.21	61.99	63.25	63.90
	Male or other	36.91	32.79	38.01	36.75	36.10
Age	24 years or less	6.90	7.61	8.94	8.13	7.89
	25-34 years	32.90	32.42	30.31	31.75	31.86
	35-44 years	24.00	22.82	23.73	22.88	23.47
	45-54 years	21.00	22.07	21.74	22.38	21.70
	55-64 years	5.00	4.49	5.59	4.00	4.83
	65+ years	10.00	10.60	9.69	10.88	10.26
	Median age	38.00	38.00	40.00	39.00	39.00
College	Less than high school graduate	1.00	0.87	1.61	1.88	1.31
	High school graduate – includes equivalence	12.00	13.84	14.91	11.00	13.03
	Some college	21.00	21.95	21.12	22.25	21.51
	2-year degree	12.00	9.48	9.32	10.75	10.41
	4-year degree	35.00	34.66	35.90	33.25	34.73
	Post-graduate degree (Master, PhD, etc.)	19.00	19.20	17.14	20.88	19.01
Income	<$25,000/year	10.00	11.72	10.68	10.88	10.88
	$25,000–$34,999/year	10.00	9.48	9.32	7.88	9.26
	$35,000–$49,999/year	13.00	11.10	12.55	11.75	12.09
	$50,000–$74,999/year	19.00	16.21	21.37	20.00	19.17

	$75,000–$99,999/year	16.00	17.96	12.17	15.25	15.34
	$100,000–$149,999/year	19.00	18.70	17.39	20.50	18.98
	$150,000–$199,999/year	6.00	6.98	7.83	5.63	6.61
	$200,000/year or more	3.00	4.24	4.72	4.50	4.21
Household size	1	17.98	15.96	16.52	18.38	17.21
	2	32.83	31.92	32.55	32.88	32.54
	3	21.47	20.57	22.86	21.75	21.66
	4	17.35	19.08	17.89	18.75	18.27
	5	6.99	7.86	6.71	5.00	6.64
	6 or more	3.37	4.61	3.48	3.25	3.68
Children	0	62.67	62.34	63.11	64.63	63.19
	1	18.10	15.21	16.77	17.00	16.77
	2	13.73	14.34	13.42	14.00	13.87
	3	4.24	5.86	4.22	2.50	4.21
	4 or more	1.25	2.24	2.48	1.88	1.96
Employment status	Employed	61.80	61.72	62.24	62.25	62.00
	Student	3.87	6.23	6.09	4.63	5.21
	Unemployed, but seeking employment	11.99	10.85	9.57	10.00	10.60
	Retired	10.86	10.35	9.19	10.13	10.13
	Free lancer	5.24	3.24	6.71	5.25	5.11
Food expenditure per month	Less than $300	14.61	15.96	14.41	16.25	15.31
	$300–$599	43.57	43.52	42.48	44.00	43.39
	$600 –$899	20.22	19.33	20.75	22.25	20.64

	$900 –$1,199	8.86	10.72	8.94	6.75	8.82
	$1,200 – $1,499	7.24	6.11	5.09	4.50	5.74
	$1,500–$1,799	2.50	1.50	3.23	1.63	2.21
	$1,800–$2,099	0.87	1.37	0.87	1.38	1.12
	$2,100 or more	0.62	0.25	0.99	1.00	0.72
Fresh fruit expenditure per month	Less than $50	34.21	37.91	33.17	38.38	35.91
	$50–$99	38.33	36.53	34.91	35.63	36.35
	$100 –$149	14.36	16.46	16.02	14.75	15.40
	$150 –$199	5.49	3.24	5.71	4.88	4.83
	$200 – $249	3.25	2.24	3.48	1.38	2.59
	$250–$299	2.00	0.62	1.74	1.63	1.50
	$300–$349	0.62	1.37	1.99	1.13	1.28
	$350 or more	1.12	0.75	1.24	1.00	1.03
Community	Rural area	19.23	24.44	21.37	22.50	21.88
	Urban area	26.72	25.06	30.31	27.75	27.46
	Suburban area	54.06	50.50	48.32	49.75	50.65
Race	White	73.41	74.31	72.55	77.00	74.31
	Black or African American	9.74	8.35	12.92	6.63	9.41
	American Indian or Alaska Native	0.75	1.50	1.12	0.88	1.06
	Asian	9.86	9.35	7.45	7.75	8.60
	Native Hawaiian and other Pacific Islander	0.37	0.25	0.25	0.38	0.31
	Hispanic or Latino	5.87	6.23	5.71	7.38	6.30
Politics	Extremely liberal	15.61	14.84	15.16	12.38	14.50

	Slightly liberal	20.47	18.70	15.65	20.25	18.77
	Moderate or middle of the road	27.22	25.69	26.46	28.13	26.87
	Slightly conservative	16.48	16.46	17.14	16.63	16.68
	Extremely conservative	8.99	8.85	9.94	9.75	9.38
	Mixed: liberal on some issues and conservative on other issues	6.87	9.48	8.20	7.38	7.98
Health status	Very unhealthy	1.75	2.24	1.74	1.88	1.90
	Somewhat unhealthy	4.49	5.36	6.46	6.50	5.70
	Slightly unhealthy	12.11	11.72	10.19	12.13	11.53
	Slightly healthy	19.60	20.70	18.88	19.13	19.58
	Somewhat healthy	44.32	43.89	44.97	43.25	44.11
	Very healthy	17.73	16.08	17.76	17.13	17.18
Physically fitness	Not at all physically fit	5.87	5.61	5.59	4.88	5.49
	Not very physically fit	20.72	21.82	21.49	21.38	21.35
	Somewhat physically fit	48.06	51.50	50.68	51.25	50.37
	Very physically fit	21.22	18.33	18.63	18.13	19.08
	Extremely physically fit	4.12	2.74	3.60	4.38	3.71
Diet-related chronic diseases	High blood pressure	21.72	22.57	23.85	21.50	22.41
	Diabetes	8.49	7.36	9.81	9.25	8.73
	High cholesterol	17.23	21.20	21.99	18.38	19.70
	Heart disease	2.50	1.75	3.23	3.50	2.74
Region	Midwest	20.60	21.07	20.87	20.63	20.79
	Northeast	17.10	17.21	16.89	17.38	17.14

	South	38.33	37.91	38.01	38.50	38.19
	West	23.97	23.82	24.22	23.50	23.88
N		801	802	805	800	3208

Table 2.4: Frequency distribution of responses to fresh fruit purchase habit questions.

*Percentage of respondents who have consumed the **fresh** fruits in the past 3 months*	
Bananas	89.01
Apples	86.64
Blueberries	79.03
Grapes	78.40
Oranges	75.66
Strawberries	73.66
Avocados	66.04
Blackberries	49.94
Raspberries	49.94
Pineapple	49.06
Watermelon	45.07
Cantaloupe	44.44
Peaches	40.95
Pears	40.45
Percentage of respondents indicated the most frequent form of consuming blueberries	
Fresh	79.28
Frozen	16.85
Dry	1.50
Juice	1.00
Canned	0.75
Other	0.62
Percentage of respondents stated how they typically consume fresh blueberries	
Raw, alone	70.29
Topping, add to granola or yogurt	45.69
Beverages, smoothies	41.32
Cooked, pancakes, muffins, waffles	38.70
Sauces, jams, jellies	13.36
Other	3.62
Cooked, savory, BBQ, salsa	3.00
Percentage of respondents indicated how often they purchase blueberries	
Once every 2 to 3 weeks	25.72
Once a week	23.97
Less than once a month	20.35
Once a month	14.36
2-3 times per week	12.36

4-6 times per week	3.25
Percentage of respondents indicated the way they purchase fresh blueberries	
Regular planned purchased	63.05
Impulse purchase	21.85
Planned purchase triggered by recipe, special occasion	15.11
Percentage of respondents indicated how often they consume blueberries	
2-3 times per week	21.60
Less than once a month	15.98
Once a day	15.48
4-6 times per week	12.23
Once every 2 to 3 weeks	11.86
Once a week	9.86
Several times a day	6.49
Once a month	6.49
Percentage of respondents indicated they consume less than once a month, and reasons	
Price	54.36
I buy blueberries at least once a month	22.72
Short shelf-life at home	22.07
Freshness	20.07
No fresh blueberries were available	18.08
Package size	16.08
Taste	12.22
Expiration/ sell-by date	8.60
Texture	8.48
Color	6.73
Smell	5.61
Brand	5.11
Place of origin	3.99
Not available organically grown	3.49
Other	2.74
Non-GMO not available	1.87
Percentage of respondents indicated where do they buy fresh blueberries regularly	
Grocery store (e.g., Kroger, Albertsons, Safeway, local chains)	50.81
Supercenter (e.g., Walmart, Target, Meijer)	23.10
Limited assortment (e.g., Grocery Outlet, Lidl, Aldi's, Save-A-Lot)	6.37
Warehouse Club (e.g., Costco, Sam's Club, BJ's)	6.12

Specialty, organic (e.g., Whole Foods, Trader Joe’s, New Seasons, Sprouts)	4.49
Farmers market	4.12
Retailer online (e.g., Safeway, Walmart, Kroger)	1.62
Discount stores (e.g., WinCo, Fareway)	1.25
Online-only food stores (Amazon Grocery, Fresh Direct, Local Harvest, ShopFoodEx)	0.87
Other	0.50
Drug Store (e.g., CVS pharmacy, Walgreens, Rite Aid)	0.25
Ethnic Food Store (e.g., H-mart, 99 ranch market, Hong Kong Supermarket)	0.25
Online Third Party (e.g., Instacart)	0.25
Convenience Store (e.g., 7-Eleven, Circle K, Speedway, Casey’s General Store)	0.00
Weighted average poundage of fresh blueberries purchased stated in one shopping occasion	
Fresh blueberries	0.80

Table 2.5: Weighted average of ratings of importance of different issues related to fresh blueberry consumption.

Weighted average importance of fresh blueberry attributes (1= most important, 5= least important)	
Freshness	1.73
Blueberries appear free from defects	2.62
Ripeness	2.63
Phytonutrient content	2.69
Sweetness	2.82
Weighted average importance of fresh blueberry characteristics to be improved (1= most important, 5= least important)	
Fresh blueberries with improved eating quality traits	2.44
Fresh blueberries with improved visual quality traits	2.65
Fresh blueberries that stay fresh longer	2.75
Blueberry plants with improved response to climate change	3.20
Fresh blueberries with improved nutritional traits	3.28
Weighted average of how the statement associate fresh blueberry consumption (1= not at all associated, 5= strongly associated)	
A natural product	4.27
Can eat at breakfast	4.23
Good source of vitamins and minerals	4.20
Fast to prepare	4.18
Can eat with family	4.07
Can use it as dessert ingredient	4.04
Suitable for children	4.02
Can preserve by freezing	4.02
Boosts immunity	3.89
Can eat with friends	3.89
Beneficial for brain health/memory	3.80
An environmentally friendly product	3.80
Supports overall gut health and healthy digestive system	3.73
Grown in the United States	3.64
Could be grown in most regions of the United States	3.54
Good source of fiber	3.50
Reduces risk factors associated with heart disease	3.46
Reduces risk of certain cancers	3.40
Helps maintain normal blood sugar levels, helps prevent Type 2 diabetes	3.39
Lowers bad cholesterol levels	3.37
Lower blood pressure	3.36

Affordable to consume everyday	3.32
Aids in exercise performance and recovery	3.15
Helps maintain healthy bones	3.09
Reminds me of childhood	3.06
Weighted average to attention to labels (1= Totally Irrelevant, 5= Crucial)	
Pesticide-free	3.13
Domestic product	3.08
Not genetically engineered	3.02
Healthy benefits	2.91
Organic	2.77
Sustainable agriculture	2.75
Local origin	2.74
Eco-label	2.64
Farmer owned	2.63
Name of the blueberry variety	2.14
A private brand	1.77

Table 2.6: Comparison of model specifications.

Treatment	Control N=801		Crunchy N=802		Fresh N=805		Sweety N=800		Pooled N=3,208	
Model	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Fruit-specific price effect	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N. of parameters	422	435	422	435	422	435	422	435	422	435
N. of choices	4806	4806	4812	4812	4830	4830	4800	4800	19248	19248
AIC	72305	72295	71848	71845	73584	73585	72481	72458	292035	291960
AICC	72386	72382	71929	71931	73665	73672	72562	72545	292054	291980
BIC	75038	75113	74582	74663	76319	76405	75214	75276	295355	295381
Loglikelihood	-35730.5	-35712.5	-35502	-35487.5	-36370	-36357.5	-35818.5	-35794	-145596	-145545

Table 2.7: Summary statistics of pooled sample and across treatment groups.

		Pooled sample N=3,208	Treatment samples			
			Control N=801	Crunchy N=802	Fresh N=805	Sweety N=800
Liberal[1]	1 if liberal; 0 otherwise	0.33	0.36	0.34	0.31	0.33
Conservative[1]	1 if conservative; 0 otherwise	0.26	0.26	0.25	0.27	0.26
Northeast	1 if northeast; 0 otherwise	0.17	0.17	0.17	0.17	0.17
West	1 if west; 0 otherwise	0.24	0.24	0.24	0.24	0.24
South	1 if south; 0 otherwise	0.38	0.38	0.38	0.38	0.39
Physically fit[2]	1 if physically fit; 0 otherwise	0.23	0.25	0.21	0.22	0.23
Diabetes	1 if diabetes; 0 otherwise	0.09	0.09	0.07	0.10	0.09
Cholesterol	1 if high cholesterol; 0 otherwise	0.20	0.17	0.21	0.22	0.18
Main_nutrition[3]	1 if nutrition was ranked 1, 2, or 3; 0 otherwise	0.78	0.77	0.79	0.78	0.76
Fresh fruit weekly budget	Continuous	36.75	39.42	32.10	40.85	37.08
Label_domestic	1 if label "domestic product" was marked important or crucial; 0 otherwise	0.44	0.43	0.42	0.44	0.46
Label_organic	1 if label "organic" was marked important or crucial; 0 otherwise	0.35	0.35	0.35	0.38	0.31
Label_nonGMO	1 if label "not genetically engineered" was marked important or crucial;0 otherwise	0.44	0.45	0.43	0.47	0.40
Label_healthy	1 if label "healthy benefits" was marked important or crucial; 0 otherwise	0.41	0.42	0.40	0.43	0.39

[1] The reference level for the categories Liberal and Conservative was Mixed, which referred to respondents who identified themselves as "Moderate or middle of the road" or "Mixed: liberal on some issues and conservative on other issues" when asked about their political views.
[2] Physically fit was 1 if the respondent identified themselves as physically fit.
[3] Main_nutrition was 1 if the respondent ranked health and nutrition as one of the top three factors influencing their overall food choices and eating patterns.

Table 2.8: Baseline utility estimates from the multivariate logit model.

	Probability of choosing blueberries under the sample treatments				
	Control N=801	Crunchy N=802	Fresh N=805	Sweety N=800	Pool N=3,208
Constant	-0.103	-1.150***	-0.620***	-0.168	-0.453***
	(0.188)	(0.197)	(0.187)	(0.190)	(0.093)
Price	-0.371***	-0.317***	-0.274***	-0.333***	-0.318***
	(0.02413)	(0.02406)	(0.02375)	(0.02387)	(0.01183)
Female	-0.188***	-0.029	-0.201***	0.039	-0.111***
	(0.070)	(0.072)	(0.071)	(0.070)	(0.035)
Millennial	-0.138*	-0.183**	-0.237***	-0.288***	-0.210***
	(0.073)	(0.074)	(0.074)	(0.073)	(0.036)
Income ≥$75,000/yr	-0.017	0.069	-0.027	0.129*	0.025
	(0.073)	(0.074)	(0.075)	(0.073)	(0.036)
With at least one child	0.118	-0.190***	-0.180**	-0.083	-0.088**
	(0.072)	(0.071)	(0.072)	(0.072)	(0.035)
Employed	0.273***	0.049	0.126*	0.065	0.114***
	(0.071)	(0.070)	(0.070)	(0.071)	(0.034)
4-yr college degree	-0.046	0.203***	0.079	0.126*	0.092***
	(0.072)	(0.071)	(0.071)	(0.070)	(0.035)
White	0.136*	0.327***	0.229***	0.150*	0.208***
	(0.079)	(0.079)	(0.077)	(0.081)	(0.039)
Liberal	-0.037	-0.074	0.164**	-0.181**	-0.021
	(0.077)	(0.078)	(0.079)	(0.078)	(0.038)
Conservative	-0.176**	-0.096	0.029	-0.129	-0.090**
	(0.085)	(0.084)	(0.082)	(0.083)	(0.041)
Northeast	0.271**	0.063	0.324***	0.028	0.171***
	(0.107)	(0.106)	(0.107)	(0.106)	(0.052)
West	0.080	0.135	0.095	-0.242**	0.019
	(0.100)	(0.099)	(0.098)	(0.101)	(0.049)
South	-0.124	0.035	0.176*	0.031	0.049
	(0.090)	(0.090)	(0.090)	(0.089)	(0.044)
Physically fit	0.069	0.058	0.132	0.238***	0.124***
	(0.081)	(0.084)	(0.086)	(0.083)	(0.041)
Diagnosed with diabetes	-0.278**	0.023	-0.004	-0.304**	-0.149**
	(0.122)	(0.131)	(0.118)	(0.122)	(0.060)
Diagnosed with high cholesterol	0.000	-0.203**	-0.065	-0.196**	-0.113***
	(0.093)	(0.085)	(0.086)	(0.093)	(0.043)
Health & nutrition is important	0.230***	0.544***	0.058	-0.044	0.158***
	(0.080)	(0.086)	(0.081)	(0.077)	(0.039)
Fresh fruit weekly budget	0.000	0.003***	0.003***	0.003***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Label domestic is important	0.065	-0.016	-0.051	0.082	0.019
	(0.072)	(0.073)	(0.071)	(0.072)	(0.035)
Label organic is important	0.075	0.367***	0.046	0.125	0.145***
	(0.083)	(0.083)	(0.083)	(0.085)	(0.041)
Label non-GMO is important	0.237***	0.125	0.109	-0.172**	0.089**
	(0.081)	(0.081)	(0.081)	(0.084)	(0.040)
Label healthy is important	0.233***	0.106	0.075	0.353***	0.190***
	(0.075)	(0.076)	(0.074)	(0.076)	(0.037)

[1] Single, double, and triple asterisks (*, **, ***) indicate statistical significance at the 10%, 5%, and 1% levels.
[2] Standard errors are in parentheses.

Table 2.9: Baseline utility estimates from the MVL model – Sample of respondents presented the control treatment.

	Apples	Avocados	Bananas	Blackberries	Blueberries	Cantaloupe	Grapes	Oranges	Peaches	Pears	Pineapple	Raspberries	Strawberries	Watermelon	No buy
Constant	-0.855***	-2.464***	0.804***	-1.344***	-0.103	-1.366***	-0.455**	-1.218***	-1.955***	-1.457***	-1.390***	-1.624***	-0.791***	-1.523***	-2.171***
	(0.178)	(0.184)	(0.232)	(0.204)	(0.188)	(0.256)	(0.184)	(0.178)	(0.199)	(0.201)	(0.215)	(0.198)	(0.203)	(0.243)	(0.274)
Price	-0.409***	-0.374***	-0.834***	-0.372***	-0.371***	-0.344***	-0.390***	-0.306***	-0.313***	-0.353***	-0.316***	-0.330***	-0.341***	-0.223***	-
	(0.040)	(0.037)	(0.224)	(0.029)	(0.024)	(0.042)	(0.027)	(0.037)	(0.026)	(0.038)	(0.031)	(0.028)	(0.027)	(0.030)	
Female	0.057	0.250***	-0.112	0.070	-0.188***	0.010	0.144**	-0.146**	-0.068	-0.142*	-0.127	0.331***	0.269***	0.024	0.061
	(0.071)	(0.072)	(0.078)	(0.080)	(0.070)	(0.089)	(0.071)	(0.072)	(0.078)	(0.079)	(0.083)	(0.079)	(0.072)	(0.084)	(0.135)
Millennial	0.094	0.098	-0.041	0.053	-0.138*	-0.695***	-0.149**	0.048	0.099	-0.340***	-0.296***	0.207***	0.189**	0.058	0.395***
	(0.073)	(0.074)	(0.082)	(0.082)	(0.073)	(0.089)	(0.073)	(0.074)	(0.081)	(0.082)	(0.084)	(0.080)	(0.075)	(0.086)	(0.143)
Income ≥ $75K/yr	0.064	-0.048	0.152*	-0.067	-0.017	0.267***	0.061	-0.317***	-0.054	-0.037	-0.075	0.127	0.217***	0.057	-0.054
	(0.073)	(0.074)	(0.081)	(0.082)	(0.073)	(0.091)	(0.073)	(0.074)	(0.081)	(0.083)	(0.085)	(0.079)	(0.075)	(0.086)	(0.139)
W/at least 1 child	0.124*	0.136*	-0.029	0.042	0.118	0.034	0.187***	0.087	-0.211***	0.272***	-0.208**	-0.214***	0.224***	0.316***	0.063
	(0.073)	(0.072)	(0.081)	(0.080)	(0.072)	(0.090)	(0.072)	(0.073)	(0.080)	(0.081)	(0.084)	(0.079)	(0.073)	(0.083)	(0.135)
Employed	-0.366***	0.062	0.111	0.186**	0.273***	0.053	0.040	0.186**	0.056	-0.315***	0.076	0.040	0.278***	-0.078	-0.420***
	(0.073)	(0.073)	(0.080)	(0.082)	(0.071)	(0.089)	(0.072)	(0.073)	(0.080)	(0.080)	(0.083)	(0.079)	(0.074)	(0.085)	(0.133)
4-yr college degree	0.235***	0.067	-0.168**	0.159*	-0.046	-0.192**	-0.253***	0.000	0.163**	0.206**	-0.148*	0.190**	-0.210***	0.060	-0.239*
	(0.072)	(0.073)	(0.080)	(0.081)	(0.072)	(0.090)	(0.072)	(0.073)	(0.080)	(0.082)	(0.084)	(0.079)	(0.074)	(0.085)	(0.136)
White	0.135*	0.040	-0.142	0.168*	0.136*	-0.182*	-0.011	-0.457***	-0.120	-0.184**	-0.251***	0.238***	-0.002	-0.174*	0.063
	(0.079)	(0.079)	(0.089)	(0.089)	(0.079)	(0.095)	(0.079)	(0.079)	(0.086)	(0.087)	(0.089)	(0.087)	(0.081)	(0.09)	(0.150)
Liberal	0.133*	-0.026	-0.036	-0.044	-0.037	-0.058	0.116	-0.088	0.382***	0.304***	0.008	0.148*	-0.186**	-0.521***	-0.397***
	(0.077)	(0.078)	(0.085)	(0.087)	(0.077)	(0.096)	(0.078)	(0.078)	(0.086)	(0.087)	(0.089)	(0.084)	(0.079)	(0.093)	(0.152)
Conservative	0.012	-0.141	-0.023	0.034	-0.176**	-0.283***	-0.001	-0.009	0.409***	0.095	-0.189*	-0.035	-0.043	-0.009	0.024
	(0.086)	(0.087)	(0.095)	(0.095)	(0.085)	(0.107)	(0.086)	(0.086)	(0.095)	(0.098)	(0.100)	(0.094)	(0.087)	(0.098)	(0.154)
Northeast	-0.238**	0.104	-0.069	-0.124	0.271**	0.233*	0.205*	-0.010	-0.028	0.176	0.137	-0.233**	-0.108	-0.221*	-0.146
	(0.108)	(0.112)	(0.120)	(0.122)	(0.107)	(0.138)	(0.108)	(0.108)	(0.119)	(0.118)	(0.124)	(0.117)	(0.110)	(0.127)	(0.191)
West	-0.24**	0.889***	-0.131	-0.074	0.080	0.172	0.018	-0.210**	0.004	-0.211*	-0.124	-0.032	-0.158	-0.173	-0.216
	(0.101)	(0.103)	(0.112)	(0.114)	(0.100)	(0.130)	(0.102)	(0.102)	(0.112)	(0.113)	(0.119)	(0.108)	(0.103)	(0.120)	(0.181)
South	-0.207**	0.286***	-0.036	0.142	-0.124	0.310***	0.263***	-0.146	0.003	-0.255**	0.010	-0.343***	0.087	-0.065	-0.452***
	(0.091)	(0.094)	(0.101)	(0.101)	(0.09)	(0.116)	(0.091)	(0.091)	(0.101)	(0.102)	(0.106)	(0.098)	(0.092)	(0.105)	(0.164)
Physically fit	-0.030	0.388***	-0.310***	0.013	0.069	-0.404***	-0.105	-0.001	0.074	-0.002	0.242***	-0.129	-0.186**	0.054	0.091
	(0.082)	(0.081)	(0.089)	(0.089)	(0.081)	(0.104)	(0.082)	(0.082)	(0.089)	(0.091)	(0.092)	(0.089)	(0.083)	(0.094)	(0.156)
Diag. diabetes	-0.129	-0.275**	-0.084	0.327**	-0.278**	-0.093	0.125	0.339***	0.343***	0.014	-0.104	-0.117	0.062	0.033	-0.146
	(0.124)	(0.127)	(0.135)	(0.133)	(0.122)	(0.148)	(0.123)	(0.124)	(0.129)	(0.133)	(0.143)	(0.133)	(0.124)	(0.141)	(0.252)
Diag. cholesterol	0.077	0.257***	-0.101	-0.105	0.000	-0.278**	-0.203**	0.156*	0.128	0.052	-0.300***	0.166	0.010	0.008	0.079
	(0.093)	(0.095)	(0.103)	(0.105)	(0.093)	(0.117)	(0.094)	(0.095)	(0.102)	(0.104)	(0.112)	(0.101)	(0.096)	(0.113)	(0.177)
Influenced by health & nutrition	-0.035	0.339***	0.172*	-0.131	0.230***	-0.127	-0.235***	0.075	0.158*	0.073	0.056	0.165*	0.045	-0.054	-0.507***
	(0.081)	(0.085)	(0.088)	(0.092)	(0.080)	(0.101)	(0.081)	(0.082)	(0.092)	(0.094)	(0.096)	(0.091)	(0.083)	(0.097)	(0.141)
Fresh fruit weekly budget	0.001	0.001	-0.004***	0.004***	0.000	0.000	0.000	0.000	0.002***	0.003***	0.002***	0.001*	0.001	0.004***	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Label domestic	-0.005	0.114	-0.066	-0.223***	0.065	0.025	0.073	-0.012	0.100	0.041	0.161*	0.057	-0.030	-0.307***	0.000
	(0.072)	(0.073)	(0.080)	(0.081)	(0.072)	(0.090)	(0.072)	(0.073)	(0.080)	(0.081)	(0.084)	(0.079)	(0.074)	(0.086)	(0.138)
Label organic	-0.176**	0.150*	-0.179*	0.121	0.075	-0.162	-0.045	-0.097	-0.148	0.228**	-0.304***	-0.021	-0.202**	0.463***	0.208
	(0.084)	(0.083)	(0.092)	(0.092)	(0.083)	(0.102)	(0.083)	(0.084)	(0.091)	(0.092)	(0.096)	(0.091)	(0.085)	(0.095)	(0.158)
Label non-GMO	-0.063	0.086	-0.165*	-0.032	0.237***	0.157	-0.198**	0.311***	0.146	0.010	0.120	-0.157*	0.113	0.217**	-0.096
	(0.081)	(0.081)	(0.091)	(0.090)	(0.081)	(0.100)	(0.082)	(0.082)	(0.089)	(0.091)	(0.093)	(0.089)	(0.083)	(0.094)	(0.153)
Label healthy	-0.116	-0.228***	-0.273***	0.091	0.233***	0.161*	0.058	0.149**	0.117	0.014	0.324***	-0.185**	0.117	-0.135	0.122
	(0.075)	(0.076)	(0.083)	(0.084)	(0.075)	(0.093)	(0.075)	(0.076)	(0.083)	(0.085)	(0.087)	(0.083)	(0.077)	(0.088)	(0.141)

[1] Single, double, and triple asterisks (*, **, ***) indicate statistical significance at the 10%, 5%, and 1% levels.
[2] Standard errors are in parentheses.

Table 2.10: Cross-utility effect estimates from multivariate logit model change in utility of purchasing.

	Blueberries	Blueberries	Blueberries	Blueberries	Blueberries
	Control N=801	Crunchy N=802	Fresh N=805	Sweety N=800	Pooled N=3,208
Apples	0.071	0.083*	0.235***	0.108**	0.107***
	(0.049)	(0.049)	(0.049)	(0.048)	(0.024)
Avocados	0.312***	0.423***	0.372***	0.367***	0.356***
	(0.049)	(0.048)	(0.049)	(0.049)	(0.024)
Bananas	0.293***	0.32***	0.254***	0.314***	0.286***
	(0.054)	(0.058)	(0.054)	(0.054)	(0.027)
Blackberries	0.570***	0.388***	0.566***	0.422***	0.483***
	(0.054)	(0.055)	(0.054)	(0.057)	(0.027)
Blueberries	0	0	0	0	0
Cantaloupe	0.203***	0.031	0.068	0.020	0.078***
	(0.062)	(0.059)	(0.061)	(0.060)	(0.030)
Grapes	0.459***	0.368***	0.391***	0.309***	0.363***
	(0.048)	(0.049)	(0.048)	(0.048)	(0.024)
Oranges	-0.079	-0.025	0.012	-0.106**	-0.039*
	(0.049)	(0.049)	(0.048)	(0.048)	(0.024)
Peaches	0.121**	0.104*	0.158***	0.212***	0.149***
	(0.054)	(0.054)	(0.054)	(0.054)	(0.027)
Pears	-0.059	0.168***	-0.101*	0.079	0.019
	(0.055)	(0.056)	(0.056)	(0.055)	(0.027)
Pineapple	-0.173***	-0.215***	0.254***	-0.105*	-0.056**
	(0.057)	(0.056)	(0.055)	(0.056)	(0.027)
Raspberries	0.298***	0.358***	0.448***	0.608***	0.426***
	(0.053)	(0.052)	(0.053)	(0.054)	(0.026)
Strawberries	0.935***	0.799***	0.937***	0.963***	0.904***
	(0.048)	(0.048)	(0.048)	(0.047)	(0.024)
Watermelon	0.166***	0.127**	0.142**	-0.020	0.095***
	(0.058)	(0.058)	(0.057)	(0.057)	(0.028)

[1] Single, double, and triple asterisks (*, **, ***) indicate statistical significance at the 10%, 5%, and 1% levels.
[2] Standard errors are in parentheses.

Table 2.11: Own and cross price elasticities of blueberries at mean demographics and prices implied by multivariate logit model.

Change in Price of	Quantity of				
	Blueberries Control N=801	Blueberries Crunchy N=802	Blueberries Fresh N=805	Blueberries Sweety N=800	Blueberries Pooled N=3,208
Apples	-0.037	-0.043	-0.050	-0.029	-0.04
Avocados	-0.046	-0.052	-0.039	-0.039	-0.042
Bananas	-0.023	-0.019	-0.026	-0.023	-0.022
Blackberries	-0.102	-0.100	-0.077	-0.097	-0.092
Blueberries	**-0.759**	**-0.768**	**-0.544**	**-0.736**	**-0.685**
Cantaloupe	-0.051	-0.056	-0.049	-0.039	-0.047
Grapes	-0.118	-0.117	-0.090	-0.085	-0.100
Oranges	-0.033	-0.040	-0.054	-0.018	-0.036
Peaches	-0.029	-0.029	-0.034	-0.026	-0.029
Pears	-0.017	-0.035	-0.023	-0.023	-0.023
Pineapple	-0.035	-0.031	-0.064	-0.033	-0.041
Raspberries	-0.077	-0.094	-0.087	-0.099	-0.089
Strawberries	-0.201	-0.227	-0.216	-0.224	-0.214
Watermelon	-0.056	-0.053	-0.067	-0.042	-0.052

Table 2.12: Own and cross elasticities of fruits – Sample of respondents presented the control treatment.

Quantity of	Change in Price of													
	Apples	Avocados	Bananas	Blackberries	Blueberries	Cantaloupe	Grapes	Oranges	Peaches	Pears	Pineapple	Raspberries	Strawberries	Watermelon
Apples	-0.392	-0.062	-0.066	-0.048	-0.073	-0.071	-0.153	-0.114	-0.039	-0.065	-0.078	-0.053	-0.100	-0.055
Avocados	-0.100	-0.447	-0.052	-0.129	-0.147	-0.082	-0.125	-0.068	-0.046	-0.059	-0.097	-0.100	-0.149	-0.083
Bananas	-0.065	-0.032	-0.152	-0.011	-0.046	-0.035	-0.067	-0.061	-0.015	-0.019	-0.035	-0.013	-0.060	-0.034
Blackberries	-0.055	-0.090	-0.012	-0.960	-0.229	-0.092	-0.115	-0.052	-0.064	-0.049	-0.086	-0.283	-0.169	-0.087
Blueberries	-0.037	-0.046	-0.023	-0.102	-0.759	-0.051	-0.118	-0.033	-0.029	-0.017	-0.035	-0.077	-0.201	-0.056
Cantaloupe	-0.124	-0.088	-0.062	-0.141	-0.177	-1.059	-0.221	-0.192	-0.121	-0.124	-0.216	-0.108	-0.221	-0.273
Grapes	-0.107	-0.054	-0.047	-0.071	-0.163	-0.088	-0.757	-0.108	-0.060	-0.044	-0.083	-0.064	-0.199	-0.090
Oranges	-0.138	-0.051	-0.075	-0.055	-0.079	-0.132	-0.186	-0.656	-0.088	-0.090	-0.116	-0.066	-0.165	-0.130
Peaches	-0.078	-0.057	-0.030	-0.114	-0.117	-0.140	-0.173	-0.147	-0.383	-0.122	-0.100	-0.098	-0.178	-0.127
Pears	-0.135	-0.075	-0.039	-0.090	-0.070	-0.147	-0.130	-0.154	-0.124	-0.603	-0.117	-0.088	-0.124	-0.113
Pineapple	-0.126	-0.097	-0.057	-0.123	-0.111	-0.200	-0.192	-0.156	-0.080	-0.092	-0.840	-0.147	-0.292	-0.249
Raspberries	-0.067	-0.078	-0.017	-0.315	-0.193	-0.078	-0.116	-0.070	-0.061	-0.054	-0.115	-0.785	-0.243	-0.079
Strawberries	-0.059	-0.054	-0.036	-0.087	-0.233	-0.074	-0.167	-0.080	-0.052	-0.035	-0.106	-0.113	-0.811	-0.113
Watermelon	-0.084	-0.077	-0.052	-0.116	-0.167	-0.236	-0.194	-0.164	-0.095	-0.083	-0.233	-0.094	-0.290	-0.921
No buy	0.422	0.259	0.424	0.371	0.834	0.241	0.604	0.349	0.208	0.203	0.261	0.333	0.717	0.278

CHAPTER THREE: THE IMPACT OF INTRODUCING A NEW APPLE VARIETY ON WASHINGTON STATE'S APPLE SHIPMENTS - EVIDENCE FROM A TIME SERIES ANALYSIS

Abstract

This study employs weekly shipment data from nine distinct apple varieties in Washington state. The first objective is to compare the predictive capabilities of various conventional time series models and machine learning techniques in predicting weekly apple shipments. While certain traditional time series models, specifically the Seasonal Autoregressive Integrated Moving Average (SARIMA), excel in terms of predictive accuracy and the ability to capture variability, this study recommends employing a machine learning model, specifically Facebook Prophet, due to its computational efficiency and strong predictive accuracy. Additionally, the research assesses the impact of introducing a new apple variety, Cosmic Crisp®, on the shipments of existing apple varieties through an interrupted time series (ITS) analysis. The introduction of Cosmic Crisp® is associated differently among various apple cultivars, with some exhibiting no discernible changes, while others undergo a decrease in subsequent shipment levels. Significantly, the overall apple shipments in Washington State show no association with the introduction of Cosmic Crisp®. Additionally, we observed that the shipments of Cosmic Crisp® are associated with an increase in the supply of specific apple varieties and the overall apple supply.

Keywords: Time Series, SARIMAX, Forecasting, Apple Shipments, Cosmic Crisp®

3.1 Introduction

The apple, derived from the apple tree (Malus domestica), is a highly popular fruit in U.S. and Worldwide. Consuming apples is advised due to the multiple proven health benefits, it is a source of fiber, minerals, vitamins, antioxidants, and other essential micronutrients. The appeal of apples to consumers can be attributed to factors such as their taste, associated health benefits, extended shelf life, and convenience in terms of portability (McCluskey et al., 2013; Gallardo et al., 2015). In addition, apples are rich in sugar making them a healthier alternative to sugary snacks and desserts (Bondonno et al., 2017; Hyson, 2011; Boyer et al., 2004; Larsson et al., 2013; Wang et al., 2014). Yet the per capita consumption of apples in the United States remains stagnant for the last two decades. In the U.S., the annual per capita consumption of fresh apples was 15.75 lbs in 2021, 16.3 lbs in 2020, with trivial variation compared to the 16.2 lbs reported in 2001 (USDA-ERS, 2023a). This relatively modest level of fresh apple consumption within the U.S. is evidence by the fact that it falls below the global average, which was 18.51 lbs in 2020. When looking at the per capita consumption of fresh apples in other countries, Turkey takes the lead with 78.71 lbs. Other countries with higher per capita consumption than the U.S. include China with 45.42 lbs, Germany with 39.9 lbs, the U.K. with 37.26 lbs, and Canada with 21.89 lbs (Statistics Canada, 2023; FAOSTAT, 2023). Hence, there exists the potential for U.S. consumers to augment their per capita fresh apple consumption, bringing it more in line with consumption patterns observed in other comparable countries.

The need to increase U.S. per capita consumption is further exacerbated by the U.S. fresh apple industry dependence on export markets. Approximately 23 percent of the U.S. fresh apple production being exports in 2021/22 marketing year (USDA-ERS, 2023b; U.S. Apple

Association, 2022). In 2021/22, the main U.S. fruit export markets destination included Canada, Mexico, South Korea, and Japan, with apples being among the leading commodities. While the volume of apple exports has increased over the past decade, the export value has decreased (USDA-FAS, 2023). This emphasizes the need of increasing domestic fresh apple consumption to balance the domestic excess apple supply and reduce reliance on export markets.

One strategy to increase the per capita consumption of fresh apples is to align the supply of apples with the desired sensory qualities expected by U.S. consumers. This can be achieved by developing and promoting new and improved apple varieties that cater to consumer preferences and meet their expectations. It is evident that newly improved apple varieties have garnered more consumer interest compared to existing ones (Yue and Tong 2011; Wang and Çakir. 2020), exemplified by the success of varieties like SweeTango (introduced in 2006), Honeycrisp (introduced in 1991), and Zestar (introduced in 1999).

Cosmic Crisp®, known as 'WA 38,' is a newly introduced apple cultivar originating from the Washington State University (WSU) Tree Fruit Research and Extension Center in Wenatchee, Washington. It stands out for its exceptional crispness, juiciness, and storability. This apple variety also boasts a balanced sweetness and acidity profile, which aligns well with attributes that consumers are willing to pay a premium for (Gallardo et al., 2018). This apple variety became commercially available for purchase starting from December 1, 2019 (Washington State University Fundraising News, 2019) with expectations that it would stimulate per capita apple consumption. However, this expected surge in demand has not materialized as initially projected (Gallardo et al., 2022). In the 2021/22 marketing year, the total shipments of Cosmic Crisp® amounted to 3.61 million 40-pound boxes (U.S. Apple Association, 2022). Growers in

Washington State, on the other hand, have planted 17 million trees since 2017, with an estimated harvest of more than 4.5 million 40-pound boxes of this apple variety in 2022, which exceeded the harvest of 2021 by 1 million boxes (Washington State University Insider, 2021; Washington State University CAHNRS News, 2023).

It is important to consider that the supply quantities of apples have a significant effect on their prices. As per the law of supply and demand, an increase in the supply of a particular apple variety can lead to a decrease in its price, all else constant (Gallardo et al., 2022). In fact, data from the Washington State Tree Fruit Association (Washington State Tree Fruit Association, 2023) reveals that as the quantity of Cosmic Crisp® shipments increased, the price per box decreased. Particularly, in the 2021/22 marketing year, there were a total of 3,609 shipments, totaling 3.61 million 40-pound boxes, with an average price of $36.40 per box; in 2020/21, there were 1,485 shipments, totaling 1.49 million 40-pound boxes, with an average price of $51.46 per box; in 2019/20, there were 343 shipments, totaling 0.43 million 40-pound boxes, with an average price of $72.72 per box. This surplus supply of an apple variety that aligns with consumer expectations raises questions about its potential impact on the consumption of other apple varieties. Industry needs to be prepared for the potential consequences of this effect.

The investigation holds particular significance due to a significant gap in the current body of research, as there is limited exploration into the consequences of introducing a high volume of a new apple variety on the supply dynamics of existing apple varieties that have long been part of the market. This study aims to bridge this gap utilizing the shipment quantity and price data to examine these effects, in contrast to simulated prices, as utilized by Amin (2023).

Washington state has been the leading apple-producing state, in the U.S., since the early 1920s and continues to hold its position, contributing 6.14 billion apples to the total production in 2022, which accounts for about 63 percent of the entire U.S. apple production (USDA-NASS, 2023). The primary apple varieties cultivated in this area, as a percentage of the total crop, including Red Delicious (34%), Gala (19%), Fuji (13%), Granny Smith (12%), Golden Delicious (10%), Cripps Pink (3%), Honeycrisp (3%), Braeburn (3%), and the remaining 3% consists of other apple types (Washington State University Tree Fruit, 2023).

This research utilizes weekly apple shipment data from Washington to analyze the time series pattern of apple shipments in the region and examine the factors that affect them. The study evaluates a range of traditional time series models and innovative machine learning models, comparing their performance to identify the most accurate forecasting approach for weekly quantities supplied of Washington grown apples. Moreover, an event analysis is carried out to assess the effect of the introduction of Cosmic Crisp® on quantities supplied of other apple varieties. This analysis holds significant importance for both stakeholders in the apple industry in Washington and the wider community.

3.2 Literature review

The first part of this section will involve a review of the existing literature related to the apple market, followed by an exploration of the literature related to the application of time series models.

Fresh apple markets

Research has shown that consumers are willing to pay a premium for apples that possess their preferred quality attributes, both externally and internally (Shapiro 1982 and 1983; Daillant-Spinnler et al. 1996; McCluskey et al. 2007). External attributes, including the name of apple varieties, play a significant role in influencing consumer preferences. Unlike many other product categories, apple varieties function much like brands, as they are marketed by their cultivar names, and the brand name can impact consumer perceptions and valuation of the fruit (Rickard et al., 2013, Gallardo et al., 2018; Richards and Patterson, 2000). Different from other fruits, new apple varieties are a key component of marketing strategy, as its price change and promotion can potentially impact other varieties' demand (Richards and Patterson, 2018). Regarding internal attributes, consumers highly value characteristics such as firmness, sweetness, low acidity, crispness, juiciness, and aroma in apples (Yue et al., 2017; Manalo 1990; Cliff, Stanich and Hampson 2014). Apples that excel in these sensory attributes are more likely to capture consumer attention and loyalty.

It's important to note that the impact of introducing new apple varieties on the market may take various forms. The introduction of a new variety can result in scenarios where there is no significant impact on existing varieties, an increase in supply and prices of other apple varieties, a decrease in supply and prices of other apple varieties. There are no studies analyzing the effect on the supply of apples after the introduction of new varieties. Existing studies focus on the demand side, for example Amin et al. (2021), found a market expansion effect rather than a market stealing effect when introducing a new club apple. That is, introducing club apple varieties had a demand shifter effect. The contribution of this study is to determine what effects the increase in the quantities supplied of Cosmic Crisp® would have on the quantities supplied and prices of other apple varieties supplied by Washington state.

Time series forecasting

Time series forecasting methods hold a distinct advantage when confronted with scenarios where there exists limited comprehension regarding the effect of explanatory variables on the output. The objective of time series analysis is to understand how change in time influences the dependent variables and thereby enabling the prediction of values for future time periods.

A substantial body of literature leverages time series data for forecasting, and time series analysis finds application across various domains. For instance, Gaur (2020) employed daily cumulative case data to predict confirmed COVID-19 cases. Cuaresma et al. (2004) used hourly LPX electricity spot-prices data for forecasting electricity spot-prices. Catalão et al. (2007) conducted a study using daily average electricity market prices in Spain. Capps (2022) utilized weekly shipments data to analyze avocado trends. Roznik et al. (2023) harnessed national and state-level corn yield data. Gupta et al. (2022) employed solar power generation data to predict future solar energy production. Michel and Makowski (2013) examined time series data on wheat yields.

Prominent time series regression models encompass Autoregressive (AR) models, Moving Average (MA) models, Non-seasonal Autoregressive Integrated Moving Average (ARIMA) models, Seasonal Autoregressive Integrated Moving Average (SARIMA) models, Exponential smoothing, among others (Hyndman & Athanasopoulos, 2018). For instance, Cuaresma et al. (2004) applied AR, MA, ARMA, and unobserved components models (UCM) to predict electricity spot-prices. Additionally, a burgeoning body of literature incorporates machine learning (ML) techniques into time series analysis. ML methods have garnered attention due to their ability to achieve heightened prediction accuracy compared to traditional empirical and physical models. This superiority arises from their proficiency in discerning intricate and latent patterns within data more efficiently (Noshi et al., 2018; Ning et al., 2022). Common ML applications encompass Neural networks, Facebook (FB) Prophet, Extreme Gradient Boosting (XGBoost), Long Short-Term Memory (LSTM), and other approaches. Catalão et al. (2007) utilized a neural network approach for electricity price forecasting, demonstrating its superiority over ARIMA models. Roznik et al. (2023) explored the XGBoost algorithm for corn yield forecasting, revealing that although it did not outperform the World Agricultural Supply and Demand Estimates (WASDE) forecast, it demonstrated the ability to generate reasonably accurate predictions of crop yields. Gupta et al. (2022) used solar power generation data for forecasting, employing FB Prophet and XGBoost models, with the latter demonstrating better performance. Siami-Namini et al. (2018) compared ARIMA and LSTM models using financial time series data and found LSTM to outperform ARIMA. However, To the best of the author's knowledge, there is a noticeable gap in the existing literature regarding the comparison of performance among different time series models using apple shipment data.

Interrupted-time-series (ITS) analysis is a widely adopted approach for assessing the consequences of an event using time series data. It is a commonly employed method, particularly in the evaluation of healthcare interventions (Bernal et al., 2017; Schaffer et al., 2021; Penfold and Zhang, 2013). Specifically, Bernal et al. (2017) utilized ITS to evaluate the impact of a smoking ban in public places on hospital admissions for acute coronary events. Schaffer et al. (2021) employed ITS to scrutinize the effects of a health policy change related to the refill of a new prescription of the lowest quetiapine tablet strength on inappropriate prescribing. Moreover, Xie et al. (2022) investigated immediate changes in preterm birth rates during the COVID-19 mitigation period. Bernal et al. (2013) also applied this methodology to investigate the effect of the late 2000s financial crisis on suicide rates in Spain. However, there is a notable gap in the literature when it comes to employing the ITS analysis to investigate the effect of introducing a new apple variety on the shipment patterns of pre-existing apple varieties.

3.3 Methodology

3.3.1 Forecast models

As summarized in the literature review, various time series analysis methodologies emerged in these two decades and have been widely applied in the fields including power usage, electricity price, agriculture product yields, etc. These methodologies have consistently demonstrated superior performance compared to traditional time series models (Catalão et al., 2007; Noshi et al., 2018; Ning et al., 2022; Siami-Namini et al., 2018; Roznik et al., 2023). However, there has been limited utilization of time series analysis in the context of apple shipments, highlighting a significant gap in the field. Therefore, this study employs a range of commonly used time series models to analyze apple shipment data and offers recommendations for selecting the most appropriate model for predicting apple shipments, supported by evidence from performance comparisons.

Non-seasonal Autoregressive Integrated Moving Average

One of the most widely used approaches for time series prediction is the Autoregressive Integrated Moving Average (ARIMA) model. The full non-seasonal ARIMA(p, q, d) model can be expressed as:

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \epsilon_t \tag{3.1}$$

where y_t is the time series, *p* is the order of the autoregressive (AR) part, *d* is the degree of first differencing involved, *q* is the order of the moving average (MA) part. To determine the optimal ARIMA model, a parameter search for the candidate model was conducted using the *auto_arima* function from the *pmdarima* package in Python 3.8. Following Panapongpakorn and Banjerdpongchai (2019), the ARIMA model with the lowest Akaike information criterion (AIC) was selected as the optimal model.

Seasonal Autoregressive Integrated Moving Average

Seasonal ARIMA (SARIMA) model is formed by including additional seasonal terms in the ARIMA models, it can be written as *ARIMA(p, d, q)(P, D,Q)S*, where *(p, d, q)* is the non-seasonal part and *(P, D,Q)S* is the seasonal part. Specifically, p is the non-seasonal AR order, d is the non-seasonal differencing, q represents the non-seasonal MA order, P is the seasonal AR order, D is the seasonal differencing, Q is the seasonal MA order, *S* is the number of observations per year.

The candidate parameters were identified using the *auto.arima* command in R 4.0.5. The algorithm examines various combinations of SARIMAX models and selects the one with the lowest AIC.

Exponential Smoothing

Another one of the most widely used time series forecasting approaches is Exponential smoothing model (Brown, 1959; Holt, 1957; Winters, 1960). Exponential smoothing methods rely on predicting future values by assigning varying weights to past observations, with these weights diminishing exponentially as the observations become older. There are three main types of exponential smoothing time series forecasting methods: simple exponential smoothing (SES), double exponential smoothing, and triple exponential smoothing. The SES is a time series forecasting method for univariate data without a clear trend or seasonality. The double exponential smoothing extends its capabilities to capture trends within the univariate time series. Triple exponential smoothing, also known as Holt Winter's Exponential Smoothing (HWES), further enhances its capabilities by accommodating seasonality within the univariate time series.

Single exponential smoothing

The SES model predicts the future time step by employing a linear function of past observations with exponential weighting. The SES requires a single parameter, called α, alternatively referred to as the smoothing factor or smoothing coefficient. This parameter governs the rate at which past observations influence the forecast and is typically set within the range of 0 to 1:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots \quad (3.2)$$

Large values indicate that the model pays attention mainly to the most recent past observations, whereas small values indicate that the model pays attention mainly to the most historical observations. When α=1, the forecasting is equivalent to the naïve method.

Holt Winter's Exponential Smoothing

The HWES model, also known as the Triple Exponential Smoothing method, forecasts the upcoming time step by utilizing a linear function of past observations, weighted exponentially. It incorporates considerations for trends and seasonality in its forecasting process. The HWES model comprises three smoothing equations, level l_t, trend b_t, and seasonal s_t. There are two variations to HWES models, the additive method and the multiplicative method. The choice between these methods depends on the nature of seasonal variations within the series. The additive method is favored when seasonal fluctuations remain relatively constant throughout the series, whereas the multiplicative method is more suitable when seasonal variations change in proportion to the level of the series. To assess model performance, both methods will be applied and compared. The Holt-Winters' additive method can be written as:

$$\hat{y}_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)}$$

$$l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1})$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1}$$

$$s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \tag{3.3}$$

where α, β^*, and γ correspond to the smoothing parameters for level, trend, and seasonality, respectively. *m* stands for the count of observations in a year. *k* is the integer part of $(h-1)/m$, guaranteeing that the seasonal index estimations employed for forecasting are based on the most recent year in the dataset (Hyndman & Athanasopoulos, 2018). The Holt-Winters' multiplicative method can be written as:

$$\hat{y}_{t+h|t} = (l_t + hb_t)s_{t+h-m(k+1)}$$

$$l_t = \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(l_{t-1} + b_{t-1})$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1}$$

$$s_t = \gamma \frac{y_t}{(l_{t-1}+b_{t-1})} + (1-\gamma)s_{t-m} \tag{3.4}$$

Furthermore, it's worth noting that both the additive and multiplicative Holt-Winters' methods allow for the application of damping. The HWES model, when incorporating a damped trend and additive seasonality, can be represented as follows:

$$\hat{y}_{t+h|t} = l_t + (\phi + \phi^2 + \cdots + \phi^h)b_t + s_{t+h-m(k+1)} \tag{3.5}$$

The HWES with a damped trend and multiplicative seasonality can be written as

$$\hat{y}_{t+h|t} = [l_t + (\phi + \phi^2 + \cdots + \phi^h)b_t]s_{t+h-m(k+1)} \tag{3.6}$$

XGBoost Model

XGBoost is short for Extreme Gradient Boosting and has become a popular tool for many applications for its demonstrated ability for prediction of classification and regression problems (Gupta et al., 2022). The method stems from the innovative gradient boosted decision tree-based searching method (Chen et al., 2015). The objective function on the wth iteration is:

$$Z^{(w)} = \sum_{i=1}^{n} z\left(f_i, \hat{f}_i(w-1) + g_w(r_i)\right) + \partial(g_w) \tag{3.7}$$

where z is the loss function, g_w is the wth tree output and ∂ is the regularization. A comprehensive explanation of XGBoost can be located in the resource provided by Machine Learning Mastery (Machine Learning Mastery, 2020).

It is worth noting that, to use XGBoost for time series forecasting, the time series dataset should be transformed into supervised learning problem first. The idea of transforming a time series dataset into a supervised learning problem is to use previous time steps as input variables and use the next time step as the output variable, that is, use the previous time step value to predict the next time step value, called sliding window. The XGBoost models were implemented by following the code examples available from Machine Learning Mastery (Machine Learning Mastery, 2020), and the analysis was conducted using Python 3.8. We conducted predictions by varying the time lag from 1 to 10 and ultimately selected the model that yielded the lowest Mean Absolute Error (MAE).

Facebook Prophet Model

The FB Prophet Model, often referred to as Prophet, is a forecasting procedure developed by Facebook (Facebook, 2019), available in Python and R. This model excels in handling time series data with strong seasonality and has the capability to account for holidays. Additionally, it performs admirably when confronted with time series data exhibiting missing values, variation in trends, and the identification of outliers (Jha & Pande, 2021; Gupta et al., 2022). The time series model is decomposed with three components: trend, seasonality, and the impact of holidays (Taylor and Letham, 2018). FB Prophet employs an additive regression model for forecasting time series data, with the equation formulated as follows:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \tag{3.8}$$

where $g(t)$ represents the trend factor, $s(t)$ represents the seasonality factor, $h(t)$ accounts for the effects of holidays, ε_t is the error factor. In this study, multiple FB Prophet models were executed using Python version 3.8, and the one demonstrating the highest predictive accuracy was selected and reported.

3.3.2 Predictive accuracy

As the forecasts are generated for a period with available actual historical data, the model's prediction accuracy is evaluated by comparing it to the real data using root mean squared error (RMSE) and mean absolute error (MAE):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.9}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.10}$$

where y_i refers to the actually realized shipments in period *t*, $\hat{y}_i$ represents the forecasted shipment for period *t*, *n* is the number of observations the dataset. The rationale for employing RMSE and MAE lies in their property of retaining the same units as the forecasted values (Roznik et al., 2023). These metrics are commonly used to assess forecast accuracy (Hodson, 2022; Clements and Hendry, 2002; Theil et al., 1966).

Furthermore, to assess and compare the average predictive accuracy of each model, we employ the mean absolute percentage error (MAPE). Despite its limitations, forecast accuracy is devoid of units of measurement (Lin et al., 2012; Makridakis, 1993). This enables us to compare forecast accuracy across various datasets or forecasting approaches without the influence of measurement units.

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{3.11}$$

3.3.3 Intervention analysis

To evaluate the effect of the introduction of Cosmic Crisp®, the interrupted time series (ITS) analysis, also called intervention analysis was employed (Bernal et al., 2017). As pointed out by Schaffer et al. (2021), a commonly used model to address the effect of an intervention is segmented linear regression, the simplest form of ITS analysis. However, a key assumption of linear regression is that the errors are independent and not correlated, which is often violated with time series data. Another alternative model is the Seasonal Autoregressive Integrated Moving-Average with Exogenous Regressors (SARIMAX) model which generalized the ARIMA model to incorporate both seasonality patterns and exogenous variables.

Utilizing SARIMAX modeling, as suggested by Schaffer et al. (2021), was chosen for its ability to assess the impacts of notable interventions while considering underlying trends, autocorrelation, and seasonal patterns before and after an intervention. To understand the effect of the introduction of Cosmic Crisp®, the data was divided into "pre-intervention" and "post-intervention" periods.

To take the intervention effect into consideration, following Schaffer et al. (2021), this study focused on three main types: step change, pulse, and ramp. Step represents a level shift, that is a sudden and sustained change where the time series is shifted either up or down by a given value immediately following the intervention, step equals 0 before the intervention and 1 after the intervention. Pulse represents a sudden and temporary change that is observed for one or more time points immediately after the intervention and then returns to baseline level. Pulse equals 1 of the date of the intervention and 0 otherwise. Ramp represents a slope change that happens immediately after the intervention. Ramp equals 0 prior to the intervention and increases by 1 after the intervention. The model assessed the impact of the intervention by predicting the value of Y_t in a counterfactual scenario, where the intervention did not occur and identify the deviation between the actual shipments and the predicted values.

In this study, we assumed that a level change (*step*) and a change in slope (*ramp*) would occur following the introduction of Cosmic Crisp®. As suggested by Schaffer et al. (2021), it is common for both a step change and a slope change to exist. A sudden and temporary change (*pulse*) is less likely, given that growers typically do not remove newly planted apple trees shortly after planting. Therefore, the final model will be a SARIMA model incorporating two exogenous regressors. The first predictor, labeled *step*, takes a value of 0 before the introduction of Cosmic Crisp® and 1 following its introduction. The second predictor, *ramp*, signifies the time elapsed since the intervention. Specifically, *ramp* takes a value of 1 on the day of the intervention and increases by 1 each subsequent day post-intervention, remaining at 0 before the intervention occurs.

The SARIMAX models were performed in R-4.0.5, the command *auto.arima* was used to identify candidate p, q, d, P, Q, and D parameters. The algorithm examines various combinations of SARIMAX models and selects the one with the lowest AIC.

3.4 Data

The weekly state apple shipment data for different apple varieties in Washington State is collected from the Washington State Tree Fruit Association (WSTFA) for the period from 2008 to February 2023. The data consisted of weekly shipment quantities, with each shipment representing a load of 1,000 boxes, and weekly Freight on Board (FOB) prices (measured in dollars per 40 lbs box) of nine different apple varieties shipped from Washington State. The study focuses on eight primary apple types grown in the region, namely Red Delicious, Gala, Fuji, Granny Smith, Golden Delicious, Cripps Pink, Honeycrisp, and Braeburn, along with the recently introduced variety known as Cosmic Crisp®. The harvest window for these apple varieties ranges from mid-August to end-October. Gala, Honeycrisp, and Golden Delicious being harvested early, while Red Delicious, Granny Smith, Braeburn, and Cosmic Crisp® are harvested in the mid-season, and Cripps Pink and Fuji are harvested later in the season (Washington State University Tree Fruit, 2023). To address missing values in the dataset, a linear interpolation approach was utilized. It's worth noting that linear interpolation is a common technique for managing missing data, as highlighted in prior studies (Noor et al., 2014; Picornell et al., 2021). The data was read and plotted as a time series using Python 3.8.

Table 3.1 presents the summary statistics of each apple variety. Every apple variety in the dataset has a distinct starting date, while they all share the same end date of February 27, 2023. Notably, the shipment data for the recently introduced Cosmic Crisp® variety, which was first planted in spring 2017, was not available until 2019 when it became available for sale, reflecting the fact that an apple tree starts production in Year 3 after being planted and achieves full production by year 5 (Washington State University Fundraising News, 2019).

Shipments

On average, the weekly shipments ranging from 35.38 (Braeburn) to 585.32 (Red Delicious). This data spans from August 25, 2008, which implies Gala's start date, to February 27, 2023. The final column indicates the total shipments encompassing the entire apple supply, including Cosmic Crisp®, within Washington State. The highest recorded total shipment is 3539, the lowest is 51, with an average of 2108.81 and a standard deviation of 477.94. Figure 3.1

presents the historical weekly apple shipments over the period August 25, 2008, to February 27, 2023.

FOB prices

The average FOB prices ranged between 17.80 (Red Delicious) and 60.63 (Honeycrisp), where the maximum FOB price reached to 104.09 yield by Honeycrisp and the lowest FOB price was 11.81 observed at Red Delicious variety.

3.5 Empirical analysis

Stationarity

Most statistical forecasting techniques rely on the assumption that the time series is approximately stationary. A stationary time series is defined by consistent statistical properties, such as a constant mean, variance, and autocorrelation over time. In this study, a stationary time series indicates that both the mean and variance of shipment quantity remain constant throughout the study period. In simpler terms, the observations can be seen as unrelated to time. To determine the stationarity of the time series, the Augmented Dickey-Fuller (ADF) test was employed. The outcome of the ADF test for each apple variety are presented in Table 3.2. It can be observed that except for Honeycrisp and Cosmic Crisp®, all the remaining varieties demonstrate stationarity, which indicates a non-negligible trend of the mean shipment amount can be observed for Honeycrisp and Cosmic Crisp®.

3.5.1 Predictive accuracy

The historical data used for model estimation covers the period from 2008 to the week ending on February 27, 2023. The time series data for each variety was divided into two parts: a training dataset and a testing dataset. Specifically, for each dataset, the testing set comprised the last 52 observations, which corresponded to the weekly shipments of the most recent year. Based on the model specifications, the weekly forecasts are generated for the period from the ending of March 7, 2022, to the week ending on February 27, 2023, encompassing 52 weeks, which represents a full year. The remaining data was designated as the training set. RMSE, MAE, and MAPE were calculated to assess and contrast the predictive precision of each model. Enhanced predictive precision is indicated by a lower RMSE, MAE, or MAPE value.

Table 3.3 displays the RMSE, MAE, and MAPE of each model for different apple varieties. Note that the abnormally high MAPE values in some cases, particularly for Golden Delicious and Braeburn, can be attributed to a limitation inherent in MAPE calculations when the actual values are small (Makridakis, 1993; Kim and Kim, 2016). As Figure 3.1 illustrates, there were periods for Golden Delicious and Braeburn when the actual shipments were notably smaller, as evidenced by a sharp decline in the actual shipments (the blue line), in contrast to other time periods.

The SARIMA model demonstrated the highest predictive accuracy in the case of Gala, Cripps Pink, Red Delicious, and Cosmic Crips. XGBoost excelled in the case of Fuji, and Prophet surpassed other models for Golden Delicious. Conversely, in some cases, the simpler models outperformed more complex ones. Specifically, the SES model consistently demonstrated better predictive accuracy in the case of Honeycrisp, Granny Smith, and Braeburn. This aligns with previous studies, which have shown that sophisticated models may not always outshine simpler ones (Fallahtafti et al., 2022; Rasmussen, 2004; Weron and Misiorek, 2008). It's worth noting, however, that while simpler models occasionally achieved good predictive accuracy, their forecasts resembled flat lines (Filios et al., 2020; Siregar et al., 2017). This limitation is corroborated by Figure 3.2, which illustrates their challenges in effectively capturing variations within the time series.

By utilizing MAPE, the forecast accuracy is devoid of units of measurement (Lin et al., 2012; Makridakis, 1993). This allows for the comparison of forecast accuracy between different datasets or forecasting methods, free from the impact of specific measurement units. To compare the average predictive precision of each model, we calculated the average MAPE for each model. The average MAPE were 138.580%, 30.607%, 42.503%, 40.743%, 54.911%, and 40.825%, for ARIMA, SARIMA, SES, HWES, XGBOOST, PROPHET, respectively (Table 3.3). Overall, the SARIMA model outperformed others in terms of the average MAPE, followed by HWES and Prophet.

It is important to highlight that although the SARIMA model (mean MAPE 30.607%) showed strong predictive accuracy, it involved significant operating time when using the *auto.arima* command to identify candidate parameters. As a result, the Prophet (mean MAPE 40.825%) stands out as a practical alternative, as it exhibits reasonable computational efficiency. Consequently, when assessing the overall performance, the Prophet models emerge as an attractive choice, offering enhanced predictive accuracy, effective capture of variations, and computational efficiency. In summary, machine learning models can also be efficiently applied to model time series data.

3.5.2 Intervention analysis

This section focuses on addressing the association of exogenous variables, particularly in the supply of Cosmic Crisp®. Therefore, although our findings indicate that apple supply is affected by its own past values, a characteristic inherent to time series data, we excluded the coefficients related to these variables, specifically the AR and MA terms, from the results table as they are not the main focus of this study.

The association of the shipments of other varieties the Cosmic Crisp® introduction

We first conducted the ITS analysis to investigate how the shipments of existing apple varieties associates with the introduction of Cosmic Crisp®. In this analysis, we employed shipment data for each variable to conduct the SARIMAX model, which incorporates exogenous regressors including *step* (representing immediate shifts in shipment) and *ramp* (reflecting continuous weekly changes following the event). Our rationale for including both *step* and *ramp* variables was affected by the expectation of an immediate shift in shipments coupled with gradual changes in shipments associated with the introduction of Cosmic Crisp®. This decision was guided by the fact that it takes apples approximately five years to reach full production capacity (Washington State University Fundraising News, 2019). This model acknowledged the possibility that the production of other varieties might still be evolving due to the introduction of Cosmic Crisp®.

Table 3.4 presents the outcomes. Figure 3.3 illustrates the shipments predicted by the SARIMAX model in the absence of intervention (referred to as the counterfactual scenario) in contrast to the observed values. In this figure, the blue line represents the actual shipment data, and the red line represents the predicted values by the SARIMAX model under the counterfactual scenario. Across all apple varieties, there was no immediate shift in supply levels associated with the introduction of Cosmic Crisp® on December 2, 2019. For Gala and Red Delicious, there was a gradual decline in weekly shipments (1.643 shipments, and 1.323

shipments, respectively) in comparison to the period before the introduction of Cosmic Crisp®, whereas the shipment levels of other apple varieties remained unchanged. The diverse results could be attributed to the distinctive characteristics of each apple variety. The decline in shipments of Red Delicious might be due to growers reallocating fewer acres to this variety following the introduction of Cosmic Crisp®, which had a negative impact on the supply of Red Delicious. Interestingly, there was no evidence of the association between the introduction of Cosmic Crisp® and the overall supply of apple, which encompasses the combined supply of all apple varieties. This suggests that there was no observed change in the total supply before or after the introduction of Cosmic Crisp®.

The association of the shipments of other varieties with the Cosmic Crisp® shipments

As indicated in Table 3.4, the introduction of Cosmic Crisp® on December 2, 2019, was not associated with an instant shift in the supply levels of other apple varieties. However, a significant gradual discouragement of the supplies was observed for Gala and Red Delicious. It is important to highlight that the analysis of Cosmic Crisp®'s introduction primarily focused on how the shipments of other apple varieties were associated with its presence. Nonetheless, the connections between the quantity of Cosmic Crisp supplies and the shipments of other varieties remained unexplored. Consequently, we conducted another SARIMAX model, this time introducing Cosmic Crisp® shipments as an exogenous regressor, to delve into these relationships. It's important to note that in this analysis, the Cosmic Crisp® supply is represented by the actual quantity of shipments, not just its presence in the market. Different from the earlier model, where the presence of Cosmic Crisp® was denoted as a binary variable with 1 indicating its presence and 0 otherwise, in this model, we used the actual shipment quantities of Cosmic

Crisp®. For instance, during the initial six weeks starting from December 2, 2019 (Washington State University Fundraising News, 2019), these quantities (observations) were 4, 84, 52, 18, 5, and 29 shipments, respectively.

The results are presented in Table 3.5. The shipment of Red Delicious, Golden Delicious, and Granny Smith is positively associated with the supply of Cosmic Crisp® (0.451, 0.172, and 0.350, respectively). No significant patterns were noted for the remaining apple varieties. Concerning the aggregate shipments of all apple varieties, an increase in Cosmic Crisp® shipments is associated with an increase in total apple shipments. To summarize, the influence of Cosmic Crisp®'s shipments is associated differently among various apple varieties: an increase in Cosmic Crisp® shipments is associated with an increase in the supply of specific varieties as well as the overall supply. This aligns with the conclusions drawn in prior research, suggesting that the introduction of a new apple variety may lead to a market expansion effect (Amin et al., 2021).

The association of the FOB prices of other varieties with the Cosmic Crisp® introduction

To gain a deeper understanding of how and whether the introduction of Cosmic Crisp® affected other varieties, we conducted an ITS analysis using the FOB prices data to assess the association of the presence of Cosmic Crisp® on the FOB prices of other apple varieties. Similar to the previous model, we employed the SARIMAX model using FOB price data with the inclusion of the *step* variable to signify immediate shifts in FOB prices. While FOB prices can be influenced by a multitude of factors, we did not anticipate any lasting and continuous impact resulting from the introduction of Cosmic Crisp®. Our expectation was that the FOB prices of other apple varieties would experience a sudden and immediate shift in response to the introduction of Cosmic Crisp®. Since the presence of Cosmic Crisp® itself would remain relatively stable, we did not foresee continuous changes in prices. Therefore, we opted to exclude the *ramp* variable from our analysis. Instead, considering the law of supply and demand, and the impact of shipment on FOB prices (Gallardo et al., 2022), we introduced shipment as an additional exogenous regressor. In summary, the expected impact of introducing Cosmic Crisp® can be summarized as twofold: it involves immediate shifts in shipments coupled with gradual changes in shipments, and it leads to immediate shifts in FOB prices.

The results are displayed in Table 3.6. Figure 3.4 illustrates the FOB prices anticipated by the SARIMAX model in the absence of intervention in contrast to the observed values The blue line corresponds to the actual FOB prices, while the red line signifies the SARIMAX model's predicted values in the absence of intervention. The results confirm that FOB prices tend to decrease as shipment quantities increase for all varieties. Notably, Braeburn appeared to be the variety most affected by its shipment, whereas Gala's FOB price exhibited the least impact by its own shipments. However, we did not observe any immediate shifts in FOB prices. This implies that the presence of Cosmic Crisp® is not significantly associated with altering the prices of other apples.

The results confirm that FOB prices tend to decrease as shipment quantities increase for all varieties. Notably, Braeburn appeared to be the variety most affected by its shipment, whereas Gala's FOB price exhibited the least impact by its own shipments. However, we did not observe any immediate shifts in FOB prices. This suggests that the presence of Cosmic Crisp® did not significantly alter the prices of other apples.

3.6 Discussion and conclusion

Washington state has maintained its preeminent position as the foremost apple-producing state since the early 1920s and persists in upholding this distinction. Nevertheless, it is incumbent to acknowledge that the apple production landscape in Washington has undergone a discernible diminishment in recent years.

Utilizing weekly shipments data from nine different apple varieties in Washington state, this study compares the predictive performance between multiple traditional time series models (e.g., ARIMA, SARIMA, SES, HWES) and machine learning methods (specifically, Prophet and XGBoost). The findings reveal a mixed predictive accuracy for different apple varieties. In general, SARIMA outperforms the other models, albeit with a significantly longer parameter optimization time. On the other hand, while XGBoost demonstrates superior performance in most cases, it exhibits lower predictive accuracy in certain instances. Prophet consistently demonstrates strong predictive accuracy across all cases. Therefore, the study recommends a comprehensive exploration of both traditional and ML models to determine the most suitable model for a given dataset.

The study employs ITS analysis to assess how the introduction of Cosmic Crisp®, a new apple variety, is associated with the shipments of existing apple varieties. Specifically, a SARIMAX model is employed. We first examined the association of Cosmic Crisp®'s introduction based solely on its presence, without considering the actual number of shipments. The association of this introduction diverges among different apple cultivars, as some show no noticeable changes, some demonstrated a decline in their following shipment levels. Importantly, the overall apple supply in Washington State is not associated with the introduction of Cosmic Crisp®. This divergence might be ascribed to the specific attributes of each apple variety, such as flavor and harvest timing. In addition, our results indicated there was no immediate shift in FOB prices following the introduction of Cosmic Crisp®.

Furthermore, we investigated the association between the shipments of other apple varieties and the shipments of Cosmic Crisp®. The results revealed that the association of Cosmic Crisp®'s shipments varies across different apple varieties: an increase in Cosmic Crisp® shipments is associated with an increase in the supply of specific varieties as well as the overall supply. It is important to note that the market response to the introduction to the new variety is not immediate. Moreover, the response of the industry to changes in consumers' demand is not immediate, as planting decisions are done for the long-term (15 years on average). This analysis offers perhaps a limited short-term impact of the introduction of Cosmic Crisp® on the supply of apples out of Washington state. Further, research involving longer time horizons is needed to have a comprehensive assessment of the effects of the new apple variety introduction. These analyses should consider the inclusion of additional exogenous variables, such as weather conditions, apple production in Washington state, and household income.

REFERENCES

Amin, M., Astill, G., Gallardo, R. K., McCluskey, J. J. (2023). Market-Winning Apple Attributes and the Projected Impacts of Releasing an Exclusive Variety. Working Paper. Unpublished.

Amin, M., Badruddoza, S., McCluskey, J. J., & Astill, G. M. (2021). Comparing Apples to Apples: Price Premiums of Club over Open Apple Varieties. Paper presented at the 2021 Agricultural and Applied Economic Association Annual Meetings. Austin, Texas, August 1-3.

Bernal, J. L., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: a tutorial. International journal of epidemiology, 46(1), 348-355.

Bondonno, N. P., Bondonno, C. P., Ward, N. C., Hodgson, J. M., & Croft, K. D. (2017). The cardiovascular health benefits of apples: Whole fruit vs. isolated compounds. Trends in Food Science & Technology, 69, 243-256.

Boyer, J., & Liu, R. H. (2004). Apple phytochemicals and their health benefits. Nutrition journal, 3, 1-15.

Brown, R. G. (1959). Statistical forecasting for inventory control. (No Title).

Capps, O. (2022). Forecasting Weekly Shipments of Hass Avocados from Mexico to the United States Using Econometric and Vector Autoregression Models. Econometrics-Recent Advances and Applications.

Catalão, J. P. D. S., Mariano, S. J. P. S., Mendes, V. M. F., & Ferreira, L. A. F. M. (2007). Short-term electricity prices forecasting in a competitive market: A neural network approach. electric power systems research, 77(10), 1297-1304.

Clements, M., & Hendry, D. (2002). An overview of economic forecasting. A Companion to Economic Forecasting. Oxford: Blackwell, 1-18.

Cliff, M. A., Stanich, K., & Hampson, C. (2014). Consumer research explores acceptability of a new Canadian apple–Salish™. Canadian Journal of Plant Science, 94(1), 99-108.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., ... & Zhou, T. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4), 1-4.

Cuaresma, J. C., Hlouskova, J., Kossmeier, S., & Obersteiner, M. (2004). Forecasting electricity spot-prices using linear univariate time-series models. Applied Energy, 77(1), 87-106.

Daillant-Spinnler, B., MacFie, H. J. H., Beyts, P. K., & Hedderley, D. (1996). Relationships between perceived sensory properties and major preference directions of 12 varieties of apples from the southern hemisphere. Food quality and preference, 7(2), 113-126.

Facebook. (2019). Automatic forecasting procedure.

Fallahtafti, A., Aghaaminiha, M., Akbarghanadian, S., & Weckman, G. R. (2022). Forecasting ATM cash demand before and during the COVID-19 pandemic using an extensive evaluation of statistical and machine learning models. SN computer science, 3(2), 164.

FAOSTAT. (2023). Food Balance Sheets database.

Filios, G., Katsidimas, I., Nikoletseas, S., Panagiotou, S., & Raptis, T. P. (2020, May). An agnostic data-driven approach to predict stoppages of industrial packing machine in near future. In 2020 16th International Conference on Distributed Computing in Sensor Systems (DCOSS) (pp. 236-243). IEEE.

Gallardo, R. K., Hanrahan, I., Yue, C., McCracken, V. A., Luby, J., McFerson, J. R., ... & Carrillo-Rodriguez, L. (2018). Combining sensory evaluations and experimental auctions to assess consumers' preferences for fresh fruit quality characteristics. Agribusiness, 34(2), 407-425.

Gallardo, R. K., Li, H., McCracken, V., Yue, C., Luby, J., & McFerson, J. R. (2015). Market intermediaries' willingness to pay for apple, peach, cherry, and strawberry quality attributes. Agribusiness, 31(2), 259-280.

Gallardo, R. K., McCluskey, J. J., Mittelhammer, R. C. & Winfree, J. A. (2022). Evaluation of promotion and advertising for Cosmic Crisp. School of Economic Sciences Report. Washington State University. Unpublished.

Gaur, S. (2020). Global forecasting of covid-19 using ARIMA based FB-Prophet. International Journal of Engineering Applied Sciences and Technology, 5(2), 463-467.

Gupta, R., Yadav, A. K., Jha, S. K., & Pathak, P. K. (2022, February). Time series forecasting of solar power generation using Facebook prophet and XG boost. In 2022 IEEE Delhi section conference (DELCON) (pp. 1-5). IEEE.

Jha, B. K., & Pande, S. (2021, April). Time series forecasting model for supermarket sales using FB-prophet. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 547-554). IEEE.

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. Geoscientific Model Development, 15(14), 5481-5487.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. International journal of forecasting, 20(1), 5-10.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.

Hyson, D. A. (2011). A comprehensive review of apples and apple components and their relationship to human health. Advances in nutrition, 2(5), 408-420.

Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. International Journal of Forecasting, 32(3), 669-679.

Larsson, S. C., Virtamo, J., & Wolk, A. (2013). Total and specific fruit and vegetable consumption and risk of stroke: a prospective study. Atherosclerosis, 227(1), 147-152.

Lin, M. C., Tserng, H. P., Ho, S. P., & Young, D. L. (2012). A novel dynamic progress forecasting approach for construction projects. Expert Systems with Applications, 39(3), 2247-2255.

Machine Learning Mastery. (2020). How to Use XGBoost for Time Series Forecasting.

Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. International journal of forecasting, 9(4), 527-529.

Manalo, A. B. (1990). Assessing the importance of apple attributes: an agricultural application of conjoint analysis. Northeastern journal of agricultural and resource economics, 19(2), 118-124.

McCluskey, J. J., Horn, B. P., Durham, C. A., Mittelhammer, R. C., & Hu, Y. (2013). Valuation of internal quality characteristics across apple cultivars. Agribusiness, 29(2), 228-241.

McCluskey, J. J., Mittelhammer, R. C., Marin, A. B., & Wright, K. S. (2007). Effect of quality characteristics on consumers' willingness to pay for Gala apples. Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie, 55(2), 217-231.

Michel, L., & Makowski, D. (2013). Comparison of statistical models for analyzing wheat yield time series. PLoS One, 8(10), e78615.

Ning, Y., Kazemi, H., & Tahmasebi, P. (2022). A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet. Computers & Geosciences, 164, 105126.

Noor, M. N., Yahaya, A. S., Ramli, N. A., & Al Bakri, A. M. (2014). Filling missing data using interpolation methods: Study on the effect of fitting distribution. Key Engineering Materials, 594, 889-895.

Noshi, C. I., Assem, A. I., & Schubert, J. J. (2018, December). The role of big data analytics in exploration and production: A review of benefits and applications. In SPE International Heavy Oil Conference and Exhibition (p. D012S021R001). SPE.

Panapongpakorn, T., & Banjerdpongchai, D. (2019, January). Short-term load forecast for energy management systems using time series analysis and neural network method with average true range. In 2019 First International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP) (pp. 86-89). IEEE.

Penfold, R. B., & Zhang, F. (2013). Use of interrupted time series analysis in evaluating health care quality improvements. Academic pediatrics, 13(6), S38-S44.

Picornell, A., Oteros, J., Ruiz-Mata, R., Recio, M., Trigo, M. M., Martínez-Bracero, M., ... & Rojo, J. (2021). Methods for interpolating missing data in aerobiological databases. Environmental Research, 200, 111391.

Rasmussen, R. (2004). On time series data and optimal parameters. Omega, 32(2), 111-120.

Richards, T. J., & Patterson, P. M. (2000). New varieties and the returns to commodity promotion: the case of Fuji apples. Agricultural and Resource Economics Review, 29(1), 10-23.

Rickard, B. J., Schmit, T. M., Gómez, M. I., & Lu, H. (2013). Developing brands for patented fruit varieties: Does the name matter?. Agribusiness, 29(3), 259-272.

Roznik, M., Mishra, A. K., & Boyd, M. S. (2023). Using a machine learning approach and big data to augment WASDE forecasts: Empirical evidence from US corn yield. Journal of Forecasting.

Schaffer, A. L., Dobbins, T. A., & Pearson, S. A. (2021). Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions. BMC medical research methodology, 21(1), 1-12.

Shapiro, C. (1983). Premiums for high quality products as returns to reputations. The quarterly journal of economics, 98(4), 659-679.

Statistics Canada. (2023). Table 32-10-0054-01 Food available in Canada.

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2018, December). A comparison of ARIMA and LSTM in forecasting time series. In 2018 17th IEEE international conference on machine learning and applications (ICMLA) (pp. 1394-1401). IEEE.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. The American Statistician, 72(1), 37-45.

Theil, H., Beerens, G. A. C., Tilanus, C. G., & De Leeuw, C. B. (1966). Applied economic forecasting (Vol. 4). Amsterdam: North-Holland.

U.S. Apple Association. (2022). Industry Outlook 2022. Retrieved from https://usapple.org/wp-content/uploads/2022/08/USAPPLE-INDUSTRYOUTLOOK-2022.pdf. Accessed October 17, 2023.

USDA-ERS. (2023a). Food Availability (Per Capita) Data System. Retrieved from https://www.ers.usda.gov/data-products/food-availability-per-capita-data-system/. Accessed October 17, 2023.

USDA-ERS. (2023b). Advancements in Apple Picking: An Industry Addresses Tight Farm Labor Markets. Retrieved from https://www.ers.usda.gov/amber-waves/2023/june/advancements-in-apple-picking-an-industry-addresses-tight-farm-labor-markets/#:~:text=Selective%20harvest%20machines%20are%20in,it%20with%20a%20robot%20arm. Accessed October 17, 2023.

USDA-FAS. (2023). Spotlight: Top U.S. Fruit Export Markets. Retrieved from https://www.fas.usda.gov/data/spotlight-top-us-fruit-export-markets#:~:text=In%202021%2F22%2C%20top%20U.S.,strawberries%2C%20oranges%2C%20and%20cherries. Accessed October 17, 2023.

USDA-NASS. (2023). Noncitrus Fruits and Nuts 2022 Summary. Retrieved from https://downloads.usda.library.cornell.edu/usda-esmis/files/zs25x846c/zk51wx21m/k356bk214/ncit0523.pdf. Accessed October 17, 2023.

Wang, Y., & Çakır, M. (2020). Welfare impacts of new demand-enhancing agricultural products: The case of Honeycrisp apples. Agricultural economics, 51(3), 445-457.

Wang, X., Ouyang, Y., Liu, J., Zhu, M., Zhao, G., Bao, W., & Hu, F. B. (2014). Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies. Bmj, 349.

Washington State University CAHNRS News. (2023). WSU's Cosmic Crisp® joins top 10 bestselling U.S. apple varieties. Retrieved from https://news.cahnrs.wsu.edu/article/wsus-cosmic-crisp-joins-top-10-bestselling-u-s-apple-varieties/. Accessed October 18, 2023.

Washington State University Fundraising News. (2019). Cosmic Crisp®: How a Bunch Made One Good Apple. Retrieved from https://foundation.wsu.edu/2019/12/01/cosmic-crisp/. Accessed October 17, 2023.

Washington State University Tree Fruit. (2023). Varieties – Apple. Retrieved from https://treefruit.wsu.edu/web-article/apple-varieties/. Accessed October 17, 2023.

Weron, R., & Misiorek, A. (2008). Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. International journal of forecasting, 24(4), 744-763.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. Management science, 6(3), 324-342.

Xie, Y., Mu, Y., Chen, P., Liu, Z., Wang, Y., Li, Q., ... & Zhu, J. (2022). Interrupted-time-series analysis of the immediate impact of COVID-19 mitigation measures on preterm birth in China. Nature communications, 13(1), 5190.

Yue, C., & Tong, C. (2011). Consumer preferences and willingness to pay for existing and new apple varieties: Evidence from apple tasting choice experiments. HortTechnology, 21(3), 376-383.

Yue, C., Zhao, S., Gallardo, K., McCracken, V., Luby, J., & McFerson, J. (2017). US growers' willingness to pay for improvement in rosaceous fruit traits. Agricultural and Resource Economics Review, 46(1), 103-122.

Table 3.1: Descriptive statistics of the weekly shipments of apples from Washington.

Descriptive statistic	Gala	Honeycrisp	Cripps Pink	Red Delicious	Fuji	Golden Delicious	Granny Smith	Braeburn	Cosmic Crisp	All varieties[1]
Start date	2008-08-25	2008-09-01	2008-10-27	2008-09-15	2008-09-01	2008-09-01	2008-09-01	2008-09-22	2019-12-02	2008-08-25
End date	2023-02-27	2023-02-27	2023-02-27	2023-02-27	2023-02-27	2023-02-27	2023-02-27	2023-02-27	2023-02-27	2023-02-27
Count	758	757	749	755	756	756	756	754	170	758
					Shipments					
Mean	494.00	159.72	87.25	585.32	300.93	155.86	284.58	35.38	48.94	2108.81
Median	519.00	157.00	87.00	601.00	311.00	143.50	293.50	21.59	44.50	2160.00
Maximum	952.00	531.00	241.00	1179.00	624.00	352.00	481.00	153.00	139.00	3539.00
Minimum	5.00	0.00	1.00	0.00	8.00	1.00	7.00	0.00	1.00	51.00
Std. Dev	161.09	110.62	48.89	195.54	109.82	73.92	69.26	33.96	33.50	477.94
					FOB prices					
Mean	25.34	60.63	30.23	17.80	25.73	22.25	23.93	21.33	50.42	-
Median	24.39	58.56	30.46	17.87	25.17	21.98	23.17	20.54	42.08	-
Maximum	45.28	104.09	47.40	38.28	47.67	37.44	48.45	40.07	75.62	-
Minimum	18.77	33.03	14.28	11.81	16.56	12.54	12.78	11.91	31.03	-
Std. Dev	4.33	12.92	4.79	2.97	4.75	4.77	5.28	4.97	15.77	-

[1] The total shipments encompassing the entire apple, including Cosmic Crisp®, supply within Washington State

Table 3.2: Augmented Dickey Fuller results.

	Gala	Honeycrisp	Cripps Pink	Red Delicious	Fuji	Golden Delicious	Granny Smith	Braeburn	Cosmic Crisp	All varieties
					ADF test					
ADF[1]-statistics	-5.530	-2.740	-6.985	-4.425	-8.286	-3.666	-6.576	-4.453	-1.437	-6.570
p-value	<0.001	0.067	<0.001	<0.001	<0.001	0.005	<0.001	<0.001	0.564	0.010
conclusion	stationary	non-stationary	stationary	stationary	stationary	stationary	stationary	stationary	non-stationary	stationary

[1] ADF is Augmented Dickey Fuller.

Table 3.3: Time series model predictive accuracy comparison.

Variety	Performance Metrics	Model					
		ARIMA	SARIMA	SES	HWES	XGBOOST	PROPHET
Gala	RMSE[1]	102.446	40.697	73.753	169.588	208.890	83.781
	MAE[2]	77.720	35.080	64.850	129.033	194.303	66.573
	MAPE[3]	21.911	8.354	15.908	34.029	45.397	17.588
Honeycrisp	RMSE	84.405	45.762	37.036	81.574	47.581	75.093
	MAE	76.280	40.297	30.211	67.599	35.554	60.093
	MAPE	27.602	15.385	12.071	23.683	13.940	23.174
Cripps Pink	RMSE	49.686	37.346	52.018	48.236	50.079	41.832
	MAE	43.926	32.006	41.099	41.973	39.018	37.162
	MAPE	33.462	24.816	46.980	32.765	44.914	29.578
Red Delicious	RMSE	263.698	61.682	81.664	149.794	92.268	112.398
	MAE	235.795	45.878	67.747	122.528	71.301	88.316
	MAPE	88.712	15.748	26.309	42.572	28.707	29.866
Fuji	RMSE	88.676	81.011	131.031	76.843	71.962	85.145
	MAE	72.083	63.186	115.700	58.248	60.836	65.655
	MAPE	38.027	28.593	58.143	26.380	26.888	28.173
Golden Delicious	RMSE	78.626	25.865	37.082	21.558	73.729	15.347
	MAE	64.765	21.744	30.881	18.321	60.706	12.411
	MAPE	204.457	51.596	103.719	38.000	200.559	25.164
Granny Smith	RMSE	37.925	43.487	37.948	66.980	70.114	67.034
	MAE	29.726	37.259	29.738	55.903	62.769	54.432
	MAPE	9.495	12.316	9.492	18.587	20.132	17.935
Braeburn	RMSE	21.072	3.893	3.186	5.161	3.178	9.515
	MAE	19.539	3.009	2.683	4.368	2.679	6.652
	MAPE	786.039	85.375	72.400	97.870	72.708	140.836
Cosmic Crisp®	RMSE	26.205	25.690	26.209	34.302	28.625	35.732
	MAE	21.668	21.111	21.675	26.885	23.542	28.145
	MAPE	37.514	33.282	37.508	52.803	40.950	55.111
	Mean MAPE	138.580	30.607	42.503	40.743	54.911	40.825

[1] RME is Root mean square error.
[2] MAE is Mean absolute error.
[3] MAPE is Mean absolute percentage error.

Table 3.4: Investigating the impact of Cosmic Crisp® presence on other apple varieties' shipments: A SARIMAX analysis.

	Gala	Honey Crisp	Cripps Pink	Red Delicious	Fuji	Golden Delicious	Granny Smith	Braeburn	All varieties
Event start date	2019-12-02	2019-12-02	2019-12-02	2019-12-02	2019-12-02	2019-12-02	2019-12-02	2019-12-02	2019-12-02
step	23.479	2.708	2.175	-63.718	-35.670	-6.296	-20.604	-5.736	-120.610
	(32.419)	(20.721)	(10.630)	(49.734)	(31.429)	(18.534)	(22.360)	(6.538)	(130.337)
ramp	-1.643***	-1.016	0.013	-1.323**	-0.225	0.033	0.072	0.123	-2.193
	(0.321)	(1.476)	(0.623)	(0.596)	(0.334)	(0.989)	(0.275)	(0.538)	(1.887)

[1] Single, double, and triple asterisks (*, **, ***) indicate statistical significance at the 10%, 5%, and 1% levels.
[2] Standard errors are in parentheses.

Table 3.5: Investigating the impact of Cosmic Crisp® shipments on other apple varieties' shipments: A SARIMAX analysis.

	Gala	Honey Crisp	Cripps Pink	Red Delicious	Fuji	Golden Delicious	Granny Smith	Braeburn	All varieties
Event start date	2019-12-02	2019-12-02	2019-12-02	2019-12-02	2019-12-02	2019-12-02	2019-12-02	2019-12-02	2019-12-02
Shipment of Cosmic Crisp®	-0.167	0.151	0.118	0.451**	0.277	0.172**	0.350***	0.011	2.736***
	(0.222)	(0.125)	(0.080)	(0.214)	(0.182)	(0.071)	(0.128)	(0.009)	(0.619)

[1] Single, double, and triple asterisks (*, **, ***) indicate statistical significance at the 10%, 5%, and 1% levels.
[2] Standard errors are in parentheses.

Table 3.6: Investigating the impact of Cosmic Crisp® presence on other apple varieties' FOB prices: A SARIMAX analysis.

	Gala	Honey Crisp	Cripps Pink	Red Delicious	Fuji	Golden Delicious	Granny Smith	Braeburn
Event start date	2019-12-02	2019-12-02	2019-12-02	2019-12-02	2019-12-02	2019-12-02	2019-12-02	2019-12-02
step	0.200	1.917	-1.406	-0.876	-0.256	0.370	-0.799	-0.727
	(0.752)	(2.901)	(1.401)	(0.868)	(1.244)	(0.897)	(0.935)	(1.248)
shipment	-0.002***	-0.010***	-0.011***	-0.004***	-0.005***	-0.009***	-0.006***	-0.034***
	(0.000)	(0.003)	(0.004)	(0.000)	(0.001)	(0.001)	(0.001)	(0.007)

[1] Single, double, and triple asterisks (*, **, ***) indicate statistical significance at the 10%, 5%, and 1% levels.
[2] Standard errors are in parentheses.

Figure 3.1: Weekly Shipments of Apples from Washington, August 25, 2008, to February 27, 2023.

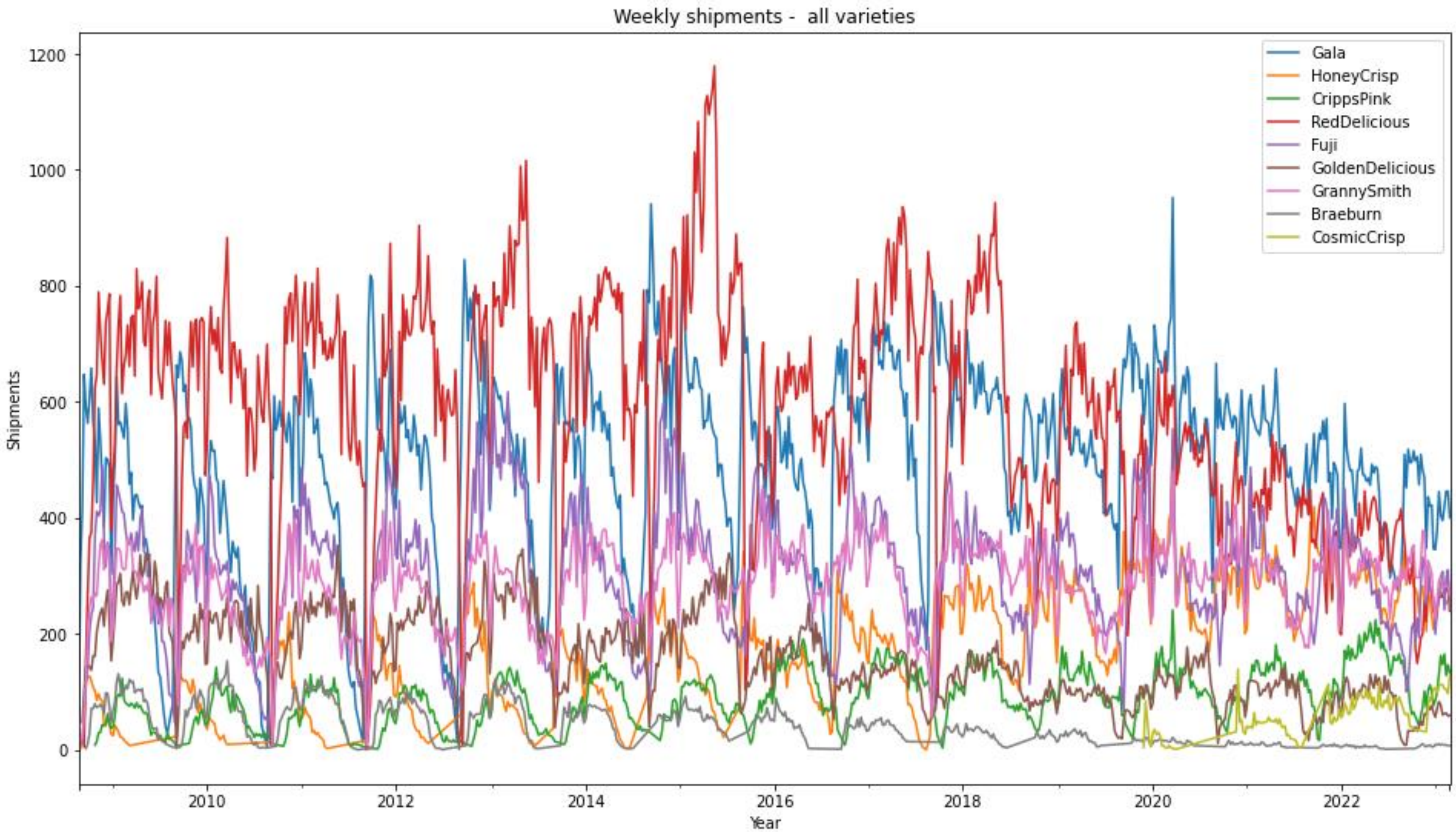
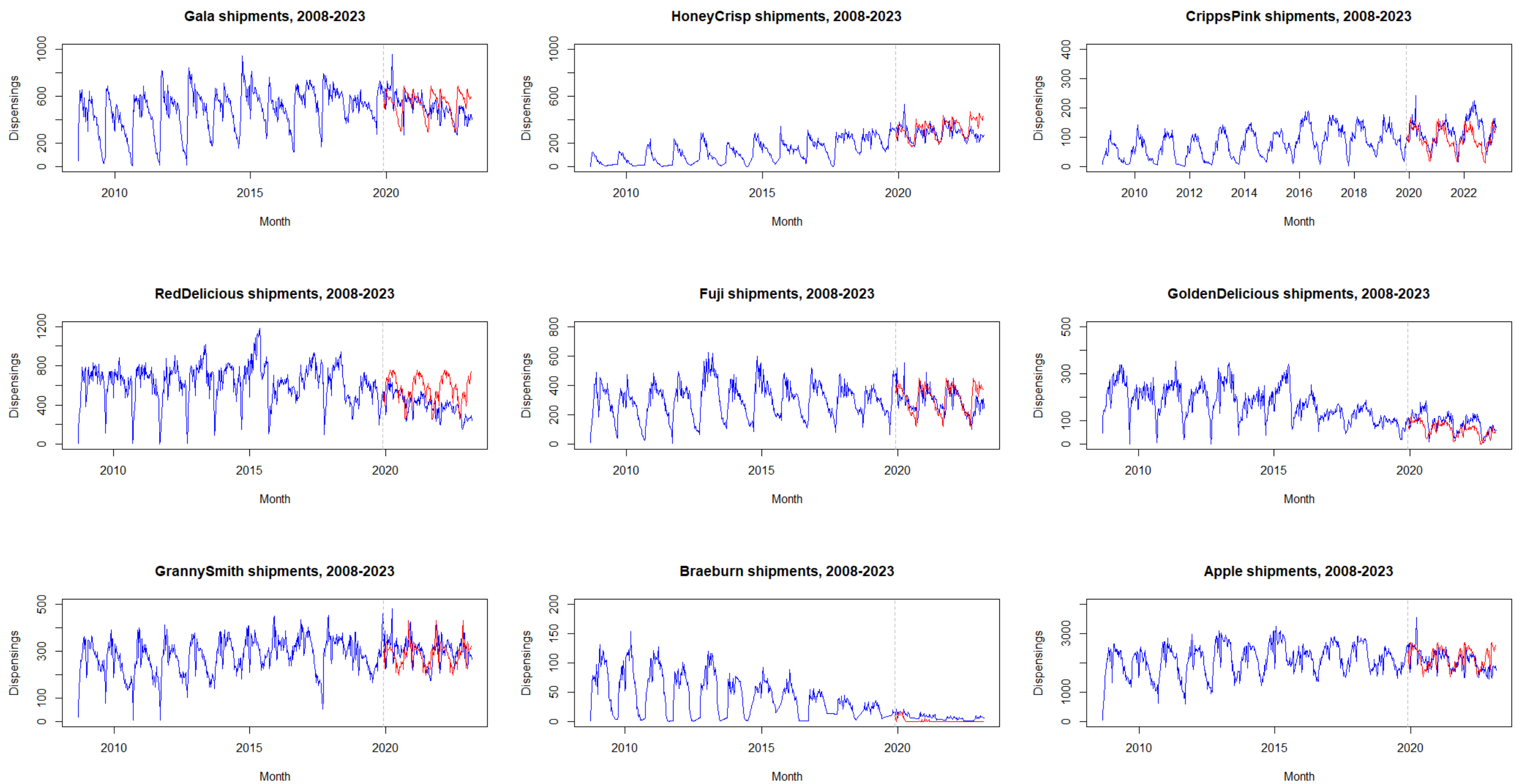


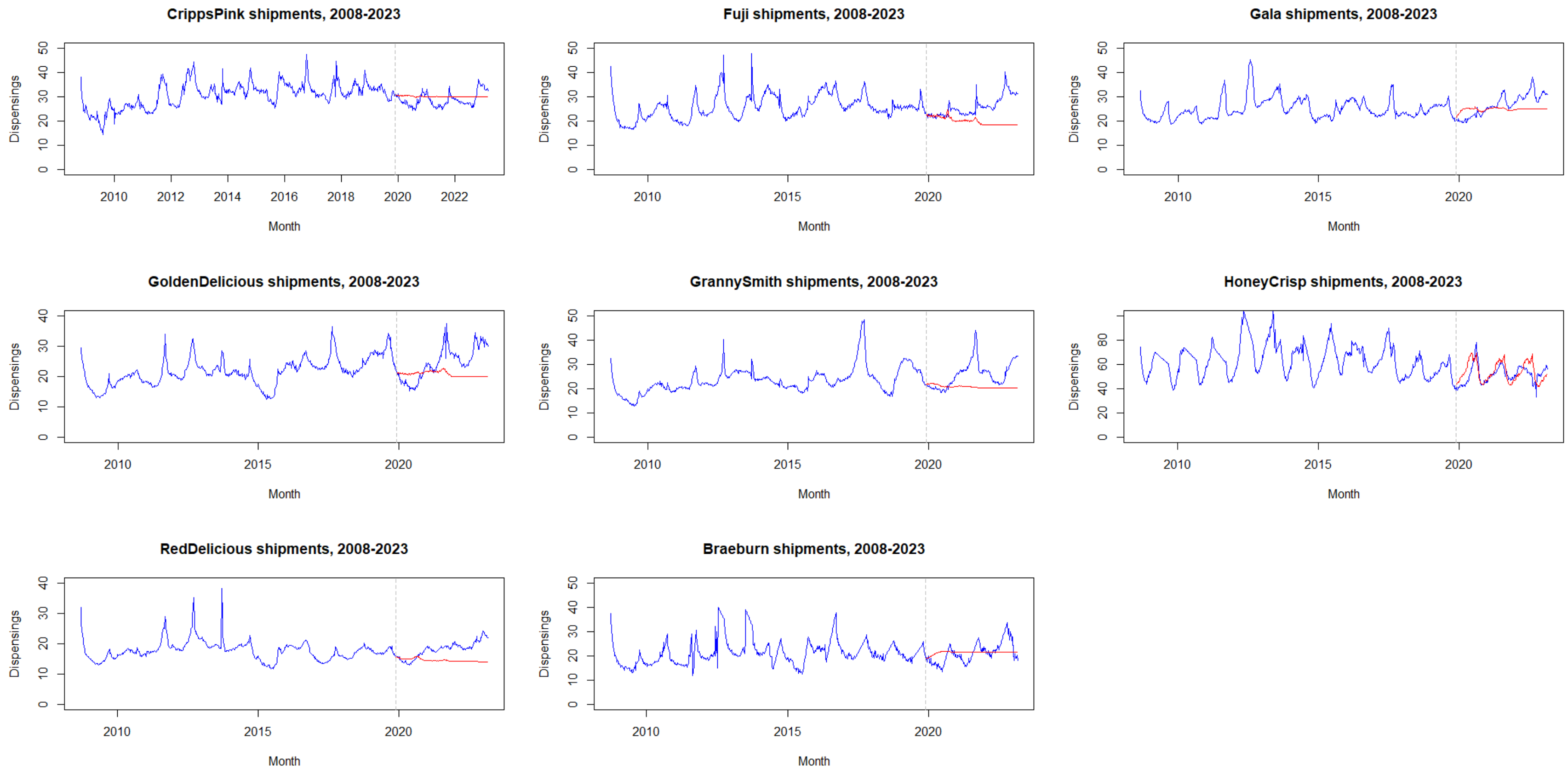
Figure 3.2: Predicted shipments vs. actual shipments

Figure 3.3: Comparison of Actual Shipments to SARIMAX Model Predictions in the Absence of Intervention.



Note: The blue line corresponds to the actual shipments, while the red line signifies the SARIMAX model's predicted values in the absence of intervention.

Figure 3.4: Comparison of Actual FOB Prices to SARIMAX Model Predictions in the Absence of Intervention.



Note: The blue line corresponds to the actual FOB prices, while the red line signifies the SARIMAX model's predicted values in the absence of intervention.

APPENDIX

APPENDIX 1.A: DRIED CRANBERRY SURVEY: NARRATIVE PROVIDED TO RESPONDENTS BEFORE COMPLETING THE DISCRETE CHOICE SCENARIOS UNDER TREATMENT 4, EXAMPLE OF A DISCRETE CHOICE SCENARIO, AND CERTAINTY SCALE – TREATMENT 4.

You will be presented with six scenarios simulating DRIED CRANBERRY sale offers. Each scenario includes two alternative bags of dried cranberries (Option A and B) that vary in levels of total sugar content, intensity of cranberry flavor, cranberry breeding technology, and prices. You will be asked to choose the ONE option you would buy as if you were facing these exact choices in a real store. The third alternative (Option C) gives you the choice of not buying any of the A or B options. Each dried cranberry option will vary by the following attributes:

(1) Total Sugar Content
(*please see line highlighted in red in image below for a reference of the "Total Sugars" information that will be presented to you*)

(2) Cranberry Flavor Intensity of cranberry flavor: Flavor refers to the overall combination of sensations, and it is influenced by the taste, aroma, look, and texture.
– Bland /weak cranberry flavor
– Full /intense cranberry flavor

(3) Breeding technology of cranberry fruit: The desired cranberry traits (e.g., sweetness, level of acidity) could be achieved by different plant breeding technologies:

- Conventional breeding: Plants with desirable traits are bred together, using existing varieties or the offspring of previous breeding programs that have the desired traits. This results in hundreds of potentially desirable plants that must be whittled down to the best candidates for commercial use. Crops improved using conventional breeding may be labelled as GMO–free or organic (if other production and certification requirements are satisfied).

- Gene editing (e.g., CRISPR): Specific genes can be altered, <u>without introducing genes from any other sources</u>. Similar to editing a word in a novel, gene editing can target specific DNA sequences in the genome for slight modification, which can improve plant traits. The USDA recently proposed that plants produced using gene editing will be treated the same as conventionally bred plants. For this study we can assume cranberries produced using gene–editing may also be labeled as GMO–free or organic (if other production and certification requirements are satisfied).

(4) Price per 6–oz bag

- $ 1.99
- $ 2.99
- $ 3.99

PLEASE KEEP IN MIND THAT

Studies have shown that answering a question about a hypothetical purchase decision, as if the purchase was for real, is difficult for many people. Usually survey respondents indicate they are more likely to state that they would buy a product when responding to a survey than when the purchase decision is real and they have to pay for the product. This happens because respondents might think “Sure, I will buy this product”, but when the decision actually involves digging into their pockets to pay for it, respondents might think instead “Do I really want to spend my money on this product?”. We ask that you try to avoid this situation and answer the following questions as you would if you were really shopping at the store and had to pay for a bag of dried cranberries.

Please carefully read the following information.

The FDA defines “Added Sugars” as sugars that are added during the processing of foods. Added sugars increase calories without contributing important nutrients. The Dietary Guidelines for Americans recommend limiting the daily amount of added sugars consumed to no more than 10% of total calories per day (which is equivalent to 200 calories or 50 grams per day). Diets lower in sugar–sweetened foods are associated with a reduced risk of developing cardiovascular disease.

Cranberries are considered a superfood due to their high nutrient and anthocyanin content. Anthocyanins are substances that can prevent or slow damage to cells caused by free radicals. The anthocyanin properties of cranberries provide multiple health benefits, including the support of cardiovascular health and reduction of the risk of some cancers.

Choose only THE ONE option (either A or B) that you WOULD REALLY BUY. Otherwise, please select Option C to indicate you would not buy option A or B.

	Option A	Option B	Option C
Total Sugars (per Serving Size: 1/4 cup)	 14 grams of sugars	 29 grams of sugars	I would not buy any of these products
Cranberry Flavor	Full /intense	Bland /weak	
Cranberry Breeding Method	Gene edited	Conventional	
PRICE ($ / 6–oz bag)	$3.99	$2.99	

	OPTION A	OPTION B	OPTION C
I WOULD CHOOSE	○	○	○

In a scale of 1 to 10, where 1 = Very uncertain and 10 = Very certain, how certain are you of your answer above?

Very uncertain Very Certain

1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○

APPENDIX 1.B: CRANBERRY JUICE SURVEY: NARRATIVE PROVIDED TO RESPONDENTS BEFORE COMPLETING THE DISCRETE CHOICE SCENARIOS UNDER TREATMENT 4, EXAMPLES OF A DISCRETE CHOICE SCENARIO, AND CERTAINTY SCALE UNDER TREATMENT 4.

You will be presented with six different scenarios simulating sale offers of Cranberry Juice. Each scenario includes three alternative juice options (100% juice, cocktail, and juice blend). These vary in the levels of total sugar content, intensity of cranberry flavor, cranberry breeding technology, and prices. You will be asked to choose the ONE option you would buy as if you were facing these exact choices at the store. The fourth alternative (Option D) gives you the choice of not buying either of the other options.

The three cranberry juice options you will see are:

- 100% Juice

You see the words "100% Juice" on the label. This is cranberry juice mixed with other fruit juices from concentrate (apple, grape, pear).

- Cocktail

You will see the word "Cocktail" on the label. This cranberry juice contains less than 100% juice with other ingredients such as water and sugar.

- Juice Blend

You will see the words "Cran–#Name of other fruit" (e.g., Cran-Apple, Cran–Cherry) on the label. This is cranberry juice that is blended with another fruit juice. This product contains less than 100% juice with other ingredients such as water and sugar.

Each cranberry juice option will vary by the following attributes:

(1) Total Sugar Content
(*see line highlighted in red in image below for a reference of the "Total Sugars" information that will be presented to you*)

(2) Cranberry Flavor Intensity of cranberry flavor: Flavor refers to the overall combination of sensations, and it is influenced by the taste, aroma, look, and texture.

- Bland /weak cranberry flavor
- Full /intense cranberry flavor.

(3) Breeding technology of cranberry fruit: The desired cranberry traits (e.g., sweetness, level of acidity) could be achieved by different plant breeding technologies:

- Conventional breeding: Plants with desirable traits are bred together, using existing varieties or the offspring of previous breeding programs that have the desired traits. This results in hundreds of potentially desirable plants that must be whittled down to the best candidates for commercial use. Crops improved using conventional breeding may be labelled as GMO–free or organic (if other production and certification requirements are satisfied).

- Gene editing (e.g. CRISPR): Specific genes can be altered, without introducing genes from any other sources. Similar to editing a word in a novel, gene editing can target specific DNA sequences in the genome for slight modification, which can improve plant traits. The USDA recently proposed that plants produced using gene editing will be treated the same as conventionally bred plants. For this study we can assume cranberries produced using gene–editing may also be labeled as GMO–free or organic (if other production and certification requirements are satisfied).

(4) Price per 64 fl oz bottle

- $ 2.49

- $ 2.99
- $ 3.49

PLEASE KEEP IN MIND THAT

Studies have shown that answering a question about a hypothetical purchase decision, as if the purchase was for real, is difficult for many people. Usually survey respondents are more likely to state that they would buy a product when responding to a survey than when the purchase decision is real and they have to pay for the product. This happens because respondents might think "Sure, I will buy this product", but when the decision actually involves digging into their pockets to pay for it, respondents might think instead "Do I really want to spend my money on this product?". We ask that you try to avoid this situation and answer the following questions as you would if were really shopping at the store and had to pay for a bottle of cranberry juice.

Page Break ---

Please carefully read the following information.

The FDA defines "Added Sugars" as sugars that are added during the processing of foods. Added sugars increase calories without contributing important nutrients. The Dietary Guidelines for Americans recommend limiting the daily amount of added sugars consumed to no more than 10% of total calories per day (equivalent to 200 calories or 50 grams per day). Diets lower in sugar-sweetened foods are associated with a reduced risk of developing cardiovascular disease.

Cranberries are considered a superfood due to their high nutrient and anthocyanin content. Anthocyanins are substances that can prevent or slow damage to cells caused by free radicals. The anthocyanin properties of cranberries provide multiple health benefits, including the support of cardiovascular health and reduction of the risk of some cancers.

Page Break ---

Choose only ONE option that you WOULD REALLY BUY. Otherwise, please select the None option.

	100% Juice	Cocktail	Juice Blend	None
Total Sugars (per Serving Size: 1 cup)	 25 grams of sugars	 12 grams of sugars	 25 grams of sugars	
Cranberry Flavor	Full /intense	Full /intense	Full /intense	
Cranberry Breeding Method	Conventional	Conventional	Conventional	
PRICE ($ / 64 fl oz bottle)	$3.49	$2.99	$2.49	
I WOULD CHOOSE	○	○	○	○

On a scale of 1 to 10, where 1 = Very Uncertain and 10 = Very Certain, how certain are you of your answer above?

Very uncertain Very Certain

1	2	3	4	5	6	7	8	9	10