DOI: 10.1111/ajae.12257

ARTICLE

Nonparametric segmentation methods: Applications of unsupervised machine learning and revealed preference

Joey Blumberg[1] | **Gary Thompson**[2]

[1]Department of Agricultural and Resource Economics, Colorado State University, Ft. Collins, Colorado, USA

[2]Department of Agricultural and Resource Economics, University of Arizona, Tucson, Arizona, USA

Correspondence
Gary Thompson, Department Head, Department of Agricultural and Resource Economics, University of Arizona, 650 N. Park Ave., Tucson, AZ 85719, USA.
Email: gdthomps@email.arizona.edu

Funding information
Economic Research Service; National Institute of Food and Agriculture; United States Department of Agriculture

Abstract
Many recent efforts by econometricians have focused on supervised machine learning techniques to aid in empirical studies using experimental data. By contrast, this article explores the merits of unsupervised machine learning algorithms for informing *ex ante* policy design using observational data. We examine the extent to which groups of consumers with differing responses to economic incentives can be identified in a context of fruit and vegetable demand. Two classes of nonparametric algorithms—revealed preference and unsupervised machine learning—are compared for segmenting households in the National Consumer Panel. Nonlinear almost-ideal demand models are estimated for all segments to determine which methods group households into segments with different expenditure and price elasticities. In-sample comparisons and out-of-sample prediction results indicate methods using price-quantity data alone—without demographic, geographic, or other variables—perform better at segmenting households into groups with sizeable differences in price and expenditure responsiveness. These segmentation results suggest considerable heterogeneity in household purchasing behavior of fruits and vegetables.

KEYWORDS
fruit and vegetable consumption, revealed preferences, segmentation, unsupervised machine learning

JEL CLASSIFICATION
C38, C55, C81, D12

The last decade has brought significant effort toward coalescing machine learning (ML) techniques with applied econometrics (Athey, 2018; Athey & Imbens, 2019; Mullainathan & Spiess, 2017;

Varian, 2014). Ongoing developments in the ML literature are providing useful tools for analytics, particularly in big data settings. The predictive power of ML algorithms continues to improve, and methods for handling complex, unstructured data types like text, images, and sound have enabled new opportunities for analysis (Gentzkow et al., 2019; Varian, 2018). So far, applied econometricians have focused primarily on using *supervised* ML predictions in service of causal inference (Athey, 2018; Mullainathan & Spiess, 2017). By contrast, in this article we compare the efficacy of *unsupervised* learning techniques with revealed preference algorithms, both of which are nonparametric methods capable of identifying groups of observations that are homogeneous by some measure. Having identified those groups of observations, we then estimate demand elasticities for each group to assess within-group homogeneity and across-group heterogeneity in tastes and preferences in the context of fruit and vegetable demand.

Most practitioners of machine learning divide approaches into supervised and unsupervised learning. Supervised learning, the center of recent focus, is principally concerned with prediction, not inference. Supervised methods involve using "features" or "predictors" (covariates) to predict outcomes of a dependent variable; these models are predominately used when the data generating process is unknown. A supervised model is calibrated on a "training" set of data, and the quality of model prediction is measured by goodness of fit in separate, "validation" subsamples. Predictions out of sample from competing individual models or "ensembles" (collections of models) are compared to choose the best models. Several methods are commonly used to avoid overfitting in sample and thereby produce better out-of-sample predictions: penalized objective functions, cross-validation, and bootstrap sampling. Penalizing objective functions, referred to as "regularization" in ML nomenclature, moderate the tendency to introduce more complexity when training models in sample. Model robustness is usually ascertained by "tuning," a process of comparing model predictions as key parameters of models are varied over grid searches. Because supervised ML approaches often conceive of models as algorithms, asymptotic properties of model estimators are often not a concern.

For policy evaluation, several approaches for incorporating supervised ML tools in estimating treatment effects have been proposed. Random forests (Breiman, 2001) and the least absolute shrinkage and selection operator, or LASSO (Tibshirani, 1996), can be useful for variable selection or propensity scoring when dealing with a large number of covariates; however, regularized estimators often lose desirable theoretical properties (Ju et al., 2019). Belloni et al. (2017) propose a multistep procedure to estimate average treatment effects where LASSO is used to select two sets of covariates, those correlated with the outcome and those correlated with the treatment, and a union of the two sets is included in an ordinary least-squares regression. Wager and Athey (2017) extend the random forest classification algorithm to develop a class of nonparametric methods—dubbed "causal forests"—for estimating heterogeneous treatment effects. Atypical of standard ML techniques, both methods above produce well-behaved estimates with valid confidence intervals. Athey and Imbens (2017) provide a comprehensive overview of recent developments in machine learning for estimating causal effects.

The other branch of ML, unsupervised learning, has received scant attention among applied econometricians (Athey & Imbens, 2019; Storm et al., 2020). In ML parlance, unsupervised learning involves data with no "labels" (outcome or dependent variables). Unsupervised learning either seeks to discern patterns among variables and observations or reduce the dimension of the covariate space when the number of candidate variables is large and potentially correlated. Principal component analysis is one unsupervised learning technique long familiar to applied econometricians that subsumes the variation of many covariates into mutually orthogonal principal components (Theil, 1971, pp. 46–55).

The strand of unsupervised learning used in this analysis is concerned with discovering groups of observations similar to one another. We will refer to these grouped observations as segments. Grouping observations into segments can be useful computationally because it partitions data sets with large numbers of observations into segments of more manageable size. More substantively, grouping observations into segments may reduce within-group heterogeneity while revealing heterogeneity across groups. Having partitioned observations into segments, separate models can be run on each segment for inference and prediction. Those separate models can yield results specific to segments that would otherwise be masked if a single model were fit using the entire sample (Lusk, 2017).

The notion of identifying distinct segments of consumers and designing appropriate policy interventions targeting those segments is not new. When employing household-level data to estimate consumer demand models, researchers frequently segment sample data. Whether implicit or explicit, segmentation is motivated by the assumption that a single utility function is not compatible with the diverse consumer behavior observed in large cross-sectional and panel data sets. Segmentation is often implemented using non-stochastic partitioning rules based on observables like income and demographic variables (Park et al., 1996; Zhen et al., 2014). Once partitioned, parametric tests for equality of parameters or elasticities across segments may be used to verify whether separate utility functions are evident in each segment (Mhurchu et al., 2013). In finite mixture models, rather than choosing the number of segments a priori, the number of segments is chosen by Akaike or Bayesian information criteria, and households are assigned to the segment for which the posterior probability is the highest (Bertail & Caillavet, 2008). In a similar spirit, Jensen and Manrique (1998) estimated Engel equations and then grouped households based on both income and homogeneity of residual variances from the Engel regressions. These approaches rely on either predetermined cutoffs or particular parametric specifications of demand relationships.

The alternative nonparametric approaches to segmentation employed in this article include revealed preference and unsupervised learning algorithms. Revealed preference algorithms partition sample observations in segments each consistent with an unknown, underlying utility function (Crawford and Pendakur, 2012a; Varian, 1982, 1983). After partitioning, separate parametric demand models may be estimated for each segment (Crawford and Pendakur, 2012a). The second approach, more often used by data analysts in industry, utilizes unsupervised clustering methods such as k-means. Cluster analysis is based largely on intuition, without appeal to consumer theory (Pradeep et al., 2019). Lusk (2017) provides one such application of using a clustering algorithm to first partition a sample of data and then estimate separate demand models for each segment. To date, the effectiveness of these nonparametric methods of segmentation in finding clusters of households that display different economic behavior has not been analyzed. We investigate the relative efficacy of theory-based partitioning with a revealed preference algorithm versus unsupervised machine learning clustering algorithms, neither of which maintain a particular parametric specification of the demand model.

Unsupervised machine learning techniques or revealed preference methods that isolate different segments may provide a useful complement to supervised ML techniques used in service of causal inference. In seeking to identify treatment effects, supervised ML models are often employed with experimental data to evaluate programs or policies *ex post*, that is, after a field or natural experiment has occurred. In contrast, unsupervised ML techniques and revealed preference methods may be used with observational data for policy design *ex ante* by identifying differing degrees of response to changes in economic incentives.

In the context of consumer demand, important differences in tastes and preferences may emerge across segments. Using household data on fruit and vegetable purchases, we find significant differences in responses to price changes and group expenditure across segments. Demand for fruits and vegetables is important because most U.S. consumers do not consume the recommended amounts (Lee-Kwan et al., 2017). Low consumption of fruits and vegetables has persisted despite decades of public programs promoting healthier diets (Casagrande et al., 2007; Lin et al., 2016). Even modest increases in fruit and vegetable consumption from current low levels could lead to significant reductions in incidence of cardiovascular disease, cancers, and all-cause mortality (Aune et al., 2017).

Differences in responsiveness across segments have important implications for policy interventions aimed at increasing consumption of fruits and vegetables. For example, consider the impact of "double-dollar" programs, which augment Supplementary Nutrition Assistance Program (SNAP) expenditures by giving consumers an extra dollar to spend on fresh fruits and vegetables for every SNAP dollar spent. For consumers in segments who are responsive to price changes, these "double-dollar" policies induce increased purchases of fresh fruits and vegetables. Double-dollar policies will be largely ineffective, however, for consumers in segments with inelastic price responses. Inducing price-inelastic consumers to purchase more fruits and vegetables would require effective educational and social policies. Successful segmentation can allow policymakers to tailor policies to specific segments without assuming a single policy will have uniform impacts across all consumers.

Using household food demand data from the National Consumer Panel, a largely representative sample of U.S. households (Muth et al., 2016), we find that 18 segments from a cross section of 28,050 households are sufficient to demarcate a wide range of price and expenditure responsiveness. We employ a diverse set of demographic, geographic, and shopping variables with the unsupervised machine learning methods. Perhaps surprisingly, segmentation using price and quantity variables *alone*, without demographics or other variables, is the most effective in identifying household segments with widely differing sensitivity to prices and expenditure on fruits and vegetables. This result points to pronounced heterogeneity in consumer behavior across households, which observables like demographic variables are not capable of isolating.

1 | ALTERNATIVE NONPARAMETRIC SEGMENTATION METHODS

1.1 | Revealed preference segmentation

Revealed preference orderings are based on the simple notion that if a consumer chooses a particular bundle of goods over all alternative bundles of equal cost, then that bundle is revealed preferred (Varian, 1982, p. 945). This comparison of the costs of different bundles of goods chosen by the same consumer over time requires some intersection of budget constraints to be able to reveal violations. If the consumer's income is increasing over time, and there is little variation in relative prices, budget constraints will not overlap, leaving each successive period's higher expenditure with no feasible alternative bundles for comparison.

Formally, the necessary and sufficient conditions for observed prices and quantities to be consistent with utility maximization require that the Generalized Axiom of Revealed Preference (GARP) be satisfied (Varian, 1982, 1983). To ascertain if GARP is satisfied, we need to define several terms. First, consider a sample of n observations on a vector of k quantities of goods and their corresponding prices denoted as $(\mathbf{p}_i, \mathbf{q}_i)$ $i = 1, \ldots, n$. A utility function $u(\mathbf{q})$ is said to *rationalize* a set of n observations on k prices and quantities, if $u(\mathbf{q}_i) \geq u(\mathbf{q})$ $\forall \mathbf{q}$ such that $\mathbf{p}_i'\mathbf{q}_i \geq \mathbf{p}_i'\mathbf{q}$ for all $i \geq 1, \ldots, n$. In other words, if a particular bundle of k goods produces at least as high utility as alternative bundles at a given set of prices that bundle must be at least as expensive as the alternative bundles. For any given pair of alternative bundles, if $\mathbf{p}_i'\mathbf{q}_i \geq \mathbf{p}_i'\mathbf{q}_j$, then the $\mathbf{q}_i$ is <u>*directly revealed preferred*</u> to $\mathbf{q}_j$ and is expressed as $\mathbf{q}_i R^o \mathbf{q}_j$. A particular bundle being *revealed preferred* over a sequence of alternative bundles, $(\mathbf{q}_i, \mathbf{q}_j, \mathbf{q}_l, \ldots, \mathbf{q}_m)$, is denoted as $\mathbf{q}_i R \mathbf{q}$ where $\mathbf{p}_i'\mathbf{q}_i \geq \mathbf{p}_i'\mathbf{q}_j$, $\mathbf{p}_j'\mathbf{q}_j \geq \mathbf{p}_j'\mathbf{q}_l, \ldots, \mathbf{p}_m'\mathbf{q}_m \geq \mathbf{p}_m'\mathbf{q}$. This sequence of inequalities is referred to as the transitive closure of R^o. Finally, if $\mathbf{q}_i R \mathbf{q}_j$ implies $\mathbf{p}_j'\mathbf{q}_j \geq \mathbf{p}_j'\mathbf{q}_i$, then GARP is satisfied. Afriat's theorem demonstrates that if the data satisfy GARP, there exists a nonsatiated, continuous, concave, monotonic utility function that rationalizes the data (Afriat, 1967; Varian, 1983). As an empirical check for consistency with GARP, Varian operationalized computation of the transitive closure by adapting an algorithm due to Warshall (Varian, 1982, p. 972).

For segmentation purposes, rather than checking for the consistency of an individual's preferences over time, we seek to find groups of households whose choice behavior could be consistent with a common utility function at any particular time (Gross, 1995a). Dean and Martin (2010) demonstrated how to find the largest number of cross-sectional observations consistent with a single utility function as indicated by revealed preferences. Recognizing a single utility function may not be adequate for characterizing the economic behavior of an entire sample, Crawford and Pendakur (2012a) developed a method to provide lower and upper bounds on the number of segments necessary to rationalize observed price-quantity combinations in sample data.

Satisfaction of GARP could be taken as a binary outcome: Either some subset of the sample data is consistent with GARP or it is not. However, acknowledging that there may be errors in data—measurement error owing to imperfect data collection and optimization error due to individuals themselves making choices not quite consistent with utility maximization (Harbaugh et al., 2001)—various methods have been developed to deal with sample observations that are not entirely consistent with GARP (Beatty & Crawford, 2011; Dean & Martin, 2016; Varian, 1985). These methods are beyond the scope of our purpose here, which is simply to use revealed preference to segment subsets of households into groups with common utility functions.

1.2 | Unsupervised machine learning methods

For segmentation, clustering algorithms are the tools we employ from unsupervised machine learning. The choice of variables to include in the cluster analysis is left to the judgment of the analyst. Unlike in revealed preference algorithms where only prices and quantities of relevant goods are used, potentially all variables in a particular data set or any subset thereof could be included for clustering. The practical problem is to decide which variables to include. Because there is no selection of a dependent variable, there is no issue of specification bias. However, some authors caution that including "trash" variables may worsen segmentation results (Kaufman & Rousseeuw, 1990, p. 14). Various procedures have been proposed for variable selection (Fraiman et al., 2008), but most practitioners appeal to domain-specific knowledge (Hastie et al., 2008)—what econometricians refer to as knowledge of the data generating process—for choosing variables.

Having chosen the variables to include for segmentation, a measure of similarity or dissimilarity must be chosen so the algorithms can calculate distances within and between clusters. The choice of dissimilarity measures is an important aspect to clustering that receives relatively less attention than clustering algorithms (Hastie et al., 2008, p. 506). When all variables are continuous, either Euclidean distance or sums of absolute differences can measure similarity. But if binary and categorical variables are also included, similarity across pairs of observations needs to be coded, typically on the zero–one interval with zero for most dissimilar and one for identical values. If the analysis mixes all three types of variables, the Gower coefficient is one way to map continuous variables to the zero–one interval (Gower, 1971, 1985). For any pair of observations on the k^{th} continuous variable, (x_{ik}, x_{jk}), the similarity measure, s_{ijk}, is calculated as $s_{ijk} = 1 - \frac{|x_{ik} - x_{jk}|}{R_k}$ where R_k is the range of values, that is, the difference between the minimum and maximum value of the k^{th} variable.[1] The overall similarity measure for any pair of observations can be obtained as a simple or weighted average of the s_{ijk}.

Once the variables to include have been chosen and their similarity measures calculated, a clustering algorithm must be selected. One of the most popular clustering algorithms is the k-means algorithm. A similar algorithm using medians rather than means is the *partitioning on "medoids"* or PAM algorithm. For a fixed number of k clusters, the algorithms iteratively assign observations to the nearest mean (median) by minimizing the within-cluster sum of squares (L2 norm) or the sum of the mean absolute deviations (L1 norm), sometimes referred to as "Manhattan" distance. The algorithms continue until the assignment of observations to a specific cluster does not change. Running the algorithm for successive values of k and plotting the total within-cluster sum of squares can help determine the number of clusters visually. Other approaches to choosing the number of clusters include bootstrapping (Hennig & Lin, 2015; Luo et al., 2019) and cross validation (Fu & Perry, 2020; Wang, 2010).

2 | DATA AND DESCRIPTIVE STATISTICS

To assess alternative segmentation methods, we focus on household demand for fruits and vegetables, employing trip-level data collected by households in the National Consumer Panel.[2] These data consist of self-reported purchase data recorded with a smart phone app or scanning device after each trip to specific grocery and food retailers. In most cases, IRI inserts the relevant price so that the shopper only needs to scan the bar code. For random-weight products with no bar code, the shopper is asked to select from a menu of generic product categories such as "lettuce" or "tomatoes." Shoppers also record the store name and trip date. When a household is initially accepted into the panel, geographic and demographic information are self-reported. IRI estimates about three quarters of households update their demographic and geographic data annually (Muth et al., 2016). IRI endeavors to maintain a nationally representative sample of consumers based on this geographic and demographic information.

The National Consumer Panel (NCP) data have several limitations. First, only purchases of food to be consumed at home are recorded. Second, as with any self-reported survey data, underreporting by households in NCP occurs. Data from all trips may not be recorded, and random-weight items requiring more time to record may be omitted. Information is less likely to be recorded the larger the number of items purchased per trip (Einav et al., 2010). Several studies have compared the potential incidence of underreporting by comparing household expenditures on various food categories and segments of consumers in governmental surveys like the Consumer Expenditure Survey and FoodAPS. The categories with the largest apparent underreporting are fruits and vegetables (Sweitzer et al., 2017; Zhen et al., 2014). Muth et al. (2016) found that households with children, households in lower income brackets, and households with heads under age 35 are the least likely to report consistently. Participants in commercial household panels may be more price sensitive than consumers not in the panel (Lusk & Brooks, 2011). Notwithstanding these deficiencies, food demand analysis with *Homescan* and Consumer Expenditure Survey data can produce very similar results (Boonsaeng & Carpio, 2019).

Despite the limitations, NCP data afford a level of specificity in products purchased not found in governmental surveys. Except for random-weight products, all are uniquely identified by universal product code (UPC) and prices are recorded, obviating the need to deal with the problems of unit values (Deaton, 1998; McKelvey, 2011). Because trips are recorded for as long as the household is a member of the panel, time of year when surveys are conducted is not a problem with NCP data.

The NCP household trip data are first aggregated temporally for each household for the calendar year to reduce the incidence of zero expenditures owing to seasonality and infrequency of purchase. Some households are observed purchasing fruit and vegetable products with prices exceeding \$15.00/lb. or lower than \$0.10/lb. Accordingly, observations with prices at the top and bottom 1% are removed. Recorded purchases of random-weight products require unit price imputation because households record total amount paid but not quantity. Average unit prices for similar fresh products with a barcode ("fixed weight") are matched to the closest random-weight product. Emulating IRI methodology, prices are imputed using retail-chain-specific data in 73 unique marketing areas. If there are no observed purchases for a product category for a specific chain within a marketing area, the average price from that entire marketing area is assigned. If there are no observed purchases for a product category within an entire marketing area, a total sample average is used. Total sample averages are imputed and assigned to less than 1% of the observations. Once unit prices are imputed for random-weight products, quantities can be generated from the total amount paid. Imputation undoubtedly introduces measurement error; however, entirely excluding expenditures on random-weight products would exclude fruits and vegetables that may be substitutes or complements from the analysis, resulting in sample selection bias.

We explicitly chose relatively aggregate categories of fruits and vegetables to avoid the additional complexity of modeling zero consumption in the 156 segments identified below. For tractability, we chose four groups of fruits and vegetables for the demand analysis: fresh vegetables (FV), storable vegetables (SV), fresh fruit (FF), and storable fruit (SF). Fresh categories include both fresh and fresh-processed products. Storable products include frozen, canned, and dried as well as shelf-stable juice products. We experimented with grouping products according to convenience—fresh processed fruits and vegetables ready to eat, microwaveable vegetables, and so on—but the incidence of zero household consumption in those categories precluded us from aggregating by convenience categories. Table 1 displays the number of unique UPCs in each of the four fruit and vegetable categories.

To assess how well segmentation methods performed, we chose data from calendar year 2016 for in-sample measures and segmented the same households out of sample for 2017, the most recent year available at the time of analysis. To segment out of sample by household, we had to exclude any households not reporting information regularly for 2016 and 2017. Though NCP is composed of roughly 120,000 households, only about half report sufficient expenditure with adequate frequency to be deemed "static" panel members by IRI.[3] In 2016, 48,521 households met the static criteria; fewer, 38,059, met the criteria for both 2016 and 2017. Selected demographic variables in Table 2 indicate that compared to the population at large (column I), static households (column II) are generally older, two-person households with proportionally more households with income ranging from $35,000 to $100,000. Fewer static households have children. Households with Hispanic and Black members are underrepresented. But comparing the static households for 2016 with those for both 2016 and 2017, the prevalence of demographic variables changes very little.

Fresh fruits and vegetables are generally more expensive than their storable counterparts (see Table 3). There is a wide range of prices paid from as low as $0.25/lb. to just over $10.00/lb. Consistent with low levels of consumption nationally, some households scarcely consume selected categories of fruits and vegetables. At the other extreme, a few households averaged between 20 and 30 pounds per week of a fruit or vegetable category purchased. Group expenditure shares are generally higher for fresh products because relatively higher quantities are consumed at higher prices. The median number of weekly store trips of 1.65 corresponds to median weekly food expenditures of $66.21 with about 13% or $8.61 spent on fruits and vegetables.

3 | EMPIRICAL APPROACH

To gauge how heterogeneity in price and expenditure elasticities can be captured by segmentation, we use various segmentation methods generating the same number of segments for the fixed number of NCP households. A nonlinear AIDS model (Deaton & Muellbauer, 1980) is then fit to each segment, imposing symmetry and homogeneity.[4] Uncompensated price and expenditure elasticities in 2016 and 2017 are catalogued and compared across segments and segmentation methods.

For segmentation with revealed preference, we check the transitive closure for GARP using the Floyd-Warshall algorithm in the *revealedPrefs* package in R (Boelaert, 2019). We also calculate the minimum and maximum number of utility functions necessary to rationalize all households in the 2016 data. The algorithm for calculating the minimum number of utility functions samples

TABLE 1 Unique universal product codes (UPCs) by category and weight type

Fixed weight			Random weight		
Category	**Count**	**Percent**	**Category**	**Count**	**Percent**
Fresh veg.	38,464	0.23	Fresh veg.	21	0.51
Storable veg.	43,394	0.26	Fresh fruit	20	0.49
Fresh fruit	36,690	0.22	Total	41	1.00
Storable fruit	51,546	0.30			
Total	170,094	1.00			

TABLE 2 Comparison of demographic characteristics, American Community Survey (ACS) vs. National Consumer Panel

Demographic variable	ACS estimate	Static panel, 2016	Static panel, 2016–2017	Static panel, 18 segments
Household size				
1 person	28.0	26.4	27.2	25.8
2 person	33.9	35.4	37.5	37.4
3 person	15.6	13.4	12.8	13.1
4+ person	22.5	24.9	22.5	23.7
Age of household head				
<35 years	18.9	14.3	11.1	11.5
35–64 years	56.1	59.2	60.3	60.0
65+ years	25.1	26.5	28.7	28.5
Annual household income				
<$15,000	11.5	7.3	6.9	6.6
$15,000–$34,999	19.2	22.0	21.4	21.1
$35,000–$59,999	20.7	22.1	21.8	22.0
$60,000–$99,999	22.3	24.0	24.4	24.7
$100,000+	26.2	24.6	25.4	25.6
Ethnicity				
Hispanic	17.8	11.0	10.2	10.7
Non-Hispanic	82.2	89.0	89.8	89.3
Race				
Black	12.3	10.5	10.5	10.9
Non-Black	87.7	89.5	89.5	89.1
Presence of children				
Yes	31.1	31.2	27.8	29.1
No	68.9	68.8	72.2	70.9
Sample size	117,716,237	48,521	38,059	28,050

Note: U.S. Census Bureau, American Community Survey (2016) estimate is calculated from *American FactFinder, American Community Survey 1-Year Estimates, Table S2501*. Static panels are calculated based on purchases of edible products, not total purchases of edible and other items.

households at random without replacement and checks for violations of GARP in a pairwise fashion. The algorithm for determining the maximum number of utility functions also samples without replacement, assigning each household to the largest segment for which GARP is not violated (Crawford and Pendakur, 2012b, p. 1–3).[5] The minimum number of utility functions was 18, and the maximum was 76. Choosing the 18 segments with the largest number of households yielded the largest segment with 4588 households and the smallest segment containing 628. The next largest segment had 573 households. At 22 segments each subsequent segment had fewer than 500 households, with the smallest segment comprising a single household.[6]

For segmentation with unsupervised learning, we use both the k-means and PAM clustering algorithms (Maechler et al., 2019). Rather than choose the number of segments through sample-based criteria, we fix the number of clusters at 18 in order to compare the results with GARP segments.[7] For comparison with GARP, we segmented limiting the segmentation variables in the k-means and PAM algorithms to just prices and quantities. We also employed several configurations of observables in the clustering algorithms to gauge the efficacy of other variables in identifying groups of households with different economic behavior.

TABLE 3 Descriptive statistics: Price, quantity, expenditure and trips, 2016

Item	Median	Minimum	Maximum
Price ($ / lb.)			
Fresh vegetables	1.71	0.28	10.24
Storable vegetables	1.20	0.36	9.21
Fresh fruit	1.49	0.28	10.64
Storable fruit	1.11	0.24	10.80
Quantity (lb. / week)			
Fresh vegetables	1.66	0.004	21.39
Storable vegetables	1.01	0.005	25.92
Fresh fruit	1.61	0.002	31.21
Storable fruit	1.28	0.005	31.80
Group expenditure (share)			
Fresh vegetables	0.325	0.003	0.960
Storable vegetables	0.145	0.001	0.940
Fresh fruit	0.284	0.001	0.951
Storable fruit	0.175	0.001	0.946
Weekly trip characteristics			
Group expenditure	$8.61	$0.17	$104.42
Trip expenditure	$66.21	$25.01	$1102.41
Group exp./trip exp.	0.131	0.04	0.69
Number of trips	1.65	0.27	6.88

Note: Group expenditure refers to expenditure on all fruits and vegetables.

The following nine nonparametric methods were compared:

1. *GARP, prices and quantities (GARP)*. *GARP* segmentation uses only prices and quantities.
2. *K-means, prices and quantities (KMeans PQ)*. Paralleling *GARP* variables, we use a k-means clustering algorithm with only prices and quantities. Prices and quantities are continuous variables, so a k-means algorithm based on Euclidean distances is appropriate.
3. *Partitioning around medoids, prices, and quantities (PAM PQ)*. To check for sensitivity to choice of algorithm, the same price and quantity variables are used with a medoid algorithm based on Manhattan distances (Maechler et al., 2019).
4. *Partitioning around medoids, demographic/geographic variables (PAM demo)*. We use demographic and geographic variables by themselves because they have been used in previous demand studies for segmentation. Most of the demographic/geographic variables are categorical; only age of household head is continuous. Gower (1985) standardization with no weighting is used to calculate a dissimilarity measure for each household.
5. *K-means, total food expenditure and number of shopping trips (KMeans ET)*. Instead of relying on prices, quantities, or demographics, we use two variables based on consumer behavior: total household expenditure on fruits and vegetables, and the number of trips made for food shopping in 2016.[8] Both food expenditure and number of trips are continuous variables so a k-means algorithm with Euclidean distance was chosen.
6. *Partitioning around medoids, total food expenditure, and number of shopping trips (PAM ET)*. Including both food expenditure and number of trips, we used a medoid algorithm with Manhattan distance as a check on sensitivity of segmentation to choice of clustering algorithm.

7. *Partitioning around medoids, all variables (PAM KS)*. All the variables just mentioned—prices, quantities, demographic/geographic, food expenditure, and number of trips—as well as income categories were included in order to assess whether using all variables available might account for heterogeneity better than any subset of variables. We refer to this segmentation approach as the "kitchen sink" or KS. Because it mixes all types of variables, we use Gower (1985) standardization with the medoid algorithm.
8. *Income*. Many previous demand studies have used income to segment different groups in demand studies. As a benchmark, we include segmentation on income without using an algorithm. There are 12 income categories in the National Consumer Panel. Those categories do not facilitate dividing households into quantiles so we simply formed 12 segments based on the number of households in each income category.
9. *Random*. As a second benchmark against which to compare the previous methods, we drew random samples without replacement to construct 18 segments with the same sample size for each segment as in the *GARP* segmentation. By virtue of the random sampling, we would expect little variation in economic behavior across these segments.

Median sample sizes of the segments generated by these nine methods consisted of at least 1000 households though several of the unsupervised methods generated segments with fewer than 500 households (see Table 4).

4 | SEGMENTATION RESULTS FOR 2016

Before estimating AIDS models for each segment, we checked whether the households in each segment not using *GARP* could be rationalized by a single utility function. In none of the 138 segments (a total of 156 less the 18 *GARP* segments) were all households in each segment rationalized by a single utility function. This result indicates none of the other segmentation methods, even the *KMeans PQ* and *PAM PQ* methods that use the same variables as *GARP*, segment in a manner consistent with revealed preference partitioning.

Another preliminary check regards differences in observables across segments. With 156 segments and at least nine demographic/geographic variables, we have too many segments and variables to make exhaustive comparisons within and between segmentation methods. Instead, we highlight several notable findings. Segments generated with *GARP* appear quite similar in terms of observables to those generated randomly. By contrast, the other seven methods tend to produce segments within

TABLE 4 Sample sizes by segmentation method

Segmentation method	Minimum	Median	Maximum	#Segments <500 households
GARP	628	1183	4588	0
KMeans PQ	131	1174	6325	5
PAM PQ	560	1457	3104	0
PAM Demo	958	1549	2382	0
KMeans ET	75	1585	3422	3
PAM ET	319	1593	2410	1
PAM KS	937	1554	2425	0
Income	277	2033	6159	2
Random	628	1183	4588	0
All	28,050	28,050	28,050	

Note: The first seven segmentation methods each have 18 segments. For Income, there are 12 segments as defined by IRI.

any given segmentation method that appear different as measured by observables. To illustrate, we compare household expenditures on fruits and vegetables across the 18 segments for each method. All methods but *GARP* and *Random* result in notable differences in mean weekly expenditure on fruits and vegetables across segments (see Table 5). Though *GARP*, *KMeans PQ* and *PAM PQ* methods all use price-quantity data for segmenting, *KMeans PQ* and *PAM PQ* cluster households with different mean expenditures, whereas *GARP* results in mean expenditures that differ little across segments.

4.1 | Expenditure elasticities

Expenditure elasticity values evaluated at segment-specific medians are displayed in Figure 1. To judge whether the estimated elasticities are different from zero, we use the sample-size-adjusted critical value of $\sqrt{\ln(n)}$ instead of the usual $\alpha = 1.96$ as a more conservative critical value (Cameron & Trivedi, 2005, p. 279).[9] All but 13 of the 624 expenditure elasticities (4 elasticities $\times$ 156 segments) are statistically distinguishable from zero. None of the expenditure elasticity estimates were negative.

A striking regularity in Figure 1 is that the three segmentation methods using only prices and quantities—*GARP*, *KMeans PQ* and *PAM PQ*—yield segments with the widest range of expenditure elasticities, from well below to well above unity. For example, for storable vegetables, *KMeans PQ* identifies one segment with a median expenditure elasticity of 0.535 and another with median value of 2.448, whereas estimation of a single AIDS model on the entire sample produced an expenditure elasticity of 0.776. By contrast, segmentation methods employing other variables produce median expenditure elasticities closely bracketing the corresponding median elasticities from the entire sample. The two alternative unsupervised learning algorithms using only prices and quantities, *KMeans PQ* and *PAM PQ*, give similar expenditure elasticity values, suggesting robustness of the results to the clustering algorithm.

Though the expenditure elasticities within a particular segmentation method appear to differ markedly, those estimates may not be very precise. To examine their precision, we compared confidence intervals about each expenditure elasticity across the 18 segments for each segmentation method.[10] The comparison involves 153 pairwise comparisons for each elasticity in each segmentation

TABLE 5 Weekly fruit and vegetable expenditure by segmentation method, minimum, and maximum across segments (dollars/week)

Segmentation method	Minimum weekly expenditure	Maximum weekly expenditure	Absolute value of standardized difference
GARP	9.21	10.73	0.259
KMeans PQ	4.10	38.67	3.283
PAM PQ	3.09	29.03	3.524
PAM Demo	6.40	13.58	1.053
KMeans ET	7.46	16.92	0.830
PAM ET	6.94	14.91	0.846
PAM KS	5.93	13.23	1.152
Income	6.67	12.55	0.758
Random	9.91	10.53	0.091
All	10.24	10.24	

Note: Mean weekly expenditure was calculated for each segment in each method. The minimum (maximum) weekly expenditure is the minimum (maximum) across all 18 segments for each method. Standardized differences were calculated between the minimum and maximum means by segmentation method using the formula in Austin (2009). If the absolute value of the standardized difference is less than 0.1, the two mean expenditures are judged as statistically indistinguishable.

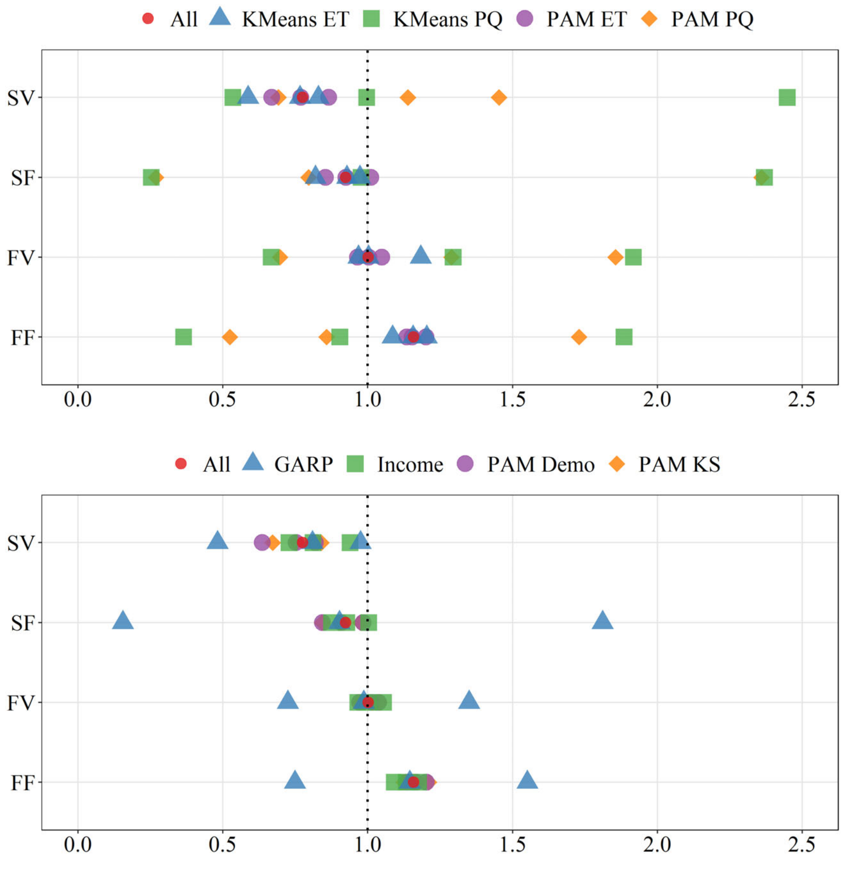

FIGURE 1 Median uncompensated expenditure elasticities by segmentation method: minimum, median, and maximum across 18 segments

method. All three segmentation methods using solely prices and quantities had a much higher proportion of non-overlapping confidence intervals. Comparing across all four expenditure elasticities and pairwise comparisons ($4 \times 153 = 612$), *GARP* segmentation produces the sharpest expenditure elasticity estimates. The percentage of non-overlapping pairs of confidence intervals on expenditure elasticity estimates was *GARP*, 56.7%; *KMeans PQ*, 31.5%, and *PAM PQ*, 43.3%. In stark contrast, the next highest percentage of non-overlapping pairs of confidence intervals was the *KMeans ET* method with just 2.5%. The *PAM KS* and *Random* methods resulted in no non-overlapping pairs of confidence intervals expenditure elasticity estimates. Recapping, only the algorithms using price-quantity data alone resulted in some proportion of expenditure elasticities with relatively small standard errors across segments. Poor results using all variables jointly, including prices and quantities—*PAM KS*—suggest that the using other variables besides prices and quantities dilutes or contaminates the segmentation process.

For policy purposes it is important to gauge whether the estimated expenditure elasticities are statistically distinguishable from 1. To check how many segments could be categorized statistically as inelastic or elastic, we conducted tests for differences from unity using the $\sqrt{\ln(n)}$ critical value. Even though *GARP* did not produce the most elastic expenditure elasticity values, *GARP* yielded the highest number of segments, 57 of 72 (79.2%), with expenditure elasticities statistically different from unity. Despite the highly elastic values generated in a few segments, *PAM PQ* and *KMeans PQ* resulted in 46 (63.9%) and 36 (50.0%) segments with elasticities different from unity. Except for

Income (36.1%), all the other segmentation methods resulted in about 60% of the segments with expenditure elasticities different from unity.

As a caveat, the foregoing comparison of expenditure elasticities uses elasticities evaluated at sample medians. Within each segment there may be sizable variation in expenditure elasticities from household to household. As a result, households in separate segments likely have very similar expenditure elasticities. But in terms of central tendency, the households within each segment identified by the segmentation methods using price and quantity differ from households in other segments. Put differently, a policy designed to affect expenditures on fruits and vegetables will generally have substantially different impacts on household consumption across segments identified by *GARP*, *KMeans PQ*, and *PAM PQ* methods.

4.2 | Own-price elasticities

Uncompensated own-price elasticities were evaluated at segment-specific medians. All own-price elasticities evaluated at median values for all segments were negative. Only 5 of the 624 own-price elasticity estimates were not significantly different from zero.

Figure 2 indicates the three segmentation methods using only prices and quantities yield the widest range of own-price elasticity values, from highly inelastic to very elastic. But in contrast to the expenditure elasticity results, *GARP* produces the widest range of own-price elasticity values. The most pronounced case is for storable vegetables where *GARP* own-price elasticities range from −0.145 to −1.689 for storable vegetables. As with expenditure elasticities, the other segmentation methods gave median own-price elasticities that tend to cluster tightly about the median estimates obtained from the entire sample (red circles in Figure 1).

To gauge the statistical precision of the own-price estimated elasticities, we again make pairwise comparisons of confidence intervals to check for overlap. Here again, elasticities from *GARP* segments exhibit more precision than those in *KMeans PQ* and *PAM PQ* segments. Percentages of elasticity pairs with non-overlapping confidence intervals are as follows: *GARP*, 57.4%; *KMeans PQ*, 33.5%, and *PAM PQ*, 39.1%. The next highest percentage of statistically different pairs is for *KMeans ET* with just 0.16%. The other segmentation methods result in own-price elasticities with totally overlapping confidence intervals across segments.

As with expenditure elasticities, we ascertained which segments could be deemed statistically significantly elastic or inelastic by testing for differences from −1. The number of inelastic segments is higher for vegetables than for fruits. Fresh fruits and vegetables also had more inelastic segments than their storable counterparts. The methods using only price-quantity data produced the highest number of inelastic segments across all four goods: *PAM PQ* with 81.5%, *KMeans PQ* with 72.6, and *GARP* with 71.4%. Nearly all segmentation methods yielded elastic segments for fruits but only *GARP* produced any segments (3) with elastic own-price elasticities for both fresh and storable vegetables.

In general, the segmentation methods using only price-quantity data produced segments that were statistically both inelastic and elastic for each of the four goods, whereas all other methods produce predominately either inelastic or elastic segments but not both. *GARP* segmentation resulted in the highest number of elasticities, 55 (76.4%), that were statistically distinguishable from −1. Both *PAM PQ* and *KMeans PQ* methods had more statistically distinguishable elasticities—73.6% and 65.3%—than all remaining methods.

The same caveat about using elasticity values evaluated at sample medians applies for own-price elasticities as for expenditure elasticities. Nonetheless, it is clear that *GARP* segmentation and, to a slightly lesser extent *KMeans PQ* and *PAM PQ* segments, display a substantial range of sensitivity to own-price changes for fruits and vegetables. A high percentage of non-overlapping confidence intervals indicates that *GARP* produces the most precise own-price elasticity estimates across segments.

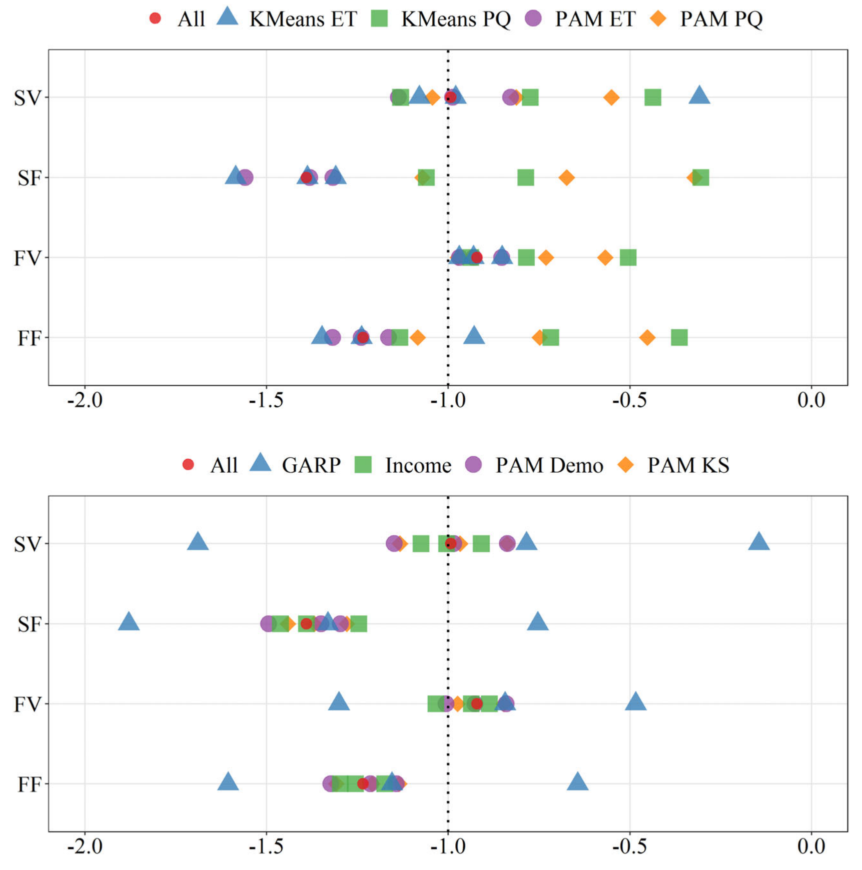

FIGURE 2 Median uncompensated own-price elasticities by segmentation method: minimum, median, and maximum across 18 segments

4.3 | Cross-price elasticities

We limit our discussion to uncompensated cross-price elasticities between the two vegetable and two fruit categories, fresh versus storable.[11] Using estimated confidence intervals as calculated for the previous elasticities, many fewer cross-price elasticities, 67.3%, are statistically different from zero. The range of cross-price elasticities displayed in Figure 3 indicates that fresh and storable vegetables are complements in nearly all segments irrespective of segmentation method. Just three fresh vegetables-storable vegetables cross-price elasticity estimates were positive, but they were not significantly different from zero. One *GARP* segment has a median cross-price elasticity of −0.787, indicating an increase in fresh vegetable prices would elicit a decrease in demand for both vegetable categories.

For fresh and storable fruit, the range of median cross-price elasticities indicates both substitution and complementary relationships depending on the segment. At one extreme, both *KMeans PQ* and *PAM PQ* methods identify a segment exhibiting a strong complementary relationship between fresh and storable fruits with cross-price elasticities of −0.972 and − 0.982 (percent change in storable fruit price eliciting a decrease fresh fruit consumption). At the other extreme, all three methods using only price and quantity identify households with moderate substitution relationships with storable-fruit price to fresh-fruit consumption elasticities of 0.468 for *GARP*, 0.377 for *KMeans PQ*, and 0.396 for *PAM PQ*.

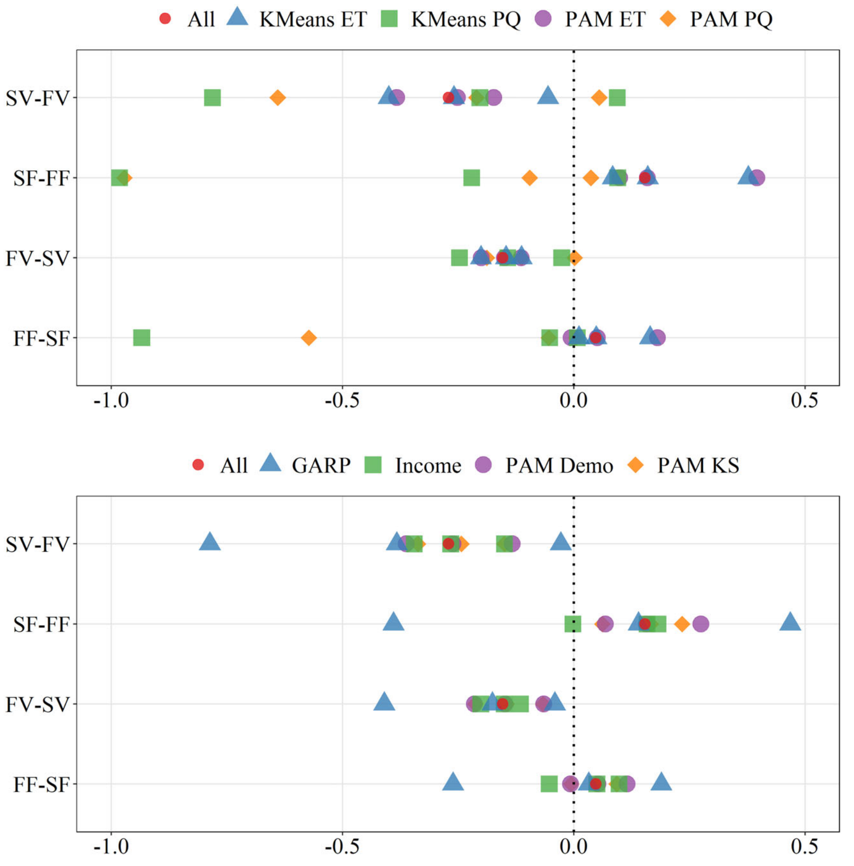

FIGURE 3 Median uncompensated cross-price elasticities by segmentation method: minimum, median, and maximum across 18 segments

Segmentation methods employing variables other than prices and quantities again tend to generate elasticities with little dispersion about those estimated from the entire sample. For several of the segmentation methods—*KMeans ET*, *PAM ET*, and Random—none of the cross-price elasticities displayed non-overlapping confidence intervals. Methods using demographic variables performed no better; the percentages of non-overlapping confidence intervals for pairs of cross-price elasticities were 1.1% for *PAM Demo* and 1.0% for *PAM KS*. By contrast, the percentages for methods relying only on prices and quantities were *GARP*, 37.3%; *KMeans PQ*, 23.2%; and *PAM PQ*, 32.0%.

5 | OUT-OF-SAMPLE SEGMENTATION FOR 2017

In supervised machine learning, obtaining good predictions out of sample is the gold standard for judging algorithm performance. Prediction out of sample with unsupervised learning is nonsensical because there is no dependent or outcome variable for comparing actual versus predicted values. Nonetheless, we will "predict" segments identified in 2016 to 2017 as a measure of how well individual segmentation methods work out of sample. Hence, we assign each household in 2017 to the same segments identified using only 2016 data.

As a preliminary measure of out-of-sample performance for *GARP*, we checked whether each household displayed consistent preferences across both years (see Figure 4). Consistent preferences mean there were no violations of GARP across years for a given household. Perhaps surprisingly, 99.2% of all households exhibited consistent preferences during the two-year span. This result is not an artifact of household budget constraints for fruits and vegetables shifting out: 50.9% of the households spent less in nominal terms in 2017. Of the 230 households with violations of GARP, 18 had some change in demographic or geographic variables, indicating changes in household income or a move to a different location. Less than 1% of households display evidence of a change in tastes and preferences as evidenced by violations of GARP.

For all segmentation methods, we compare estimated expenditure and price elasticities across the two years as a measure of whether the economic behavior of households in segments apparently changes with time. To that end, we match each elasticity for each segment across 2016 and 2017 (12 elasticities x 156 segments), and tally the percentage of elasticities with overlapping confidence

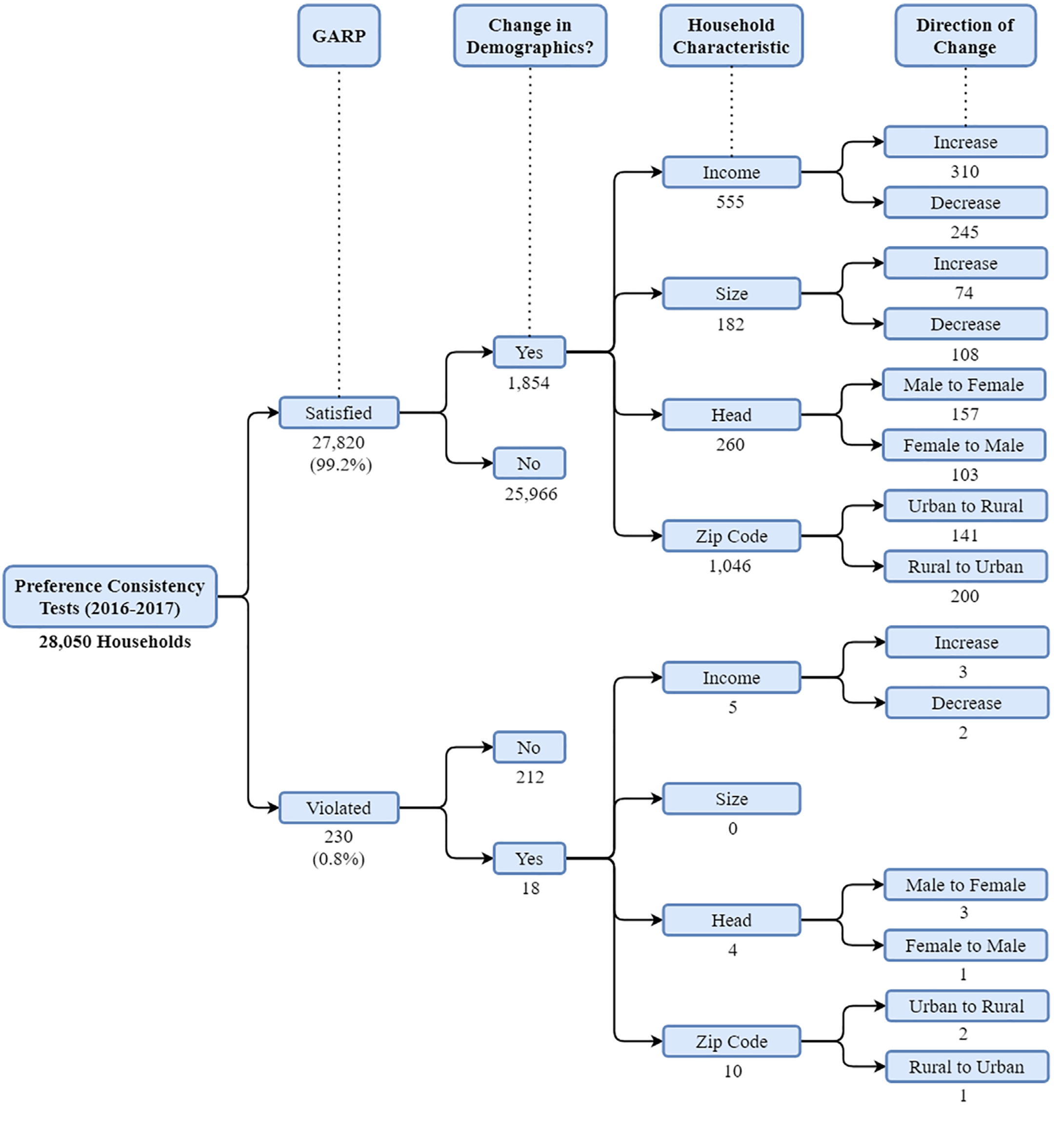

FIGURE 4 Consistency of preferences, GARP segments, individual households 2016–2017

intervals by type of elasticity and segmentation method (see Table 6). If the confidence intervals for each elasticity for 2016 and 2017 overlap, similar economic behavior will be implied for households in a particular segment across two years. The majority of the elasticities in each segment for each segmentation method overlap, ranging from 97.8% (Random) to 73.7% (PAM PQ).

The highest percentages of overlap occur for the methods that produced segments with seemingly homogeneous economic behavior. The lowest percentages of overlap occur for the three methods using solely prices and quantities. The smaller percentage of overlap is consistent with those three methods producing the widest range of elasticity values across segments in 2016. The range of elasticity values is given in the last six columns of Table 6 with minimum and maximum values irrespective of fruit and vegetable category. Though all three methods using price and quantity exhibit wider ranges than do the other methods, GARP segmentation produces the widest range of all methods in 2017 just as it did in 2016. In short, segmentation using GARP produces the widest range of elasticities with the smallest confidence intervals that are stable over the two years.

6 | GARP SEGMENTATION

To illustrate how segmentation might inform retail strategies or public policy, we examine own-price and expenditure elasticities for the18 GARP segments in Figure 5 in which heatmap colors indicate the absolute values of elasticities in each segment. Reading from the southwest to northeast in each panel, we encounter segments with higher median quantities and prices for each group of fruits and vegetables. Focusing on the left four panels with own-price elasticities, retailers would like to offer price incentives via coupons or frequent shopper discounts to households in segments in the northeast portion with yellow to red heatmap values because those households spend more on fruits and vegetables, and would respond significantly to price incentives. To boost fruit and vegetable consumption, public policies could target households with relatively lower median consumption toward the left of each panel. But providing price incentives like Supplemental Nutrition Assistance Program, "Double Dollars" and "Double Up Food Bucks" would be largely ineffective for segments with blue (own-price inelastic) values (Nourish Colorado, 2019; Wholesome Wave, 2019). Instead, social policies would need to be implemented to boost consumption fruits and vegetables for those households.

TABLE 6 Elasticity comparisons, 2016 and 2017

Segmentation method	Overlapping confidence intervals, 2016 vs. 2017 (%)	Minimum and maximum elasticities, 2017					
		Expenditure		Own price		Cross price	
		Min.	Max.	Min.	Max.	Min.	Max.
GARP	85.4	0.293	1.467	−1.504	−0.396	−0.472	0.377
KMeans PQ	77.3	0.734	1.361	−1.008	−0.473	−0.454	−0.040
PAM PQ	73.7	0.766	1.331	−0.999	−0.505	−0.476	−0.046
PAM Demo	97.3	0.618	1.211	−1.318	−0.561	−0.418	0.209
KMeans ET	97.2	0.494	1.208	−1.343	−0.577	−0.421	0.169
PAM ET	97.2	0.663	1.202	−1.322	−0.580	−0.435	0.164
PAM KS	94.6	0.628	1.222	−1.288	−0.544	−0.407	0.189
Income	91.1	0.691	1.175	−1.278	−0.585	−0.444	0.129
Random	97.8	0.696	1.197	−1.319	−0.634	−0.452	0.215
All	63.6	0.772	1.157	−1.224	−0.767	−0.321	0.092

Note: Minima and maxima of elasticities were calculated across all four categories of fresh and storable fruits and vegetables.

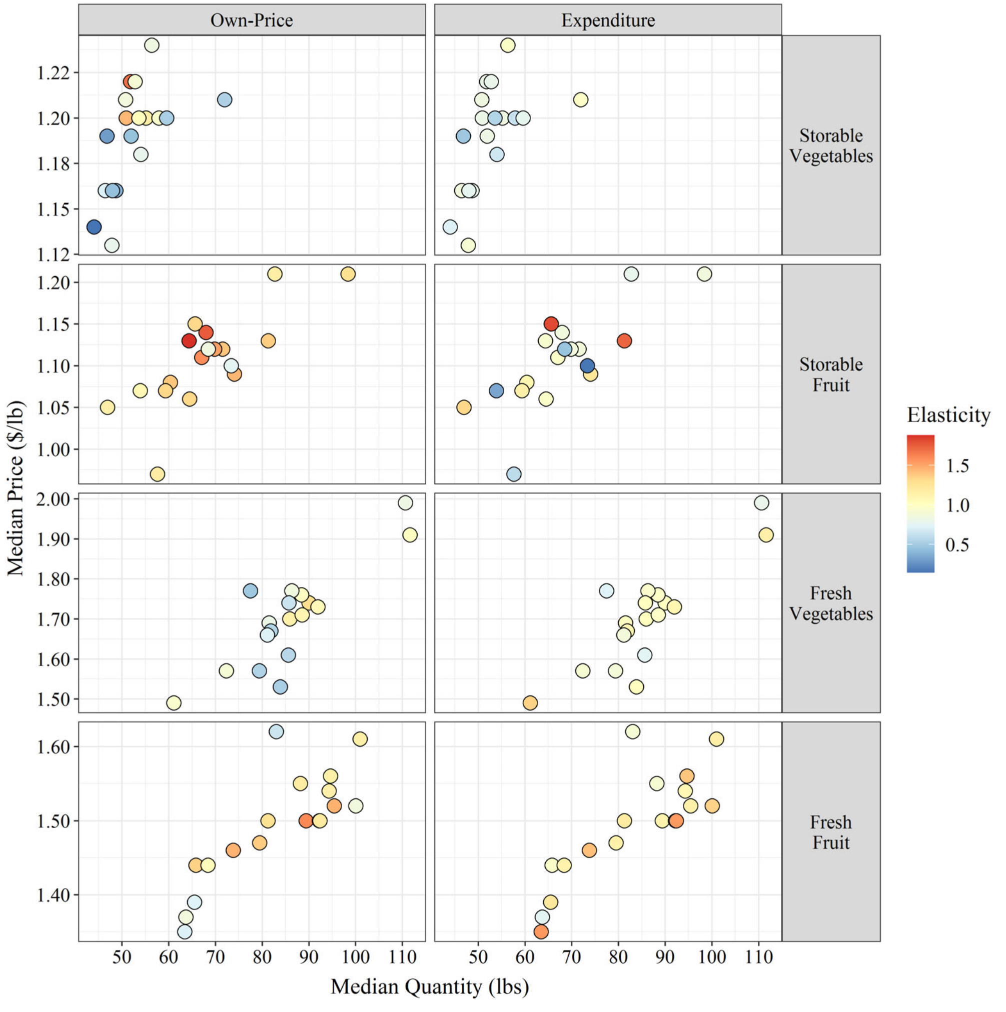

FIGURE 5 Own-price and expenditure elasticities for GARP segments by product type and median prices and quantities

Focusing on the four right panels with expenditure elasticities and comparing across all four groups of fruits and vegetables, increases in disposable income that augment expenditures on fruits and vegetables would tend to favor fresh products over storable (there are more blue segments for storable products). Even so, within each of the four panels, there is considerable variation in expenditure elasticities. For example, household segments with nearly identical median quantities and prices of storable vegetables have highly inelastic and highly elastic expenditure elasticities, evidence of heterogeneous economic behavior.

7 | ROBUSTNESS CHECKS

The foregoing in- and out-of-sample results are contingent on the nonlinear AIDS model being the correct specification. As a partial check for misspecification of the demand model, we estimated the quadratic AIDS (QUAIDS) model (Banks et al., 1997) for all segments in 2016. Likelihood ratio tests were mixed with the QUAIDS model preferred in just over half (76) of the 156 segments. But because we are interested in the elasticities rather than the parameter estimates, we checked for overlap in the confidence intervals for each AIDS elasticity with its paired QUAIDS elasticity. Most of the approximate standard errors estimated by the delta method were relatively small, yielding tight confidence intervals that seldom overlapped across segments irrespective of standard or $\sqrt{\ln(n)}$ critical values.

As a last informal comparison of differences across the two model specifications, we compared the pairs of smallest and largest elasticities for both AIDS and QUAIDS across 18 segments to check whether the range of economic behavior indicated by minimum and maximum elasticity values differed across the two specifications. For all expenditure, own- and cross-price elasticities, the differences were negligible. Though there may be statistical differences between the QUAIDS and AIDS elasticities, the ranges of price and expenditure sensitivity—that is, the economic behavior implied by those elasticities—were virtually the same. We take this as evidence that the foregoing in- and out-of-sample results appear robust to the QUAIDS alternative specification of the demand model.

8 | CONCLUSIONS

In contrast to much of the recent work integrating *supervised* machine learning with applied econometrics for purposes of *ex post* policy evaluation, we employ *unsupervised* machine learning and revealed preference methods to explore how they may be useful in policy design *ex ante*. More specifically, we examine the efficacy of these two nonparametric methods—unsupervised clustering techniques and revealed preference algorithms—in identifying segments of households with differing sensitivities to price and expenditure changes. In our application to demand for fruits and vegetables in a panel of 28,050 households, we find that algorithms using solely prices and quantities are the most effective at identifying segments displaying the largest range of elasticity values. Given persistent low levels of fruit and vegetable consumption in the United States, the heterogeneous economic behavior across segments suggests boosting consumption would require a combination of price, income, and social policies. By contrast, algorithms using demographic, geographic, and trip-related variables, even jointly with prices and quantities, produce segments of households with very similar elasticity values. That demographic variables may not be good proxies for differences in preference is consistent with Gross' findings of heterogeneity in preferences for local public goods (Gross, 1995b).[12]

Both the revealed preference algorithm, GARP, and the two clustering algorithms using only prices and quantities—k-means and partitioning on medoids—produced segments with wide ranges of both expenditure and price elasticities. However, the elasticities estimated in GARP segments tended to be more precise with tighter confidence intervals. Greater precision led to a higher proportion of elasticities being statistically significant and different from unity. Additionally, GARP segments produced the most stable elasticity estimates between 2016 and 2017. Given the ability to check for the minimum number of utility functions necessary to rationalize all households with revealed preference algorithms, segmentation with GARP seems to hold an edge over clustering algorithms, which can produce highly varying numbers of segments depending upon the measure chosen.

Implementing these nonparametric methods on a representative sample of data can enable researchers and policymakers to gauge how specific policy measures may influence different segments of the population. One drawback of our findings is that households from anonymized data sources cannot be individually targeted using their price and quantity data. If significant heterogeneity in preferences arises, a wide breadth of incentive mechanisms that allows people to self-select based on their preferences would be necessary to match policies with household preferences.

ACKNOWLEDGMENTS

We gratefully acknowledge support from the United States Department of Agriculture, National Institute of Food and Agriculture Specialty Crop Research Initiative program award (2015-51181-24283), Subaward No. 201504249-02. We also thank the United States Department of Agriculture, Economic

Research Service for facilitating access to the *Consumer Network* data compiled and maintained by Information Resources, Inc. (IRI). The analysis, findings, and conclusions expressed in this article should not be attributed to IRI.

ENDNOTES

1 Other transformations of the variables using z-scores or correlation coefficients have been developed.

2 We would like to thank the Economic Research Service (ERS) for granting us access to the NCP data collected by Information Resources, Inc. (IRI). National Consumer Panel data are now jointly supported by Information Resources, Inc. (IRI) and Nielsen (https://www.ncppanel.com/). For years prior to 2008, ERS purchased panel data from Nielsen who market the data as *Homescan*. From 2008 to the present, ERS has purchased the national panel data from IRI who refer to the data as *Consumer Network*.

3 "Static" households must report at least one purchase for 11 out of 13 four-week reporting periods and meet a minimum weekly spending requirement depending on household size: \$25/week for single-person households, \$35/week for two-person households, and \$45/week for households of three persons or more.

4 The nonlinear AIDS model is sufficiently well known that reproducing the notation for the model does not merit space here. Citations of Deaton and Muellbauer's seminal 1980 article introducing the nonlinear AIDS model indicate how well known the AIDS model is: Google Scholar citations, 6345; Web of Science, 1756 citations (accessed 8/17/2020).

5 The results of both algorithms depend on the order in which the households are drawn so we reran the algorithm 5000 times to check robustness of the results. We also ran the algorithms using only the 2017 household data and to our surprise found 18 utility functions to be the minimum number necessary to rationalize the data.

6 With 18 segments, some of the unsupervised machine learning algorithms yielded segments with as few as 75 households. If we had chosen a larger number of segments for those algorithms, we would have had to estimate nonlinear AIDS models on samples of untenably small numbers of households.

7 Several data-driven methods for determining the number of clusters were tested; however, results were widely inconsistent depending on method, similarity variables, and clustering algorithm. Using k-means with only price and quantity variables resulted in an optimal k varying from 3 using "silhouettes" (Rousseeuw, 1987) to 30 using the Caliński-Harabasz index (1974). The most consistent method across different variable combinations was the commonly used "elbow" method, which involves plotting the total within-cluster sum of squares for different values of k and choosing a value for k where the slope changes from steep to gradual (Hastie et al., 2008, p. 519). When testing $k = 2$ to 100, many variable combinations resulted in a choice of k from 10 to 20 using the elbow method.

8 We contemplated segmentation similar to that of Jensen and Manrique but without a continuous measure of income, we could not specify Engel curves as a basis for segmentation. Though there are methods for imputing continuous income values based on income categories (see Ferrier and Zhen, 2017), we choose to employ the categorical income measures as given in NCP.

9 We considered calculating confidence intervals with a nonparametric bootstrap but the computational burden of doing so with 20 uncompensated elasticities in each of 156 segments was excessive.

10 We do not perform hypothesis tests for equality of elasticity medians across segments. Hypothesis tests of equality of means are not equivalent to checking whether the standard confidence intervals of the two means overlap (Goldstein & Healy, 1995). However, checking for overlapping confidence intervals for each pair of elasticity medians is conservative for two reasons. First, any pair of estimates with overlapping standard confidence intervals may still be statistically different at a given significance level. Second, we use wider adjusted confidence intervals based on the Bayesian information criterion. If these wider confidence intervals do not overlap, equality of medians will always be rejected. However, because we are making multiple comparisons, we recognize that we should control for false-discovery rates. We account indirectly for false-discovery rates using $\sqrt{\ln(n)}$ instead of 1.96 as a critical value for hypothesis testing and in construction of confidence intervals. Calculating the family-wise error rate for differences in all pairs of elasticities (Efron & Hastie, 2016, p. 274) is beyond the scope of this study. We thank an anonymous referee for alerting us to the issue of false-discovery rates in this context.

11 Although some degree of substitution or complementarity may exist between fruits vis-à-vis vegetables, in the context of our four categories, we think substitution and/or complementarity between perishable and storable products is more plausible and interesting.

12 Gross states, "However, if demographics are not suitable proxies for differences in taste, traditional estimates of price and income elasticities may not be reliable" (Gross, 1995b, p. 104).

REFERENCES

Afriat, Sidney. 1967. "The Construction of a Utility Function from Expenditure Data." *International Economic Review* 8: 67–77.

Athey, Susan. 2018. "The Impact of Machine Learning on Economics." In *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb, 507–47. Chicago: University of Chicago Press. https://www.nber.org/system/files/chapters/c14009/c14009.pdf

Athey, Susan, and Guido Imbens. 2017. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives* 31(2): 3–32.

Athey, Susan, and Guido W. Imbens. 2019. "Machine Learning Methods that Economists Should Know About." *Annual Review of Economics* 11: 685–725.

Aune, D., E. Giovannucci, P. Boffetta, L.T. Fadnes, N. Keum, T. Norat, D. Greenwood, E. Rioli, L. Vatten, and S. Tonstad. 2017. "Fruit and Vegetable Intake and the Risk of Cardiovascular Disease, Total Cancer and all-Cause Mortality—a Systematic Review and Dose-Response Meta-Analysis of Prospective Studies." *International Journal of Epidemiology* 46 (3): 1029–56.

Austin, Peter C. 2009. "Balance Diagnostics for Comparing the Distribution of Baseline Covariates between Treatment Groups in Propensity-Score Matched Samples." *Statistics in Medicine* 28: 3083–107.

Banks, James, Richard Blundell, and Arthur Lewbel. 1997. "Quadratic Engel Curves and Consumer Demand." *Review of Economics and Statistics* 79(4): 527–39.

Beatty, Timothy K.M., and Ian A. Crawford. 2011. "How Demanding Is the Revealed Preference Approach to Demand?" *American Economic Review* 101(6): 2782–95.

Belloni, A., V. Chernozhukov, I. Fernández-Val, and C. Hansen. 2017. "Program Evaluation and Causal Inference with High-Dimensional Data." *Econometrica* 85(1): 233–398.

Bertail, Patrice, and France Caillavet. 2008. "Fruit and Vegetable Consumption Patterns: A Segmentation Approach." *American Journal of Agricultural Economics* 90(3): 827–42.

Boelaert, Julien. 2019. "Revealed Preferences and Microeconomic Rationality" Package "revealedPrefs." https://cran.r-project.org/web/packages/revealedPrefs/revealedPrefs.pdf.

Boonsaeng, Tullaya, and Carlos E. Carpio. 2019. "A Comparison of Food Demand Estimation from Homescan and Consumer Expenditure Survey Data." *Journal of Agricultural and Resource Economics* 44(1): 117–40.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45: 5–32.

Caliński, Tadeusz, and Joachim Harabasz. 1974. "A Dendrite Method for Cluster Analysis." *Communications in Statistics* 3(1): 1–27.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge, UK: Cambridge University Press.

Casagrande, Sarah Stark, Youfa Wang, Cheryl Anderson, and Tiffany L. Gary. 2007. "Have Americans Increased their Fruit and Vegetable Intake? The Trends between 1988 and 2002." *American Journal of Preventive Medicine* 32(4): 257–63.

Crawford, Ian, and Krishna Pendakur. 2012a. "How Many Types Are There?" *Economic Journal* 123: 77–95.

Crawford, Ian, and Krishna Pendakur. 2012b. "Technical Appendix to How Many Types Are There?" *Economic Journal* 123: 1–6.

Dean, Mark and Daniel Martin. 2010. "How Rational Are your Choice Data?" *mimeo*. http://www.martinonline.org/daniel/DeanMartin_30June2010.pdf

Dean, Mark, and Daniel Martin. 2016. "Measuring Rationality with the Minimum Cost of Revealed Preference Violations." *Review of Economics and Statistics* 98(3): 524–34.

Deaton, Angus, and John Muellbauer. 1980. "An Almost Ideal Demand System." *American Economic Review* 70(3): 312–26.

Deaton, Angus. 1998. "Quality, Quantity, and Spatial Variation of Price." *American Economic Review* 78(3): 418–30.

Efron, Bradley, and Trevor Hastie. 2016. *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. Cambridge, UK: Cambridge University Press.

Einav, Liran, Ephraim Leibtag, and Aviv Nevo. 2010. "Recording Discrepancies in Nielsen Homescan Data: Are they Present and Do they Matter?" *Quantitative Marketing and Economics* 8: 2017–239.

Ferrier, Peyton M., and Chen Zhen. 2017. "The Role of Income in Explaining the Shift from Preserved to Fresh Vegetable Purchases." *Journal of Agricultural and Resource Economics* 42(3): 329–49.

Fraiman, Ricardo, Ana Justel, and Marcela Svarc. 2008. "Selection of Variables for Cluster Analysis and Classification Rules." *Journal of the American Statistical Association* 103(483): 1294–303.

Fu, Wei, and Patrick O. Perry. 2020. "Estimating the Number of Clusters Using Cross-Validation." *Journal of Computational and Graphical Statistics* 29(1): 162–73.

Gentzkow, M., B.T. Kelly, and M. Taddy. 2019. "Text as Data." *Journal of Economic Literature* 57(3): 535–74.

Goldstein, Harvey, and Michael J.R. Healy. 1995. "The Graphical Presentation of Collection of Means." *Journal of the Royal Statistical Society A* 158(Part 1): 175–7.

Gower, J.C. 1971. "A General Coefficient of Similarity and some of its Properties." *Biometrics* 27: 857–74.

Gower, J.S. 1985. "Measures of Similarity, Dissimilarity, and Distance." In *Encyclopedia of Statistical Sciences*, Vol 5, edited by S. Kotz, N.L. Johnson, and C.B. Read, 397–405. New York: Wiley.

Gross, John. 1995a. "Testing Data for Consistency with Revealed Preference." *Review of Economics and Statistics* 77(4): 701–10.

Gross, John. 1995b. "Heterogeneity of Preferences for Public Goods: The Case of Private Expenditure for Public Educations." *Journal of Public Economics* 77(4): 701–10.

Harbaugh, William T., Kate Krause, and Timothy R. Berry. 2001. "GARP for Kids: On the Development of Rational Choice Behavior." *American Economic Review* 91(5): 1539–45.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2008. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.

Hennig, Christian, and Chien-Ju Lin. 2015. "Flexible Parametric Bootstrap for Testing Homogeneity against Clustering and Assessing the Number of Clusters." *Statistics and Computing* 25: 821–33.

Jensen, Helen, and Justo Manrique. 1998. "Demand for Food Commodities by Income in Indonesia." *Applied Economics* 30: 491–501.

Ju, Cheng, Richard Wyss, Jessica M. Franklin, Sebastian Schneeweiss, Jenny Häggström, and Mark J. van der Laan. 2019. "Collaborative-Controlled LASSO for Constructing Propensity Score-Based Estimators in High-Dimensional Data." *Statistical Methods in Medical Research* 28(4): 1044–63.

Kaufman, Leonard, and Peter J. Rousseeuw. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ: John Wiley and Sons, Inc.

Lee-Kwan, Seung Hee, Latetia V. Moore, Heidi M. Blanck, Diane M. Harris, and Deb Galuska. 2017. "Disparities in State-Specific Adult Fruit and Vegetable Consumption—United States, 2015." *MMWR Morbidity and Mortality Weekly Report* 66(45): 1241–7.

Lin, Biing-Hwan, Jean C. Buzby, Tobenna D. Anekwe, and Jeanine T. Bentley. 2016. *U.S. Food Commodity Consumption Broken Down by Demographics, 1994-2008*, Washington, DC: U.S. Department of Agriculture, Economic Research Service ERR-206, March.

Luo, Xing, Zhu Xu, and Eng Gee Lim. 2019. "A Parametric Bootstrap Algorithm for Cluster Number Determination of Load Pattern Categorization." *Energy* 180: 50–60.

Lusk, Jayson L. 2017. "Consumer Research with Big Data: Applications from the Food Demand Survey (FooDS)." *American Journal of Agricultural Economics* 99(2): 303–20.

Lusk, Jayson L., and Kathleen Brooks. 2011. "Who Participates in Household Scanning Panels?" *American Journal of Agricultural Economics.* 93(1): 226–40.

Maechler, Martin, Peter Rousseeuw, Anja Struyf, Mia Hubert, Kurt Hornik, Matthias Studer, and Pierre Roudier. 2019. "Package 'cluster'," version 2.0.9. https://svn.r-project.org/R-packages/trunk/cluster.

McKelvey, Christopher. 2011. "Price, Unit Value, and Quality Demanded." *Journal of Development Economics* 95(2): 157–69.

Mhurchu, Cliona Ni, Helen Eyles, Chris Schilling, Qing Yang, William Kaye, Murat Genç, and Tony Blakely. 2013. "Food Prices and Consumer Demand: Differences across Income Levels and Ethnic Groups." *PLoS One* 8(10): e75934.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31(2): 87–106.

Muth, Mary K., Megan Sweitzer, Derick Brown, Kristen Capogrossi, Shawn Karns, David Levin, Abigail Okrent, Peter Siegel, and Chen Zhen. 2016. *Understanding IRI Household-Based and Store-Based Scanner Data*. Washington, DC: U.S. Department of Agriculture, Economic Research Service, Technical Bulletin.

Nourish Colorado. 2019. Double Up Food Bucks. https://doubleupcolorado.org/

Park, John L., Rodney B. Holcomb, and Kellie Curry Raper, and Oral Capps Jr. 1996. "A Demand Systems Analysis of Food Commodities by U.S. Households Segmented by Income." *American Journal of Agricultural Economics* 78(2): 290–300.

Pradeep, A.K., Andrew Appel, and Stan Sthanunathan. 2019. *AI for Marketing and Product Innovation: Powerful New Tools for Predicting Trends, Connecting with Customers, and Closing Sales*. Hoboken, NJ: John Wiley and Sons, Inc.

Rousseeuw, Peter J. 1987. "Silhouettes: A Graphical Aid in the Interpretation and Validation of Cluster Analysis." *Journal of Computational and Applied Mathematics* 20: 53–65.

Storm, Hugo, Kathy Baylis, and Thomas Heckelei. 2020. "Machine Learning in Agricultural and Applied Economics." *European Review of Agricultural Economics* 47(3): 849–92.

Sweitzer, Megan, Derick Brown, Shawn Karns, Mary K. Muth, Peter Siegel, and Chen Zhen. 2017. *Food-at-Home Expenditures: Comparing Commercial Household Scanner Data From IRI and Government Survey Data*. Technical Bulletin Number 1946, U.S. Department of Agriculture, Economic Research Service, Washington, DC, September.

Theil, Henri. 1971. *Principles of Econometrics*. New York: John Wiley and Sons, Inc.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society Series B (Methodological)* 58(1): 267–88.

U.S. Census Bureau, American Community Survey 2016. American Community Survey 1-Year Estimates, Table S2501, generated using American FactFinder. http://factfinder.census.gov

Varian, Hal R. 1982. "The Nonparametric Approach to Demand Analysis." *Econometrica* 50(4): 945–74.

Varian, Hal R. 1983. "Nonparametric Tests of Consumer Behaviour." *Review of Economic Studies* 50(1): 99–110.

Varian, Hal R. 1985. "Nonparametric Analysis of Optimizing Behavior with Measurement Error." *Journal of Econometrics* 30(1): 445–58.

Varian, Hal. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28(2): 3–27.

Varian, Hal. 2018. "*Artificial Intelligence, Economics, and Industrial Organization*" NBER Working Paper 24839. http://www.nber.org/papers/w24839.

Wager, Stefan, and Susan Athey. 2017. "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests." *Journal of the American Statistical Association* 113(523): 1228–42.

Wang, Junhui. 2010. "Consistent Selection of the Number of Clusters via Crossvalidation." *Biometrika* 97(4): 893–904.

Wholesome Wave. 2019. https://www.wholesomewave.org/what-we-do (accessed: August 12, 2021).

Zhen, Chen, Eric A. Finkelstein, James M. Nonnemaker, Shawn A. Karns, and Jessica E. Todd. 2014. "Predicting the Effects of Sugar-Sweetened Beverage Taxes on Food and Beverage Demand in a Large Demand System." *American Journal of Agricultural Economics* 96(1): 1–25.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Blumberg, Joey, Gary Thompson. 2021. "Nonparametric segmentation methods: Applications of unsupervised machine learning and revealed preference." *American Journal of Agricultural Economics* 1–23. https://doi.org/10.1111/ajae.12257